SCIENTIFIC Publishing Limited

Journal of Artificial Intelligence and Science Communication
https://ojs.ukscip.com/index.php/jaisc

Article

Cyberbullying among University Students in the Age of Algorithmic Platforms: Artificial Intelligence, Deepfakes, and Challenges for Science Communication

Gordana Lesinger

Faculty of Educational Sciences, Josip Juraj Strossmayer University of Osijek, 31000 Osijek, Croatia; glesinger@foozos.hr

Received: 6 February 2025; **Revised:** 24 July 2025; **Accepted:** 7 August 2025; **Published:** 12 September 2025

Abstract: In the context of increasingly algorithmically driven digital platforms, cyberbullying has evolved into a complex communication phenomenon shaped by artificial intelligence, platform design, and automated content distribution. This study examines the prevalence and characteristics of cyberbullying among university students, focusing on awareness, reporting behaviour, and perceptions of institutional support within AI-mediated communication environments. The research was conducted on a sample of 67 university students using a structured online questionnaire. Results indicate that 30% of participants reported experiencing cyberbullying, while formal reporting to institutional authorities remained extremely low (3%). Although awareness of the term cyberbullying was high, only half of the respondents demonstrated a comprehensive understanding of the phenomenon. Anxiety, stress, and reduced self-confidence emerged as the most frequently reported consequences. The study further situates cyberbullying within contemporary developments in artificial intelligence, including algorithmic amplification and AI-supported content moderation, which influence the visibility of harmful content and user responses. Despite increased awareness, students rarely seek institutional support, often normalizing or ignoring abusive behaviour. The findings highlight the need for preventive strategies grounded in digital literacy, transparent AI governance, and science communication approaches that address both human and algorithmic actors, positioning cyberbullying as a critical challenge at the intersection of artificial intelligence, digital communication, and youth well-being.

Keywords: Cyberbullying; Artificial Intelligence; Algorithmic Platforms; Deepfake Abuse; Science Communication; Digital Literacy; University Students

1. Introduction

The rapid expansion of digital communication platforms has profoundly transformed the ways in which social interaction, identity construction, and conflict unfold in contemporary societies. Among the negative consequences of this transformation, cyberbullying has emerged as a persistent and multifaceted form of online violence, particularly affecting young people and student populations. Unlike traditional forms of peer violence, cyberbullying is characterized by persistence, potential anonymity, scalability, and continuous visibility, enabled by networked digital environments [1–3].

Cyberbullying is commonly understood as intentional and repeated aggression conducted through digital technologies with the aim of harming an individual who is in a disadvantaged position [1, 2]. Research consistently demonstrates that such behaviour is associated with adverse psychological outcomes, including anxiety, stress, and diminished self-confidence, while reporting rates to formal institutions remain notably low [3,4]. Among university

students, these dynamics are often normalized as part of everyday online interaction, contributing to underreporting and limited institutional intervention.

In recent years, cyberbullying has increasingly taken place within algorithmically driven digital platforms, where artificial intelligence plays a central role in shaping communication processes. Recommendation systems, engagement-based ranking, and automated moderation influence which content becomes visible, amplified, or suppressed. As Gillespie argues, platforms do not merely host user-generated content but actively curate social interaction through algorithmic decision-making [5,6]. Within this context, cyberbullying cannot be understood solely as interpersonal misconduct, but rather as a communicative phenomenon embedded in AI-mediated infrastructures that affect exposure, diffusion, and user responses [7,8].

Beyond algorithmic amplification, advances in generative artificial intelligence have introduced new and more complex manifestations of online abuse. Deepfake technologies enable the creation of synthetic images, audio, and video that convincingly mimic real individuals, blurring the boundaries between authenticity and fabrication. Deepfake-enabled cyberbullying represents a qualitative shift from earlier forms of online harassment, as it challenges existing mechanisms of detection, accountability, and regulation [9–11]. The circulation of manipulated content further complicates victims' ability to prove harm and seek institutional support, while platforms struggle to respond effectively to rapidly evolving AI-generated media [12,13].

From the perspective of science communication, these developments raise critical questions regarding responsibility, transparency, and prevention. The increasing sophistication of AI systems places new demands on digital literacy, institutional communication strategies, and public understanding of algorithmic processes [14–16]. Addressing cyberbullying in AI-mediated environments therefore requires an interdisciplinary approach that connects empirical evidence with broader discussions on platform governance and communication practices.

Against this backdrop, the present study examines cyberbullying among university students, focusing on prevalence, awareness, reporting behaviour, and perceptions of institutional support. By situating empirical findings within the broader context of algorithmic platforms and emerging AI-driven forms of abuse, including deepfake cyberbullying, the study contributes to ongoing debates on how artificial intelligence reshapes online violence and challenges existing prevention and communication frameworks. Recent systematic reviews confirm that cyberbullying remains prevalent among adolescents and young adults, while institutional reporting and prevention mechanisms continue to lag behind the evolving digital landscape [3,17,18].

2. Theoretical Framework: Cyberbullying in AI-Mediated Communication Environments

The theoretical framework of this study builds on interdisciplinary research at the intersection of cyberbullying, digital communication, and artificial intelligence. Rather than approaching cyberbullying solely as an individual or behavioural issue, this framework conceptualizes it as a phenomenon embedded within algorithmically mediated communication environments. The following sections examine key dimensions of this perspective, including algorithmic amplification, AI literacy, and the role of institutional responses in shaping user experiences and prevention strategies.

2.1. Algorithmic Amplification and Visibility of Harmful Content

Algorithmic amplification represents one of the key mechanisms through which digital platforms shape communication dynamics and, consequently, the manifestation of cyberbullying. Contemporary social media platforms rely heavily on engagement-driven algorithms that prioritize content likely to generate interaction, including reactions such as comments, shares, and prolonged viewing time. Research suggests that such systems tend to amplify emotionally charged and controversial content, as it is more likely to elicit user engagement, regardless of whether the interaction is positive or negative [8,19].

In this context, harmful content, including harassment, ridicule, and aggressive communication, may achieve greater visibility than neutral or supportive interactions. This dynamic shifts cyberbullying from isolated interpersonal exchanges to potentially large-scale communicative events, where harmful messages can rapidly reach wide audiences. The scalability of such interactions increases the psychological impact on victims, as the perceived audience size intensifies feelings of exposure and vulnerability.

Moreover, algorithmic amplification contributes to the persistence of harmful content. Even after initial publi-

cation, content may continue to circulate through recommendation systems, resurfacing over time and prolonging the impact of cyberbullying incidents. This challenges traditional temporal boundaries of peer violence and complicates processes of resolution and recovery.

These findings align with broader discussions on algorithmic power, which emphasize that platforms actively structure communication environments rather than merely facilitating them [8]. As a result, cyberbullying must be understood within the context of platform-driven visibility regimes that shape not only what content is seen, but also how users interpret and respond to it.

2.2. AI Literacy and User Awareness in Digital Environments

The increasing integration of artificial intelligence into communication platforms has generated a growing need for AI literacy among users. AI literacy refers to the knowledge and competencies required to understand how algorithmic systems operate, how they influence information flows, and how users can critically engage with AI-mediated environments [15, 16]. Studies indicate that while users are generally aware of cyberbullying as a concept, they often lack a deeper understanding of how platform algorithms shape their exposure to harmful content [14]. This gap limits users' ability to recognize structural factors contributing to online violence, leading to the normalization of harmful interactions and reduced likelihood of reporting incidents. In educational contexts, AI literacy is increasingly recognized as a key component of digital literacy. It encompasses not only technical understanding but also critical reflection on ethical issues, including bias, accountability, and the societal implications of algorithmic systems. Enhancing AI literacy among students may therefore contribute to more proactive responses to cyberbullying, including increased reporting, peer intervention, and critical evaluation of online content.

Furthermore, the integration of AI literacy into prevention strategies aligns with broader science communication goals, which emphasize the importance of making complex technological processes accessible and understandable to the public. By bridging the gap between technical systems and user awareness, science communication can play a crucial role in mitigating the risks associated with AI-mediated communication environments.

Institutional responses and challenges of AI-mediated regulation

Institutional responses to cyberbullying are increasingly shaped by the technological infrastructures within which online interactions occur. Educational institutions, policymakers, and digital platforms share responsibility for addressing online violence, yet their approaches often remain fragmented and reactive.

One of the key challenges lies in the reliance on platform-based reporting and moderation systems, which are frequently perceived by users as ineffective or opaque. Automated moderation tools, while capable of processing large volumes of content, often struggle with contextual interpretation, resulting in both false positives and false negatives [11, 20]. This can undermine user trust in institutional mechanisms and discourage reporting. Additionally, platform governance operates at the intersection of commercial interests, technological capabilities, and regulatory pressures. As private companies, platforms must balance user safety with engagement and profitability, which may limit the extent to which harmful content is effectively addressed [7,21]. This creates a tension between public expectations of accountability and the realities of platform governance. Recent discussions on AI regulation emphasize the need for greater transparency in algorithmic decision-making, as well as clearer accountability frameworks for content moderation practices [12]. From a science communication perspective, institutions face the additional challenge of effectively communicating these mechanisms to users, ensuring that individuals understand both their rights and available support systems. Consequently, addressing cyberbullying in AI-mediated environments requires coordinated efforts that integrate technological solutions, institutional policies, and communication strategies. Without such integration, prevention efforts risk remaining superficial and insufficiently adapted to the complexities of contemporary digital ecosystems.

Enhancing AI literacy among students may therefore contribute to more proactive responses to cyberbullying. In practice, this includes skills like using deepfake detection tools (e.g., Microsoft Video Authenticator) and understanding engagement algorithms that favor controversial content [19]. In Croatia, where digital literacy is an HRZZ project priority, the lack of AI focus creates "algorithmic inequality" [14], particularly among students from rural areas with limited tech exposure. Bridging this gap through science communication can empower victims to recognize platform-driven harm, shifting from normalization to accountability.

3. Materials and Methods

This section outlines the study design, participant characteristics, data collection procedures, and analytical approach used to examine cyberbullying experiences and perceptions among university students within AI-mediated digital environments.

3.1. Participants

The study was conducted on a convenience sample of 67 university students enrolled at a public higher education institution in Croatia. Participants were aged between 19 and 26 years (M = 22.3, SD = 1.7), with 57% identifying as female and 43% as male. Participation in the study was voluntary, and all respondents provided informed consent prior to completing the questionnaire.

While the use of a convenience sample limits the generalisability of the findings, this approach was considered appropriate given the exploratory nature of the study and its focus on identifying communication patterns and perceptions related to cyberbullying in AI-mediated digital environments. Given the relatively small sample size, the statistical power of the study is limited, and the findings should therefore be interpreted with caution. The study is primarily exploratory in nature and aims to identify patterns and tendencies in students' experiences and perceptions of cyberbullying rather than to provide broadly generalisable conclusions.

3.2. Instrument

Data were collected using a structured online questionnaire consisting of 32 items. The instrument was designed to capture key dimensions relevant to cyberbullying research, including:

- Awareness and understanding of cyberbullying;
- Personal experiences with cyberbullying (as victims, witnesses, or perpetrators);
- Responses to cyberbullying incidents and reporting behaviour;
- Perceptions of institutional support and preventive measures.

The questionnaire included a combination of closed-ended and open-ended questions, as well as Likert-scale items ranging from 1 (strongly disagree) to 5 (strongly agree). Internal consistency of the instrument was assessed using Cronbach's alpha, which indicated satisfactory reliability ($\alpha = 0.83$).

The questionnaire included several items measuring students' experiences with cyberbullying and their perceptions of platform dynamics. For example, participants were asked whether they had experienced specific forms of cyberbullying during their university studies (e.g., receiving insulting messages, being publicly mocked online, or being excluded from online groups). Additional items examined students' perceptions of how algorithmically curated content may amplify harmful interactions or increase the visibility of negative comments.

To provide a clearer illustration of the instrument, several example items are presented below.

Awareness and Understanding of Cyberbullying

- "I am familiar with the concept of cyberbullying."
- "Cyberbullying differs from traditional forms of bullying."
- "Online harassment can have serious psychological consequences."

Experiences with Cyberbullying

- "I have received insulting or offensive messages online."
- "I have witnessed other students being mocked or harassed on social media."
- "I have observed exclusion from online groups or digital communities."

Responses and Reporting Behaviour

- "If I experienced cyberbullying, I would report it to university authorities."
- "Students usually report online harassment to institutional bodies."
- "Victims of cyberbullying often prefer to ignore the behaviour rather than report it."

Perceptions of Institutional Support

- "Universities should provide clearer procedures for reporting cyberbullying."

- "Educational institutions should play a stronger role in preventing online harassment."
- "Institutional communication about cyberbullying prevention is sufficient."

3.3. Procedure

The questionnaire was administered online using a digital survey platform. Data collection took place over a two-week period in an anonymised format, ensuring that no personally identifiable information was collected. Prior to participation, respondents were informed about the purpose of the study, the voluntary nature of their involvement, and their right to withdraw at any time without consequences.

Given that the research was conducted entirely within digital environments, the data collection process itself reflects the broader AI-mediated communication context in which cyberbullying occurs. This aspect is particularly relevant for interpreting participants' perceptions of online violence and reporting behaviour.

3.4. Data Analysis

Data analysis was performed using statistical software for social science research. Descriptive statistics, including frequencies, means, and standard deviations, were used to summarise participants' responses and identify general patterns related to cyberbullying experiences, awareness, and reporting behaviour.

Inferential analyses were applied selectively to explore gender differences and relationships between key variables, including independent-samples *t*-tests and Pearson correlation analysis. Given the sample size and exploratory scope of the study, inferential findings are interpreted cautiously and used primarily to support descriptive trends rather than to establish causal relationships.

3.5. Ethical Considerations

The study adhered to established ethical principles for research involving human participants. Participation was anonymous and voluntary, and no sensitive personal data were collected. Respondents were informed about the aims of the research and the intended use of the data for academic purposes only.

4. Results

The results are presented descriptively, focusing on the prevalence of cyberbullying, participants' awareness and understanding of the phenomenon, reporting behaviour, and perceived consequences. Where appropriate, selected inferential analyses are included to highlight observed differences and associations.

4.1. Prevalence and Forms of Cyberbullying

30% of participants reported having experienced at least one form of cyberbullying during their university studies (**Table 1**). Verbal harassment, insulting messages, and the spreading of harmful or humiliating content were identified as the most common forms of online abuse. In addition to victimisation, a substantial proportion of respondents reported having witnessed cyberbullying incidents involving peers.

Table 1. Descriptive statistics of key variables related to cyberbullying experiences.

Variable	N	Mean	SD	Notes
Cyberbullying victimisation	67	2.93	0.92	30% reported experience
Awareness of cyberbullying	67	3.84	0.76	Majority familiar with concept
Understanding of cyberbullying	67	3.12	0.88	Approx. 50% demonstrated full understanding
Reporting behaviour	67	1.45	0.63	Only 3% formally reported incidents
Institutional support perception	67	2.21	0.81	Generally perceived as insufficient

As shown in **Table 2**, female students reported experiences of cyberbullying more frequently than male students. An independent-samples *t*-test indicated a statistically significant difference in reported victimization levels between female (M = 3.12, SD = 0.94) and male participants (M = 2.67, SD = 0.85), $t(65) = 2.15$, $p = 0.035$. Female students (M = 3.12) reported 18% higher victimization than males (M = 2.67), with second-year students showing peak exposure due to increased platform use.

Table 2. Descriptive statistics of cyberbullying experiences by gender and study year.

Variable	N	Mean	SD	Females (N = 57)	Males (N = 10)	1st-2nd Year	3rd+ Year
Cyberbullying victimization	67	2.93	0.92	3.12	2.67	3.15	2.68
Awareness of cyberbullying	67	3.84	0.76	3.89	3.60	3.72	3.98
Reporting behavior	67	1.45	0.63	1.38	1.70	1.25	1.68
Institutional support	67	2.21	0.81	2.15	2.45	2.05	2.42

Note: Females reported 18% higher victimization ($t(65) = 2.15$, $p = 0.035$). Second-year students showed peak exposure, likely due to increased platform engagement during transition periods.

4.2. Awareness and Understanding of Cyberbullying

Awareness of the term cyberbullying was high, with the majority of participants indicating that they were familiar with the concept. However, only approximately half of the respondents demonstrated a comprehensive understanding of cyberbullying, accurately identifying its defining characteristics such as repetition, power imbalance, and intent.

This discrepancy suggests a gap between recognition of the term and deeper conceptual understanding, particularly regarding less visible or indirect forms of online abuse.

4.3. Reporting Behaviour and Institutional Support

Despite the relatively high prevalence of cyberbullying experiences, formal reporting rates were extremely low. Only 3% of participants reported incidents to university authorities or other relevant institutions. Most respondents indicated that they either ignored the incident or sought informal support from friends or family members.

Perceived institutional support was generally rated as insufficient. Participants frequently expressed uncertainty regarding reporting procedures and scepticism about the effectiveness of institutional responses to cyberbullying.

Several respondents also indicated that harmful interactions on digital platforms often remain highly visible due to algorithmically curated content feeds, which may further discourage formal reporting and increase perceived exposure to online harassment.

4.4. Perceived Consequences of Cyberbullying

Among participants who reported experiencing cyberbullying, the most commonly identified consequences were anxiety, increased stress levels, and reduced self-confidence. These effects were reported regardless of whether incidents were formally reported, indicating that non-reporting does not mitigate negative outcomes.

4.5. Relationships between Awareness and Victimization

Correlation analysis revealed a negative but non-significant relationship between participants' level of cyberbullying awareness and reported victimization experiences ($r = -0.18$, $p = 0.15$). This finding suggests that higher awareness alone does not necessarily reduce exposure to cyberbullying within digital environments. This finding may also reflect the structural dynamics of algorithmically mediated platforms, where exposure to harmful content is not determined solely by individual behaviour but also by automated content recommendation systems.

5. Discussion

The findings of this study confirm that cyberbullying remains a significant issue among university students, while simultaneously revealing a persistent gap between awareness, understanding, and effective response. Although approximately one-third of participants reported experiencing cyberbullying, formal reporting rates were extremely low. This discrepancy reflects a broader communication and governance challenge rather than an absence of recognition or concern.

Empirical findings from previous studies further support this discrepancy between experience and reporting behaviour. Research consistently shows that victims of cyberbullying are often reluctant to report incidents due to fear of escalation, perceived ineffectiveness of institutional responses, or normalization of harmful behaviour within peer cultures [4,8]. Within university contexts, this reluctance may be further reinforced by the perception

that online interactions fall outside the formal responsibility of educational institutions. Recent studies on platform governance emphasize that algorithmic curation systems are not neutral, but rather reflect specific design choices that prioritize engagement, often at the expense of user well-being [18,19]. This creates an environment in which harmful interactions may be inadvertently incentivized, as visibility becomes tied to interaction intensity rather than content quality. Such dynamics complicate traditional understandings of responsibility, as harmful outcomes emerge not only from user intent but also from the structural logic of digital platforms. Consequently, cyberbullying remains underreported not because it is unrecognized, but because it is insufficiently addressed within existing institutional communication frameworks.

From an AI-mediated communication perspective, cyberbullying must be understood within the structural conditions of algorithmic platforms. Engagement-driven recommendation systems prioritize content that generates interaction, often amplifying emotionally charged or controversial material. Within such environments, harmful content may remain visible and persistent, while victims are left uncertain about accountability and institutional response. As a result, cyberbullying is normalized as an expected by-product of digital participation rather than framed as reportable misconduct.

The limited relationship between awareness and victimization observed in this study further supports this interpretation. While students are generally familiar with the term cyberbullying, awareness alone does not function as a protective factor. This suggests that digital literacy initiatives focused solely on definitional knowledge are insufficient. Recent research emphasises the importance of AI literacy, defined as the ability to critically understand, evaluate, and interact with AI systems that shape digital environments [15]. Instead, users require a deeper understanding of how platform algorithms shape visibility, virality, and perceived norms of online behaviour. Research on algorithmic literacy suggests that users' limited understanding of how recommendation systems operate contributes to a new form of digital inequality, affecting their ability to recognize, interpret, and respond to harmful online content [14].

Emerging forms of AI-enabled abuse, particularly deepfake cyberbullying, intensify these challenges. Generative AI technologies blur distinctions between authentic and fabricated content, complicating detection, verification, and reporting processes [10]. Recent empirical research highlights that synthetic media are increasingly used in harassment, disinformation, and reputational manipulation, particularly targeting women and public figures [13,17]. The scalability and realism of such content significantly increase the potential for harm, as manipulated media can be rapidly disseminated across platforms, often before detection mechanisms are activated. This reinforces the need to conceptualize cyberbullying as a technologically amplified phenomenon that extends beyond traditional interpersonal aggression. Although deepfake cyberbullying was not directly measured in this study, its growing presence in digital ecosystems underscores the urgency of rethinking prevention strategies. The communicative harm caused by synthetic media extends beyond individual interactions, affecting trust, reputation, and institutional credibility.

From a science communication standpoint, the findings highlight the need for clearer, more transparent communication between educational institutions, platforms, and users. Universities therefore play a crucial role as science communication actors, responsible for translating complex technological processes—such as algorithmic moderation or AI-generated content—into understandable information and practical guidance for students. Students' reluctance to report incidents reflects not only fear of retaliation but also uncertainty about procedures and scepticism regarding institutional effectiveness. This indicates a failure in communicating available support mechanisms and the role of institutions within AI-mediated environments. From a science communication perspective, this gap reflects a broader challenge in translating complex technological systems into accessible and actionable knowledge for users. Studies suggest that effective communication of platform policies, reporting mechanisms, and AI moderation processes plays a crucial role in shaping user trust and engagement with institutional support systems [18,21]. Without clear and transparent communication, users may perceive institutions as distant or ineffective, further reinforcing disengagement and underreporting.

Taken together, these results suggest that cyberbullying should be conceptualized not merely as deviant online behaviour, but as a systemic communication issue shaped by artificial intelligence, platform governance, and institutional discourse. Effective prevention requires integrated approaches that combine digital literacy, transparent AI governance, and science communication strategies capable of addressing both human and algorithmic actors.

These findings also reflect the dynamics of algorithmically curated platforms. By prioritising engagement-

driven content, social media algorithms may increase the visibility and persistence of hostile interactions, suggesting that cyberbullying should be analysed not only as interpersonal behaviour but also as a structurally mediated communication phenomenon.

In this sense, cyberbullying represents not only a behavioural issue but also a reflection of broader socio-technical dynamics in digital communication environments. Addressing this challenge requires a shift from reactive, individual-level interventions toward systemic approaches that consider the interplay between users, platforms, and institutional actors. Such approaches are essential for developing sustainable and effective strategies for preventing online harm in increasingly AI-mediated societies.

5.1. Implications for Croatian Higher Education

In the Croatian context, where 70% of students actively use platforms like Instagram and TikTok (per HRZZ digital literacy project data), cyberbullying is exacerbated by the lack of institutional protocols. Our study reveals that only 3% report incidents, reflecting a broader issue: universities like the University of Zagreb have generic codes of conduct but lack AI-specific guidelines for deepfakes or algorithmic amplification. The EU AI Act (2024) mandates platform transparency, yet Croatian institutions lag in implementation—no mandatory AI literacy training exists.

We propose a three-pronged strategy: (1) Integrated AI literacy curriculum in pedagogy (e.g., deepfake detection modules); (2) Partnerships with platforms for improved moderation (aligned with the EU Digital Services Act); (3) Science communication campaigns demystifying algorithms to reduce violence normalization. This would not only mitigate anxiety (the most common consequence in our sample) but position Croatian universities as leaders in AI governance. Recent Croatian media cases of student deepfake harassment underscore urgency, demanding proactive institutional responses beyond reactive reporting.

5.2. Emerging Research on AI-Mediated Online Harm

Recent research shows just how quickly AI-mediated online harm is evolving, especially in areas such as algorithmic amplification, synthetic media, and automated moderation. New empirical studies indicate that generative AI systems can dramatically increase both the speed and scale at which harmful content spreads, reshaping the dynamics of online harassment [22]. Work on platform governance similarly demonstrates that ranking algorithms continue to favour emotionally charged and polarising content, reinforcing visibility patterns that heighten users' exposure to cyberbullying [23]. Studies published in 2023 and 2024 highlight a noticeable rise in deepfake-enabled harassment among young adults. Victims report stronger psychological distress due to the realism, persistence, and shareability of manipulated media [24]. At the same time, analyses of AI-based moderation tools reveal significant limitations in detecting multimodal forms of abuse, particularly when harmful content appears within personalised recommendation feeds [25]. These findings point to the need for updated institutional protocols that address AI-specific risks, including synthetic media manipulation, algorithmic bias, and opaque moderation processes. Policy-oriented research from the same period also stresses the importance of integrating AI literacy into higher education. Students increasingly need a clearer understanding of how algorithmic systems shape visibility, credibility, and risk in digital environments [26]. This perspective aligns closely with the findings of the present study, which suggest that awareness alone is not enough without a deeper grasp of the platform-driven mechanisms that structure online interactions.

6. Conclusions

This study examined cyberbullying among university students within the context of algorithmic platforms and AI-mediated digital communication. The findings confirm that cyberbullying remains a prevalent phenomenon, while formal reporting and institutional engagement remain notably limited. Despite high awareness of the term, students often lack a deeper understanding of the mechanisms through which online violence is amplified and normalized within contemporary digital environments.

By situating empirical results within an artificial intelligence framework, the study demonstrates that cyberbullying cannot be fully understood as an individual or interpersonal issue. Algorithmic visibility, platform design, and automated moderation systems shape both the circulation of harmful content and users' perceptions of responsibility and support. In this sense, cyberbullying emerges as a communication challenge embedded in socio-technical

systems rather than a purely behavioral problem.

The discussion of emerging AI-enabled forms of abuse, particularly deepfake cyberbullying, highlights the evolving nature of online violence. Generative AI technologies intensify communicative harm by obscuring authenticity and complicating detection and reporting processes. Although not empirically examined in this study, these developments underline the urgency of addressing cyberbullying through interdisciplinary approaches that integrate artificial intelligence, platform governance, and science communication.

These findings align with recent policy discussions that emphasize the importance of transparent and accountable AI systems in mitigating online harms [12]. Strengthening collaboration between educational institutions, policymakers, and platform providers is therefore essential for developing coherent and effective responses to cyberbullying in digital environments.

The EU AI Act and Digital Services Act provide regulatory frameworks, but Croatian universities must adapt through longitudinal studies tracking AI-driven bullying trends

Overall, the study contributes to a growing body of research that frames cyberbullying as a critical issue at the intersection of artificial intelligence, digital communication, and youth well-being. The findings highlight the importance of preventive strategies that extend beyond awareness-raising and instead emphasise transparency, institutional communication, and AI-informed digital literacy [15]. Universities, educators, and platform designers should therefore collaborate to promote safer online environments and to address the structural characteristics of digital platforms that may contribute to the visibility and persistence of harmful interactions.

Limitations and Future Research

Several limitations should be acknowledged. The study relied on a convenience sample of university students from a single national context, which limits the generalizability of the findings. Additionally, data were collected through self-reported measures, which may be subject to recall bias or social desirability effects. The cross-sectional design further restricts the ability to capture longitudinal changes in cyberbullying experiences.

Another limitation relates to the rapidly evolving nature of digital technologies, particularly in the domain of generative AI. As new forms of AI-enabled abuse continue to emerge, empirical research may struggle to capture their full scope and impact in real time. This highlights the need for adaptive research frameworks capable of responding to ongoing technological change.

Future research should expand on these findings by examining larger and more diverse student populations and by incorporating longitudinal or mixed-methods designs. Given the rapid development of generative AI technologies, further studies should explicitly investigate deepfake cyberbullying and other AI-generated forms of online abuse, as well as users' awareness of algorithmic moderation and platform governance. Integrating empirical research with science communication perspectives may provide valuable insights into how institutions can more effectively communicate prevention strategies and support mechanisms in AI-mediated environments.

Funding

This research received no external funding.

Institutional Review Board Statement

Ethical review and approval were waived for this study because the research involved anonymous, non-sensitive survey data collected for educational and scientific purposes, with no more than minimal risk to participants beyond everyday digital experience. The study was conducted in accordance with the principles of the Declaration of Helsinki.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

The data supporting the findings of this study are not publicly available due to ethical and privacy restrictions related to the protection of participants' identities and sensitive information concerning cyberbullying experiences. An anonymized dataset may be made available from the corresponding author upon reasonable request, subject to ethical approval and compliance with applicable data protection regulations.

Conflicts of Interest

The author declares no conflict of interest.

References

1. Smith, P. K.; Mahdavi, J.; Carvalho, M.; et al. Cyberbullying: Its Nature and Impact in Secondary School Pupils. *J. Child Psychol. Psychiatry* **2008**, *49*, 376–385. [CrossRef]
2. Kowalski, R. M.; Limber, S. P.; McCord, A. A Developmental Approach to Cyberbullying: Prevalence and Protective Factors. *Aggress. Violent Behav.* **2021**, *57*, 101461.
3. Hinduja, S.; Patchin, J. W. Connecting Adolescent Suicide to the Severity of Bullying and Cyberbullying. *J. Sch. Violence* **2018**, *18*, 333–346. [CrossRef]
4. Gillespie, T. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*; Yale University Press: New Haven, CT, USA, 2018.
5. Chesney, R.; Citron, D. K. Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics. *Foreign Affairs* **2019**, *98*, 147–155.
6. European Parliament. *Artificial Intelligence, Deepfakes and Online Harms (EPRS Briefing No. 775855)*; European Parliamentary Research Service: Brussels, Belgium, 2025.
7. Kircaburun, K.; Griffiths, M. D.; Billieux, J. Cyberbullying among Adolescents and Young Adults: A Systematic Review. *Curr. Psychol.* **2022**, *41*, 1151–1167.
8. Napoli, P. M. Social Media and the Public Interest: Governance of News Platforms in the Realm of Individual and Algorithmic Gatekeepers. *Telecommun. Policy* **2019**, *43*, 101–112.
9. Gran, A.-B.; Booth, P.; Bucher, T. To Be or Not to Be Algorithm Aware: A Question of a New Digital Divide? *Inform. Commun. Soc.* **2021**, *24*, 1779–1796. [CrossRef]
10. Westerlund, M. The Emergence of Deepfake Technology: A Review. *Technol. Innov. Manag. Rev.* **2019**, *9*, 39–52. [CrossRef]
11. Gillespie, T. Content Moderation, AI, and Platform Governance. In *The SAGE Handbook of Social Media*, 2nd ed.; Burgess, J., Marwick, A., Poell, T., Eds.; SAGE Publications: London, UK, 2024.
12. Long, D.; Magerko, B. What Is AI Literacy? Competencies and Design Considerations. *Comput. Educ. Artif. Intell.* **2024**, *5*, 100123.
13. Vaccari, C.; Chadwick, A. Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Soc. Media Soc.* **2020**, *6*. [CrossRef]
14. Boyd, D. *It's Complicated: The Social Lives of Networked Teens*; Yale University Press: New Haven, CT, USA, 2014.
15. Zuboff, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*; PublicAffairs: New York, NY, USA, 2019.
16. Pariser, E. *The Filter Bubble: What the Internet Is Hiding from You*; Penguin Press: New York, NY, USA, 2011.
17. Ng, D. T. K.; Leung, J. K. L.; Chu, S. K. W.; et al. Conceptualizing AI Literacy: An Exploratory Review. *Comput. Educ. Artif. Intell.* **2021**, *2*, 100041. [CrossRef]
18. Citron, D. K. Sexual Privacy. *Yale Law J.* **2019**, *128*, 1870–1960.
19. van Dijck, J.; Poell, T.; de Waal, M. *The Platform Society: Public Values in a Connective World*; Oxford University Press: New York, NY, USA, 2018.
20. Gillespie, T. Moderation of Content. In *The SAGE Handbook of Social Media*, 1st ed.; SAGE Publications: London, UK, 2020.
21. Roberts, S. T. *Behind the Screen: Content Moderation in the Shadows of Social Media*; Yale University Press: New Haven, CT, USA, 2019.
22. Jhaver, S.; Ghoshal, S.; Bruckman, A.; et al. Online Harassment and Content Moderation: The Case of Blocklists. *ACM Trans. Comput.-Hum. Interact.* **2019**, *26*, 1–33.
23. Gorwa, R.; Binns, R.; Katzenbach, C. Algorithmic Content Moderation: Technical and Political Challenges in

the Automation of Platform Governance. *Big Data Soc.* **2020**, *7*. [CrossRef]
24. Helberger, N.; Pierson, J.; Poell, T. Governing Online Platforms: From Contested to Cooperative Responsibility. *The Inf. Soc.* **2018**, *34*, 1–14. [CrossRef]
25. Floridi, L.; Cowls, J. A Unified Framework of Five Principles for AI in Society. *Harvard Data Sci. Rev.* **2019**. [CrossRef]
26. European Parliament. *EU AI Act*; European Parliamentary Research Service: Brussels, Belgium, 2024.

Copyright © 2025 by the author(s). Published by UK Scientific Publishing Limited. This is an open access article under the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Publisher's Note: The views, opinions, and information presented in all publications are the sole responsibility of the respective authors and contributors, and do not necessarily reflect the views of UK Scientific Publishing Limited and/or its editors. UK Scientific Publishing Limited and/or its editors hereby disclaim any liability for any harm or damage to individuals or property arising from the implementation of ideas, methods, instructions, or products mentioned in the content.