Cybersecurity and its Impact on Users' Digital Safety: An Analytical Study on Social Media Threats

NAKIBUUKA JAMIRAH*, MOHAMAD FADLI ZOLKIPLI

School of Computing, College of Arts and Sciences, Universiti Utara Malaysia,
06010 Sintok, Kedah Darul Aman, Malaysia
*Email: *nakibuuka085@gmail.com, m.fadli.zolkipli@uum.edu.my | Tel: *+60132370875 | Tel: +60177247779|*

Abstract

This rapid expansion of social media has transformed human communication, interaction and information sharing but it has created fertile ground for new cybersecurity threats that compromise users' digital safety. This study investigates the impact of cybersecurity threats particularly cyberbullying on users' privacy, identity, reputation and mental health across platforms like Instagram, X, TikTok, Facebook, LinkedIn and others. It examines the cybersecurity implications of social media threats by reviewing scholarly literature and case studies to identify measures. Findings reveal that age, language and platform design significantly influence vulnerability to cyberbullying while current mitigation strategies remain fragmented and reactive. Also reveals that weak privacy settings, anonymity and poor digital illiteracy contribute to the escalation of social media attacks, while emerging technologies such as artificial intelligence (AI) and machine learning are being leveraged to detect and prevent cyberbullying and misinformation. Additionally, emerging technologies such as AI-based content detection demonstrate potential but remain limited by language barriers, data bias and inconsistent enforcement. This study proposes an integrated digital-safety framework that combines platform governance, legal reform, user awareness and AI-driven monitoring. The insights contribute toward building safer social media ecosystems that promote accountability, digital well-being and cybersecurity resilience.

Keywords: **Cybersecurity, Digital Safety, Social media threats, cyberbullying, Identity theft, Mental health, Online Harassment.**

1 Introduction

1.1 Background and problem statement

Digital safety refers to the ability of individuals to use online technologies and social media responsibly while safeguarding their personal data, privacy and psychological wellbeing. It encompasses awareness of online risks, ability to identify harmful activities and the consistent use of safe digital practices that reduce exposure to cyber threats and malicious content (Ray et al., 2024). Social media threats are harmful behaviors or technical exploits that occur on platforms such as Facebook, X, Instagram and TikTok. These threats undermine user trust, compromise privacy and cause psychological, reputational and financial harm (Kadhim et al., 2025). The evolution of social media platforms into daily life has intensified risks by embedding social media into deeper complexes threatening digital safety (Rana & Shukla, 2024). Platforms like X, Instagram, Facebook and TikTok have become central to personal and professional life yet their openness and algorithmic design create an environment where misuse, harassment and data exploitation can occur easily (Han et al., 2024). Studies show that cyberbullying significantly affects mental health, privacy and digital trust particularly among adolescents and young adults (Collins, 2024). The anonymity provided by online platforms allows perpetrators to operate without accountability while harmful messages and images can circulate widely before moderation systems respond (Obaida et al., 2024). Moreover, users' limited cybersecurity awareness and poor digital literacy further heighten their vulnerability to cyberattacks and online harassment (Mbonimpa et al., 2024). Additionally, looking into legal frameworks and inconsistent enforcement across jurisdictions where countries have established legal mechanisms like the communications and multimedia act initiatives against cyberbullying, enforcement remains fragmented and reactive (Kamarudin et al., 2025). Weak institutional collaboration and limited awareness campaigns have hindered effective implementation with these evolving social media technologies are protected by stronger integration between law, technology and user education is essential to ensure digital resilience (Hendrayani, 2024).

The study bridges technical, behavioral, and policy perspectives, this research contributes to the growing disclosure on cybersecurity in social media. It offers actionable insights for platform developers, educators and policymakers, emphasizing the ethical balance between free speech and protection from harm. So this study aims to empower stakeholders to create safer and more accountable digital environments. Ultimately, the study explores the relationship between cybersecurity and users' digital safety within social media environments, seeks to examine the prevalence and typologies of social media threats, assesses their effects on privacy, identity and mental well-being and also evaluating emerging mitigation strategies such as AI-driven detection systems, policy interventions and public awareness programs. The findings aim at informing educators, policymakers and technology developers in strengthening online safety, reducing cyberbullying and promoting responsible digital citizenship.

1.2 Objectives and research questions

This paper pursues four objectives: (i) Studying how cyberbullying travels on social media and the eligibility release process. (ii)To see how social media threats are causing problems for people's lives, rights and justice. (iii) To summarize how technology, law and policy advice for dealing with cyberbullying and (iv) Demand for a digital-safety system increases regulation, technology and consumer intelligence. We address:

- RQ1: What increases the vulnerability of social media users in terms of cyberbullying?
- RQ2: What are the implications of social media authentication requests for cyberbullying?
- RQ3: What are the competencies for combating cyberbullying such as use of technology, law and human knowledge?
- RQ4: Where should platforms, rules, and social media users' intelligence practices adapt to digitally posting authentication?

1.3. Scope and contributions

We review peer-reviewed literature years (2022–2025) on academic, healthcare, technology, and social media assurances, preceding on different books, how practical results on how to combat cyberbullying, human qualifications, and affirmation schemes. Our contributions are: (1) Advertising strategy pathways from exposure on social media to health, eligibility and public assurance; (2) Study and analyze how technologies and programs have been developed to combat cyberbullying and other forms of social media abuse; (3) developing a framework for working together on technology, law, and consumer intelligence (hybrid approach) in combating harassment; and (4) Advice on how to request authentic, non-legal and irrelevant authentication for social media, and also to store and maintain administrative and system data.

2 Literature Review

2.1 Defining Cyberbullying

Cyberbullying has merged emerged as one of the most pervasive threats to digital safety on social media platforms. Defined as the intentional use of digital communication to harass, intimidate or humiliate individuals. It manifests in various forms ranging from flaming and exclusion to impersonation and cyberstalking (Ray et al., 2024). Scholars have extensively explored the nature, prevalence and consequences of these threats while others have examined emerging technological and policy interventions aimed at addressing them. This section critically reviews existing studies under six key dimensions that is, definitions, prevalence trends, theoretical models, impacts, mitigation, strategies and research gaps.

2.2 Overview of Cyberbullying

According to (Kadhim et al., 2025), common forms include flaming, exclusion, impersonation and cyberstalking. These acts often exploit the lack of effective monitoring and weak enforcement of online conduct policies. (Han et al., 2024) shows with deep learning models that cyberbullying often involves aggressive language and emotional hostility on platforms like X and Facebook. Similarly, (Hendrayani, 2024) also finds that anonymity is a key factor in online aggression, as users use pseudonyms to avoid accountability. Collectively , these studies highlight that cyberbullying is both a social and cybersecurity issue because it takes advantage of digital communication systems to cause harm.

2.3 Prevalence Trends

The prevalence of cyberbullying is well-documented across global and regional contexts(Kustanti et al., 2024) show that young people are the most vulnerable to social media abuse. In Malaysia, (Ooi et al., 2022) observed that over 35% of college students had experienced cyberbullying or perpetrated it against peers. In Vietnam, (Do & Pham, 2025)

observed that tertiary school students were more likely to be involved in social media violence, with this influence being watered down by the use of recording boards and rampant text messages. According to (RUSLI et al., 2024) found that Malaysian students recognize cyberbullying as a major problem, but do not report it due to fear and uncertainty about how they can be helped. These findings collectively demonstrate that cyberbullying is becoming increasingly ubiquitous across Southeast Asia, amplified by socio-economic insecurities, limited digital literacy and inadequate systemic innervations.

2.3.1. Theoretical Frameworks
Security theoretical frameworks address how cyberbullying takes place like the Online Disinhibition Effect, discussed by (Hendrayani, 2024) which mentions that anonymity, invisibility and asynchronous communication diminish social inhibitions people who consume information online do not conform to their own culture, and the lack of other people's faces leads them to commit sins of treason. Another theory called the Routine Activity Theory stated by (Ray et al., 2024) addresses cyberbullying through the convergence of the three like a motivated offender, a suitable technological medium and absence of capable guardians such as platform moderators. Legal enforcement mechanisms or digital literacy safeguards. These theories collectively underscore that cybersecurity the nature of cyber offences talking about the killer, the unfriendly tool, and the absence of standby "guardians" with technical or legal experience. These theories collectively underscore that cyberbullying emerges from a complex interplay of individual psychology, technological affordances and environmental vulnerabilities , thereby necessitating integrated mitigation frameworks.

2.3.2. Impacts of Cyberbullying
Cyberbullying and social media violence are often to blame for people's intellectual and mental stability. (Rana & Shukla, 2024) argue that victims of bullying have a compromised identity, fear, and some withdraw from the sites. (Mbonimpa et al., 2024) observed that students with no cybersecurity background do not even know how to report violence, and this makes them more vulnerable. (Redha Mutar, 2024) also points out that many victims are caught up in violence when they are the first to reveal their secrets.

(Obaida et al., 2024) demonstrates the capability of deep learning algorithms to detect cyberbullying with high accuracy across multiple social media platforms. This shows that machine intelligence can access emotional hostility in real time, enabling proactive intervention. Region-specific technological studies underscore linguistic diversity challenges; (Singh & Othman, 2025) propose a transformer-based detection model tailored for Malay-language content, demonstrating improved performance on local datasets. Beyond law and technology, (Collins, 2024) and (Shah et al., 2023) emphasize the importance of moral education and culturally adaptive digital community frameworks, arguing that behavioral change and values-based learning are essential for long-term prevention.

2.3.4. Mitigation Strategies
Preventing cyberbullying calls for the use of law, network security, and machine intelligence (AI). (Nurulhuda Ahmad Razali & Nazli Ismail Nawang, 2022) point out that in Malaysia, laws such as the Communications and Multimedia Act help in terms of cyberbullying, but do not equate to effective laws that should reduce cyberbullying. (Kamarudin et al., 2025), argue that the government needs to impose uniform standards and unify leadership in cyber legal matters. In the field of intelligence, (Obaida et al., 2024) shows that the use of artificial intelligence (AI) helps to embrace real-time violent rumors. (Han et al., 2024) also argues that AI is used in negotiating and foreseeing the harmful effects of violence, while (Singh & Othman, 2025) show that the approach used in local languages (including Malay) uses different skills. This suggests that both law and intelligence need to unite in preventing cyber harm.

Although many researchers have addressed cyberbullying, there are still some unclear issues. Most researchers use only one method which is either technologically or sociologically without combining the two. Those that deal with how users perceive risks or with what villages can avoid have recently been adapted in those areas. In order to address these problems, it is expected that researchers will be a combination of expertise from technology, law, and pedagogy in integration, to ensure that cybersecurity receives quality leadership**.**

3   Methodology

The study uses a qualitative content analysis, based on publications written between 2022 and 2025. Its focus is on the use of information on social media in a way that complains about the safety and security of users. This approach is chosen because it can aggregate and bring together data from a variety of qualitative sources, reporting in an aggregate way how cyberbullying and other social media threats impact security in this country. The data used in this study came from peer-reviewed journals and conference proceedings that come from the harvest of subjects such as IEEE Access, Elsevier and Quantum Journal. Articles discussing issues of cyberbullying, digital safety, social media threats, and online safety laws are relied upon. The sources of these articles were selected in order to send a powerful message that outlines how law, technology and education intersect in preventing cyber threats.

Table 1: Criteria used to select high-quality and relevant academic materials for the review

Criteria	Inclusion Criteria	Exclusion Criteria	Source Databases
Publication Period	Articles published between 2022-2025	Publications before 2022	IEEE, Scopus, Google Scholar
Type of document	Peer-reviewed journals, conference papers	Blogs, news or non-academic reports	AMC, Elseveir
Topic Relevance	Studies focusing on cyberbullying, digital safety or social media threats	Articles outside cybersecurity or unrelated topics	IJLGC, Quantum journal
Language	English-Language publications	Non-English works	English journals
Research Quality	Peer-reviewed, verified sources	Non-reviewed, opinion-based works	University Repositories

Accepted publications were selected by thematic synthesis for meanings related to five key issues including cyberbullying patterns, frequency and hotspots, conflicts between law and technological authority, online safety implications, and harm avoidance and consumer protection practices. This selection has been used to conduct comparative analysis to see how different books deal with certain topics, and to find out the meaning that becomes significant in subsequent research (Han et al., 2024). The selected publications were understood by descriptive and comparative analysis. Strong terms such as "cyberbullying," "digital safety," "awareness," and "AI detection" were used in the selection of publications from the subject databases. Books discussing technologies such as AI and machine learning were placed in the corner of power to reflect the dynamics of smart security.

Figure2:Conceptual Framework illustrating the relationship between social media platforms, threats, user vulnerability, platform design, legal frameworks and mitigation strategies toward digital safety.

Although this study summarizes the meaning of power, some constraints apply. The lack of new data without selection on different populations was one of the problems that appeared, because this study relied on secondary data. The heterogeneity of legislation in different countries such as Malaysia, Vietnam and Uganda also had limited effects in modeling the existing (Kamarudin et al., 2025). Also, the records captured cover the years 2022 to 2025, and the future data could be unaltered in this study. These barriers were alleviated by attempts to use quality books from factual sources and a selection of selected academically responsible messages. More specifically, the study uses a tourism and documented assessment approach, to understand the impact of cyberbullying and social media violence on safety. This

initiative contributes to the demonstration that the power of technology, law, and pedagogy combine to protect people in different countries in the best way possible.

4 Discussion & Analysis

These results indicate that users' safety on social media sites is significantly affected by cyberbullying, lack of awareness about security, and the absence of strong protective legislation. The literature written between 2022 and 2025 shows an increase in malicious activity on the internet, especially in countries with increased technological and communication skills. A review of various literatures shows that cyberbullying is not just a social media sensation, but has become a problem that affects public safety and health. (Kadhim et al., 2025) points out that cyberbullying has changed from negative speech on social media to a form of harassment that affects privacy, safety of avoidance, and life expectancy. In the study of Ooi et al. (2022), many young Malaysians behave irresponsibly on social media, with some not even knowing how to maintain their safety. This suggests that lack of crime, lack of adequate legislation, and lack of consumer education have led to excessive safety. In a technological angle, (Han et al., 2024) point out that the use of machine intelligence (AI) in recognizing and handling cyberbullying does occur, but the under-understood characteristics of those devices have led to a lack of effective compliance. (Obaida et al., 2024) add that AI may foresee threats in time, but most of the facilities did not have specified legal courses on the use of technology in forecasting threats to networks. This suggests that there is a need for technological and legal integration to determine direct protection.

The selected literature also indicates that cyberbullying relates to issues of health safety, dignity, and trustworthiness online. (Rana & Shukla, 2024) shows that victims of cyberbullying can suffer from psychosis, academic decline, and negative self-concept. (Mbonimpa et al., 2024) also find that in universities, students with no cybersecurity background are more susceptible to hostility due to their lack of ability to sort out malicious messages. This meant that cybersecurity is not just a technological issue, but is an area that requires pedagogy, oversight, and effective communicated regulations. This assessment has further shown that changes in the use of laws on security have differences in different countries. (Nurulhuda Ahmad Razali & Nazli Ismail Nawang, 2022) point out that Malaysia does not have a single specified law on cyberbullying, although an existing set of laws such as the Communications and Multimedia Act recognizes the problem. (Kamarudin et al., 2025) call for the inclusion under a uniform set of rules to protect people online, and this is complemented by public education and improving experience. In order to understand how cyberbullying is used and its impact on safety, Online Disinhibition Effect theory and Routine Activity Theory proved to be very useful. This theory suggests that the lack of a capable guardian and the absence of laws defining technological components of networks contribute to the growth of cyber offences. This suggests that users' online behavior, as well as legal capabilities and expertise, are prerequisites for a collective approach to individual security.
In details, the results of the survey meant that cyberbullying and social media threats are becoming the most increasing problem every year in various countries, all of which are made worse by the lack of user experience and security assurances. These risks affect users physically, psychologically, and culturally, and require a hybrid assessment that revolves around technology, law, and education. The use of machine intelligence (AI), legal security, and enhanced user experience on social media could help in mitigating these risks.

differences in different countries(Nurulhuda Ahmad Razali & Nazli Ismail Nawang, 2022) point out that Malaysia does not have a single specified law on cyberbullying, although an existing set of laws such as the Communications and Multimedia Act recognizes the problem. (Kamarudin et al., 2025) call for the inclusion under a uniform set of rules to protect people online, and this is complemented by public education and improving experience.
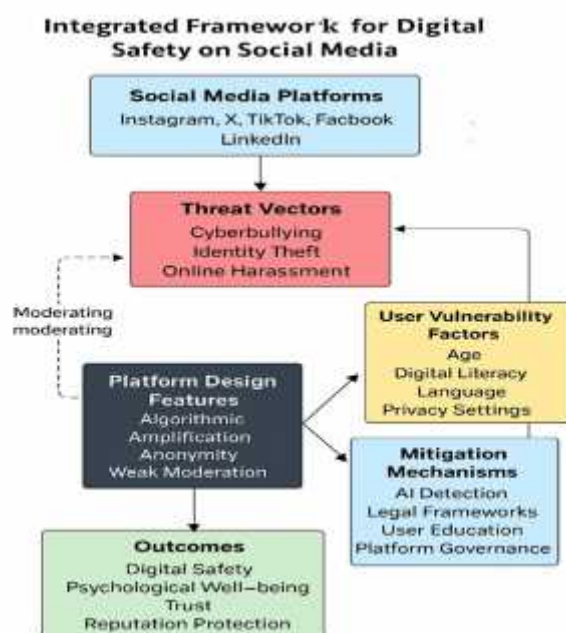

Figure 2: An Integrated Framework for Digital Safety on Social Media

5 Conclusion and Future Work

In order to understand how cyberbullying is used and its impact on safety, Online Disinhibition Effect theory and Routine Activity Theory proved to be very useful. This theory suggests that the lack of a capable guardian and the absence of laws defining technological components of networks contribute to the growth of cyber offences. This suggests that users' online behavior, as well as legal capabilities and expertise, are prerequisites for a collective approach to individual security. In details, the results of the survey meant that cyberbullying and social media threats are becoming the most increasing problem every year in various countries, all of which are made worse by the lack of user experience and security assurances. These risks affect users physically, psychologically, and culturally, and require a hybrid assessment that revolves around technology, law, and education. The use of machine intelligence (AI), legal security, and enhanced user experience on social media could help in mitigating these risks.

Acknowledgment

References

Collins, A. (2024). Mitigating Cyberbullying among Teens: The Crucial Role of Moral Education in School Curricula. *IAA JOURNAL OF EDUCATION*, *10*(1), 35–39. https://doi.org/10.59298/IAAJE/2024/10135.39

Do, J., & Pham, C. (2025, June 11). *EXPERIENCE OF CYBERBULLYING AND PRACTICES TO COPE WITH IT AMONG HIGHSCHOOL STUDENTS IN HANOI*. https://doi.org/10.17501/26138417.2025.8103

Han, H., Asif, M., Awwad, E. M., Sarhan, N., Ghadi, Y. Y., & Xu, B. (2024). Innovative deep learning techniques for monitoring aggressive behavior in social media posts. *Journal of Cloud Computing*, *13*(1), 19. https://doi.org/10.1186/s13677-023-00577-6

Hendrayani, P. F. A. (2024). DIGITAL LITERACY AND ONLINE TRUST TO AVOID CYBERBULLYING ANONYMITY AMONG TWITTER ROLE PLAYERS COMMUNITY. *Interaksi: Jurnal Ilmu Komunikasi*, *13*(1), 37–59. https://doi.org/10.14710/interaksi.13.1.37-59

Kadhim, A. A., Abdalrdha, Z. K., & Khuder Naser, W. A. (2025). Subject Review: Cyberbullying and Detection Methods. *International Journal of Advances in Scientific Research and Engineering*, *11*(03), 26–37. https://doi.org/10.31695/IJASRE.2025.3.3

Kamarudin, A., Kamarudin, I., & Abdullah Sanek, S. K. (2025). BRIDGING THE GAPS IN MALAYSIA'S CYBERBULLYING LAWS: CHALLENGES AND REFORM PROPOSALS. *International Journal of Law, Government and Communication*, *10*(39), 133–147. https://doi.org/10.35631/IJLGC.1039008

Kustanti, E., Afiatin, T., & Febriani, A. (2024). Overview of Bullying and Cyberbullying among Adolescents : Preliminary Study. *Proceedings of the 5th International Conference on Psychological Studies, ICPSYCHE 2024, 24 – 25 July 2024, Semarang, Central Java, Indonesia*. https://doi.org/10.4108/eai.24-7-2024.2354325

Mbonimpa, T., Richard, N., Innocent, M. J., & Priscilla, M. (2024). Investigating Security Awareness and Incident Reporting levels at Mbarara University of Science and Technology. *Indonesian Journal of Innovation and Applied Sciences (IJIAS)*, *4*(3), 208–216. https://doi.org/10.47540/ijias.v4i3.1482

Nurulhuda Ahmad Razali, & Nazli Ismail Nawang. (2022). AN OVERVIEW OF THE LEGAL FRAMEWORK GOVERNING CYBERBULLYING AMONG CHILDREN IN MALAYSIA. *IIUM Law Journal*, *30*(S1), 207–228. https://doi.org/10.31436/iiumlj.v30iS1.704

Obaida, M. H., Elkaffas, S. M., & Guirguis, S. K. (2024). Deep Learning Algorithms for Cyber-Bulling Detection in Social Media Platforms. *IEEE Access*, *12*, 76901–76908. https://doi.org/10.1109/ACCESS.2024.3406595

Ooi, P. B., Ahrumugam, P., Teh, P. L., & Chan, N. N. (2022). Factors Associated with Adolescent Cyberbullying Perpetration and Victimization in Malaysia. *International Academic Symposium of Social Science 2022*, 109. https://doi.org/10.3390/proceedings2022082109

Rana, J. S., & Shukla, S. (2024). IMPACT OF CYBERBULLYING ON SOCIAL MEDIA PRIVACY AND CYBERSECURITY: A CRITICAL APPRAISAL. *ShodhKosh: Journal of Visual and Performing Arts*, *5*(6), 1617–1621. https://doi.org/10.29121/shodhkosh.v5.i6.2024.2482

Ray, G., McDermott, C. D., & Nicho, M. (2024). Cyberbullying on Social Media: Definitions, Prevalence, and Impact Challenges. *Journal of Cybersecurity*, *10*(1). https://doi.org/10.1093/cybsec/tyae026

Redha Mutar, J. (2024). Cyberbullying Messages Detection Using Machine Learning and Deep Learning. *International Journal of Advances in Scientific Research and Engineering*, *10*(03), 19–29. https://doi.org/10.31695/IJASRE.2024.3.3

RUSLI, R., MAHMUD, M. M., RAMLI, N. F. M., TAJIZAN, F. N., & ZAKARIA, S. F. (2024). ADDRESSING CYBERBULLYING: PERSPECTIVES ON UNIVERSITY ACTIONS FROM MALAYSIAN UNDERGRADUATES. *Quantum Journal of Social Sciences and Humanities*, *5*(5), 362–372. https://doi.org/10.55197/qjssh.v5i5.522

Shah, V., Sinha, A., Navalkar, N., Gupta, S., Gonsalves, P., & Malik, A. (2023). ML and Natural Language Processing : Cyberbullying Detection System for Safer and Culturally Adaptive Digital Communities. *Journal of Smart Internet of Things*, *2023*(2), 193–205. https://doi.org/10.2478/jsiot-2023-0020

Singh, S., & Othman, S. H. (2025). An Effective Cyberbullying Detection Model for the Malay Language Using Transformer Model in Social Media Platform X. *International Journal of Innovative Computing*, *15*(1), 63–71. https://doi.org/10.11113/ijic.v15n1.520