Roadmap for Emerging Cyberbullying Mitigation: Integrating AI-Based Solutions, Ethics, and Policy

Atif Mahmood[1], Shaik Shabana Anjum[2]*, Umm E Mariya Shah[3]*, Pavani Cherukuru[4]*, Javid Iqbal[5], Sarah Bukhari[6]
Faculty of Data Science and Information Technology, INTI International University, 71800 Nilai, Negeri Sembilan, Malaysia[1]
School of Computer Science-Faculty of Innovation and Technology, Taylor's University, Kuala Lumpur, Malaysia[2]
Department of Finance-Kulliyyah of Economics and Management Sciences,
International Islamic University, 53100 Kuala Lumpur, Malaysia[3]
Department of Information Science, Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka, India[4]
Department of Data Science and Artificial Intelligence-Faculty of Engineering and Technology (FET),
Sunway University, Kuala Lumpur, Malaysia[5]
Department of Information and Communication Technology, Bahauddin Zakariya University Multan, Pakistan[6]

***Abstract*—Cyberbullying is one of these challenges that are most found among the younger users of social media which affects the mental health. Artificial Intelligence (AI) is rapidly developing and has enormous potential to mitigate cyberbullying. Therefore, this chapter will talk about the role AI has started playing in strengthening the efforts to combat cyberbullying. Cyberbullying includes all forms of deliberate aggressive behaviour that aims to inflict social, psychological or physical pain in a digital space and AI detection technologies have a lot of potential to detect, predict and prevent cyberbullying in real time. Other critical components of the chapter are how the advances in Natural Language Processing (NLP) technologies, machine learning, images and videos, behavioural analytics make AI an emerging innovation to prevent cyberbullying and provide better services in a timely manner. There are positive trends that make it clear how Safer AI can help in improving the safety of future digital environments. More advanced NLP models will be able to identify the nuances of cyberbullying involving indirect attacks and sarcasm. The chapter will also discuss the hazards associated with AI-based solutions, such as privacy, the zero-sum game of AI morality against AI effectiveness, and the importance of explaining and assigning responsibility for every AI decision. It shows how AI is changing our approach to online safety and helps us identify cyberbullying in a variety of media, including text, video and images. This article gives an overview of roadmap for cyberbullying mitigation with the assistance of AI and ethical practices.**

Keywords*—*Cyberbullying; human computer interaction; artificial intelligence; natural language processing; machine learning; content moderation; predictive analytics; online safety; youth protection; mental health; mental illness; cybercrime

I. Understanding Cyberbullying

A. Definition and Categories of Cyberbullying

Cyberbullying is the intentional use of internet communication to harass, menace or humiliate another person. Unlike conventional bullying, cyberbullying originates in the digital space and is thereby widespread and persistent since it can cross boundaries of geography and time [1]. Anonymity of online sites can sometimes give abusers more confidence. Lack of face-to-face contact makes the problem worse because the offenders may not consider the emotional impact of their actions on the victim.

Several studies have looked into the issue in depth; others have broken down the broad term "cyberbullying" into specific types including flaming, harassment, impersonation, cyberstalking, flooding, masquerade, denigration, outing and stalking [2,3].

1) Sending offensive or threatening messages repeatedly is harassment. For example, sending hate emails or messages every day.

2) Flaming means participating in rude internet arguments. For instance, commenting offensive things in comment sections.

3) Impersonation is when you assume the identity of another person to harm their reputation. For example, setting up a fake social media account and posting inappropriate content under someone else's name.

4) Cyberstalking is when you continuously monitor or follow someone online with malicious intent to intimidate or control them from their social media activity.

5) Flooding in digital worlds is when someone sends too many unwanted messages, emails or comments to someone. Intended to scare, annoy or silence the target, this is another form of harassment.

6) Masquerade is when someone pretends to be another online to damage relationships or reputation. This can be creating fake profiles, forwarding damaging messages pretending to be someone else, or acting dishonestly to undermine the trust or reputation of the victim.

7) Denigration is the spreading of false or defamatory online information to discredit someone. This can be abusive comments, lies or gossip sent on posts, comments or messages.

8) Outing is when you publicly share private, sensitive or personal information without consent. This can affect the victim's personal or professional life or shame and distress.

9) Cyberstalking is continuous, invasive monitoring or harassment of a person online. It usually means monitoring the victim's activity, sending continuous negative messages or instilling fear through digital means.

*Corresponding authors.

B. Social, Psychological and Physical Effects of Cyberbullying in Digital Spaces

Cyberbullying has a huge impact on people's social, psychological and physical lives [4]. Often those who are socially isolated retreat from both online and offline events because of fear of being ridiculed or more harassment. This disengagement can ruin personal and professional reputation, and affect jobs and social networks and other areas. You can find many more negative psychological effects. Many times, victims suffer from low self-esteem, anxiety and depression. Studies show that cyberbullying victims show symptoms of depression twice as much as non-victims [5]. Persistent long-term stress from cyberbullying can lead to mental health problems like headaches, fatigue, sleep disturbances, high blood pressure and heart conditions. Extreme cyberbullying can increase the risk of suicidal thoughts and self-harm especially in younger population. In extreme cases cyberbullying has led to suicides [6].

These social, psychological and physical effects are more harmful to vulnerable groups, that's why we need to address and prevent cyberbullying to protect ourselves.

C. Impact of Cyberbullying on Younger Social Media Users

Cyberbullying is particularly common among younger social media users, such as teens and preteens, due to their developmental stage and reliance on social recognition [7]. Negative online interaction can have a significant impact on such individuals. Bullied children may skip school to avoid encountering their oppressors, which could result in a decline in academic performance and an increase in absenteeism [8]. Negative online behaviour including creating fake profiles to avoid being found out could be due to fear of rejection or exclusion. Teenagers' growing brains make them more susceptible to mental illness from internet addiction [9]. Public humiliation on sites like Instagram can totally crush people's self-esteem. Studies show that because of its focus on visual content and social validation, sites like Instagram and TikTok are hotspots for cyberbullying [10,11]. A 2021 survey found that almost 60% of kids said they experienced some form of online harassment [12].

D. Current Global Efforts to Combat Cyberbullying

Among the many approaches to cyberbullying are legislative actions, technological innovations, educational programs and support systems. These programs help victims and reduce online abuse.

Legislative actions are needed to combat cyberbullying globally and nationally. The US, UK and Australia have already passed laws on this. The Australian "e-Safety Commissioner" can help reduce online abuse by removing harmful internet content [13]. Meanwhile the European Union's Digital Services Act has a comprehensive plan to increase member internet security. This proposal requires online platforms to limit harmful content so they are accountable in keeping the digital space healthy [14,15].

Partially technology has helped reduce cyberbullying. On social media networks like Facebook, content moderation tools and reporting mechanisms allow users to flag objectionable content. Facebook uses artificial intelligence to aggressively search for hate speech and inflammatory language [16]. Meanwhile artificial intelligence-powered platforms like Rethink give consumers quick cues to rethink before they share potentially dangerous content [17].

The fight against cyberbullying is mostly focused on educational activities. Non-governmental organisations (NGOs) and educational institutions are implementing projects to increase digital literacy and empathy among young people. The "Delete Cyberbullying" project teaches teens about the effects of online harassment (https://www.endcyberbullying.net/). Meanwhile events like Safer Internet Day brings together educators, politicians and tech companies to discuss practical solutions and raise awareness on cyberbullying.

Support systems are needed for cyberbullying victims to manage their psychological and emotional response to online harassment. Helplines and initiatives like Childline (https://childline.org.uk/) in the UK are available for victims. Meanwhile social media platforms have introduced tools like Instagram's "Restrict" feature which allows users to block interactions with potential attackers so victims have some sense of security [16]. These combined strategies are a total solution to cyberbullying.

E. Limitations of Current Approaches

Despite all the work, current approaches to cyberbullying have big flaws that make them not work. Legislative is a main challenge. Cyberbullying is global, law enforcement projects are across borders. Tracking down anonymous offenders across different countries is a real challenge for law enforcement.

To make it worse, many social media platforms rely on users to report cyberbullying. This means delays in response to incidents [18]. Many times, victims complain about platforms not responding to objectionable content. Many times, algorithms used to detect negative behaviour can't understand context, sarcasm or cultural sensitivity. AI systems can flag innocent banter as harassment so making it harder to detect cyberbullying.

The shame of cyberbullying stops victims from seeking help. Many delay reporting because of fear of being labelled as "oversensitive" or retaliation. This stigma stops victims from getting the help and support they need [13]. Plus, there are no educational initiatives to increase awareness and stop cyberbullying. Many projects don't keep up with the fast pace of digital interactions so new online environments like BeReal don't have enough security and user services. These limitations highlight the need for continuous work to reduce cyberbullying.

F. The Importance/Potential of AI in Mitigating Cyberbullying

As cyberbullying has emerged as a significantly widespread problem, it is required to adopt the efficient detection and prevention measures. Artificial intelligence is essential in tackling cyberbullying. It provides innovative and adaptable solutions that are scalable as well. It facilitates in real-time monitoring and reviewing the content for automated detection and classification of hazardous messages. Thus, assisting in the identification and reduction of cyberbullying events.

Furthermore, AI can also help to increase awareness among digital space to combat with cyberbullying.

This paper discusses the importance of AI in cyberbullying mitigation in Section I followed AI-powered solutions for prevention of cyberbullying in Sections II and III respectively. The future recommendations and directives have been elaborately discussed in Section IV followed by concluding remarks in Section V.

II. AI-Powered Solutions for Cyberbullying Prevention

AI can be utilized to identify cyberbullying in social networks. It can offer a range of solutions to combat cyberbullying. Such solutions are comprised of real-time detection and intervention, predictive modelling and risk assessment, content moderation and filtering and victim support. Moreover, AI-based educational tools and personalized learning approaches could be incorporated to spread awareness and educate individuals to address such situations and make decisions [19].

AI solutions are made up of text analysis, multimedia content analysis and user behaviour analysis as depicted in Fig. 1. Natural language processing (NLP), machine learning and data mining are used in AI powered cyberbullying solutions. These can help build a robust cyberbullying detection system.

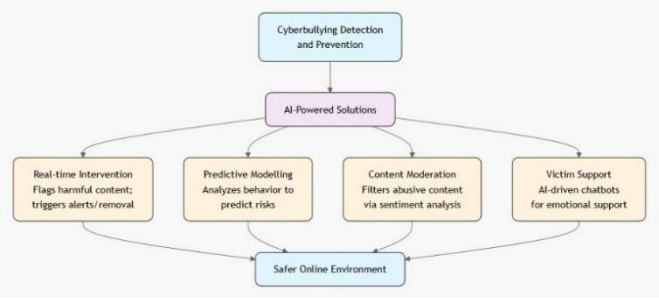

Fig. 1. AI powered solution.

A. Real-Time Intervention Mechanisms

For text analysis, linguistic insights can be extracted using NLP. It will help identify taunts, insults, threats and other harsh or abusive language in the text.

Multimedia content is audio, video and image content so cyberbullying can be detected through various methods. Images and videos can be analysed by AI algorithms involving facial expression recognition, body language analysis, object recognition and image manipulation detection [20]. Audios can be analysed for tone, pitch and volume variation and recognising sound relevant to violence or distress. Combining text, audio and visual indicators in multimodal analysis helps to understand the background and intention of the content [16].

Datamining can help to reveal the broader context of human behaviour to uncover hidden patterns related to cyberbullying. It involves examining user profiles, posting behaviour of the user, interaction pattern and social contacts to identify possible victims or cyberbullies [21].

In this way text messages, online comments and social media posts could be automatically flagged by the AI-powered system. Moreover, several interventions could be triggered to combat cyberbullying like sending alerts to the human reviewers or moderators, warning notifications to the offenders, automatic removal of the hazardous content, connecting the victims to the support channels, helplines or counselling services [22].

B. Predictive Modelling and Risk Assessment

Proactive prevention of cyberbullying relies much on the AI-driven predictive modelling and risk assessment strategies. These strategies observe the online activities and interaction pattern to analyse user behaviour. Based on the analysis, system can predict who would be more likely to engage in cyberbullying as an offender or a victim. Such proactive approach enables platforms to take necessary measures to minimise the damage thus fostering a safer online community.

Machine learning could help to develop predictive models based on predefined labels and large datasets that can classify and detect cyberbullying based on word frequency, the use of particular emojis, or the overall message sentiment [23]. Nevertheless, the scarcity of publicly available dataset, especially having multimedia content and the limitation of several studies based on only textual data poses a significant challenge in the development of cyberbullying detection systems capable of analysing images and videos [20]. Furthermore, there is no standardized definition of cyberbullying and are no specified rules or guidelines for data labelling as well. This could also lead to inefficient cyberbullying detection algorithm.

III. Advances in AI Technologies for Cyberbullying Prevention

Artificial intelligence has emerged as a powerful tool in combating the growing problem of cyberbullying [24]. Recent developments in deep learning and natural language processing have enabled the creation of sophisticated models that can effectively detect and mitigate the impact of cyberbullying incidents [25]. One key area of focus is the use of sentiment analysis and behavioural analytics to identify potential cyberbullying threats. By analysing the tone, context, and language patterns of online interactions, AI-powered systems can flag concerning content and alert authorities or community moderators, allowing for timely intervention.

Moreover, machine learning algorithms have garnered attention for their ability to moderate user-generated content, filtering out harmful or abusive material and curbing the spread of cyber-violence. Beyond detection, AI techniques are being leveraged to enhance the proactive prevention of cyberbullying. Intelligent agents and chatbots, powered by natural language generation, can engage with users at risk, offering emotional support, guidance, and strategies to build resilience.

A. Natural Language Processing (NLP)

1) Text analysis and sentiment analysis: Utilizing NLP techniques to identify bullying language patterns, tone, and context. Recognition and classification of dangerous or abusive content on websites and social media [26]. Creating sophisticated deep learning models to decipher the subtleties of

language used in online conversations, such as transformer models and recurrent neural networks [26]. Although natural language processing has been a major area of interest, scientists are investigating a variety of cutting-edge strategies in the quickly growing field of artificial intelligence for cyberbullying prevention [27].

2) Identifying cyberbullying patterns and language: Examining textual data to find signs of cyberbullying, including aggressive exchanges, derogatory language, and emotional trends [24,26]. Recent developments in deep learning architectures, like transformer models and recurrent neural networks, have shown impressive potential in precisely identifying and categorising cyberbullying content. Researchers have been able to build systems that can detect and distinguish cyberbullying from regular online interactions by using these models with big data [28, 29,56].

3) Recognizing sarcasm and indirect attacks: One of the challenges in detecting cyberbullying is finding indirect or sarcastic forms of harassment which can be more subtle and harder to identify. To detect these more advanced forms of cyber aggression AI powered systems are being developed to detect contextual information and subtle language cues.

To better understand the cyberbullying landscape some studies have looked into multifaceted approaches, combining text analysis with other sources of information like images and videos [24].

B. Machine Learning

Machine learning is being used more and more in risk identification and content moderation to prevent cyberbullying. By looking at user generated content these algorithms can identify and remove offensive or dangerous content and stop cyberviolence from spreading.

By looking at online trends and interactions behavioural analytics has also been explored as a way to predict who will experience cyberviolence. Scientists have looked at various feature sets to develop strong machine learning models for early detection of cyberbullying incidents - content driven, profile based, mobile app data etc [30].

More precise and effective cyberbullying detection systems have been made possible by machine learning. To use the power of neural networks in text data analysis and cyberbullying pattern recognition researchers have tried out various deep learning algorithms - Gated Recurrent Units, Long Short-Term Memory, Bidirectional LSTM.

1) Predictive modelling to prevent cyberbullying: Machine learning-based predictive models were also implemented to the cyberbullying prevention system. These models can help predict people at risk of being-involved in CB from past patterns and information.

Based on this group of at-risk individuals, targeted interventions, such as AI-based interventions, including personalised support, counselling, and education materials, can be provided to help them build resilience and encourage positive online behaviours [25, 30].

One promising approach to address the challenging problem of cyberbully prevention is integrating the proactive intervention techniques and AI-based predictive modelling [29].

2) Data analysis-based methods for user behaviour study: In order to obtain better cyberbullying detection, researchers have been studying data-driven approaches that exploit peer influence and per user feature set beyond the well-known detection and classification strategies. In order to offer a more accurate and personalised detection of cyberbullying, these tailored models can also take into account a user's individual online activity, social media profile and interaction specificity [31].

These AI-based systems can recognize new trends in cyber-heists and also act more quickly by incorporating more granular data about users as well as peer-to-peer dynamics. This could halt the dynamics of the situation from exacerbating and reduce harm to victims [30].

C. Image and Video Analysis

While most AI-based cyberbullying prevention related work has focused on text analysis, it is increasingly recognized that these approaches need to be extended to other modalities, like images and videos. The emerging area of computer vision and multi\-media analysis have been employed by researchers to address visual facets of cyberbullying such as circulation of embarrassing or humiliating images and videos. Deep-learning algorithms also trained on massive numbers of cyberbullying-related photos and videos can help locate and flag harmful media, leading to pre-emptive measures and better content moderation.

AI-powered systems that integrate sophisticated text analysis and visual recognition capabilities provide a more complete method for tackling the complexity of cyberbullying by covering both written and visual aspects [32].

1) Detecting harmful content in multimedia: The increased presence of user-generated multimedia content, such as videos and images, has also provided a new challenge in detecting instances of cyberbullying. To address this issue, researchers have several projects that develop AI solutions to analyse visual data to identify or detect potentially harmful or abusive content. These systems can use various computer vision approaches including object detection, face recognition or detection, and scene understanding to detect and identify images or videos that could potentially be representative of a cyberbullying related incident.

2) Detecting cyberbullying through visual cues (e.g., body language, facial expressions): In addition to identifying harmful media, there are a few AI-powered systems that are currently investigating using visual cues to detect cyberbullying behaviour. By analysing the context, body language, and emotional expressions in images and video, these AI models can be leveraged to identify patterns of behaviour that can act as indicators of cyber-aggression, even when there is no text content [27][57].

The combination of visual-based analysis with a traditional text-based approach may result in a more comprehensive picture of the cyberbullying phenomenon, which can help inform more effective detection and intervention protocols [33].

D. Behavioural Analytics

Behavioural analytics have taken root as an instrumental resource in the prevention of cyberbullying by utilizing AI to assess patterns of online interactions and behaviour. Algorithms focused on behavioural analytics consider the frequency of communication, tone of communication, and social networks, which are valuable predictors of whether individuals are defenceless or if their behaviour is a risk for cyberbullying. The combination of AI-driven behavioural analytics and personalized intervention approaches has potential to empower victims, support positive online communities and encourage citizenship online.

1) Analysing user behaviour patterns to find at-risk cyberbullying victims or perpetrators: Research has been done which aims to harness AI-powered behavioural analytics to discover early warning signs of cyberbullying such as changes in online activities, communications, and social network interactions. AI can help monitor and analyse user behaviour data to discover individuals who may be prone to cyberbullying victimhood or may display concerning behaviours which could escalate into cyber-aggression.

In even a preliminary use of behavioural analytics to detect potential for harm, early identification of these behaviour patterns could facilitate timely interventions with at-risk individuals leading to connection with appropriate supports and resources to help them build resilience and positive online engagement.

2) Tracking digital interactions to identify aggression and conflict: More than just identifying victims and offenders - AI behavioural analytics can be used to track interactions in digital spaces for signs of aggression, conflict, and rising tensions. If AI systems analyse tone, sentiment and dynamics of the digital conversation, they may show early warning signals for cyberbullying so the inventible mechanism can kick in. This proactive data-driven model to moderation and community management can completely alter how we consider - and respond to - cyberbullying and increase levels of acceptability of online spaces.

To summarize, where there is great potential for preventing cyberbullying is through radical data taking using AI-powered solutions. More specifically, there are an unlimited number of opportunities for real-time monitoring of online environments whether through imaginative text analysis methods, multimedia recognitions or through developing personalized behavioural analytics and access to the patterns of users across platforms, AI holds a lot of promise to help detect, prevent, and intervene, when being victimized online [24,25,34].

Practitioners and scientists may want to consider how they can leverage AI in developing digital wellbeing interventions that may result in much more efficient, scalable, and adaptable services. They can also bring some responsibility back to owners of digital spaces to create more inclusive and safer online communities and places where it is safe to innovate [24, 34, 35, 36].

E. Challenges and Considerations

There is tremendous promise in AI-Supported cyberbullying prevention, although researchers and practitioners have found several important challenges and considerations. Among the most important are ethical issues, privacy implications, and needing different disciplines to work together to substantiate responsible and effective uses of these technologies.

F. Privacy Considerations

The collection and analysis of user data for the detection of cyberbullying and for intervention raise serious privacy implications. AI systems need to have strong privacy and security protocols in place to protect personal information, particularly for minors who could be vulnerable to cyberbullying [37].

In order to establish user trust, and ensure compliance with privacy laws, it is important to deliberately consider the processes for collecting, storing, processing, and being transparent and accountable about how data is collected and used by these systems [24].

1) Data collection and use for AI models: When developing AI-based cyberbullying prevention and detection tools, it is often important to collect and analyse user data, such as information from online chats, online social networking communications, and patterns of digital behaviour [38].

This information has to be processed with utmost care and vigilance to protect personal privacy and avoid possible abuse or exploitation—even if it is critical for training and validating these AI systems. In order to ensure that personal information is only used for lawful and ethical purposes, developers and researchers must implement strict data management frameworks that provide clear rules on data collection, user consent, and data de-identification [39].

2) Safeguarding user privacy and online anonymity: Consideration for personal safety and online anonymity should be explored when developing AI systems to prevent cyberbullying, especially in regard to children and youth (i.e., vulnerable users). It is important chatbot and AI systems do not breach user privacy and jeopardize the anonymity many users need to feel safe, secure and empowered to navigate their online lives [34].

Developers must provide a fair balance in protecting user privacy and online safety while identifying and addressing cyberbullying incidents.

G. Ethical Considerations

The use of AI-powered solutions to combat cyberbullying raise a variety of ethics-related questions that will need to be taken seriously. One question to consider is the obligation to ensure that these advancements are developed and deployed in ways that honour the fundamental rights of individuals and cultural norms, along with requisite algorithmic calculation, and the concern of unintended consequences.

1) Algorithmic bias and fairness: Cyberbullying prevention measures must be thoroughly assessed for bias and fairness to ensure that they do not unjustly single out or disfavour specific people or groups. Discriminatory judgements that further disenfranchise already marginalised groups and perpetuate complaints can be caused by biases in algorithmic decision-making, model design, or training datasets [40].

2) Transparency and accountability in AI-enabled decision-making: AI-enabled cyberbullying control or prevention systems must be transparent and accountable to users, stakeholders, and regulators for its decisions and actions. Processes and procedures must be put in place, and detailed, to evidence, show, and or demonstrate the workings of the system, the decision processes, and how the system responds to instances of cyberbullying.

In the absence of similar transparency and accountability, these technologies may be perceived as opaque, unreliable, and perhaps subject to abuse or inappropriate misuse [41].

3) The potential for AI to be repurposed for malicious ends: The aim of the AI-powered cyberbullying prevention programs is to render online spaces safer and comparatively friendly, but there is a possibility for them to be misappropriated for illegal ends such as censorship, surveillance, or to purposely harass certain individuals or groups. To ensure that these technologies are utilized for appropriate and ethical purposes and do not infringe upon fundamental human rights or democratic freedoms, protections and governance structures must exist [42].

H. The "Zero-Sum Game" of AI Morality and Effectiveness

There might be a disjunction between the moral and ethical consequences of AI's expanding abilities to detect and respond to cyberbullying and the efficiency of systems employed to address this issue. In employing AI-enabled cyberbullying surveillance systems, there must be a cognitive divorce between AI tools preventing cyberbullying with moral and ethical soundness with high expectations of privacy, equity, and protection of human rights on one side and the use of AI tools for efficacy and convenience against cyberbullying on the other side.

However, the more successful these systems are at detecting and addressing instances of cyberbullying, the greater amount of user information may be required to gather and examine, which could cause privacy issues. This "zero-sum game" between AI efficacy and morality is a serious issue that needs to be resolved by careful, multi-stakeholder cooperation and the creation of strong governance structures.

AI-powered preventative cyberbullying methods that are socially conscious and have an impact on safer online environments will only be successful if the proper balance is struck between ethical considerations and technological efficacy [24, 25, 43, 44].

In the end, preventing cyberbullying with AI will need a multipronged strategy that emphasises both modern technology and moral user-centred design. Researchers and developers can endeavour to create solutions based on AI that are not only successful at recognising and reducing cyberbullying but also in line with fundamental liberties and societal values by tackling the important issues concerning information privacy, bias within algorithms, and transparent accountability [24, 44, 45, 46].

1) Balancing the need for accurate detection with the risk of false positives: There is always a chance of false positives, in which innocuous or normal interactions are mistakenly reported as cyberbullying, even though AI-powered cyberbullying detection systems can be very successful at spotting potentially harmful content or behaviours.

The AI-based cyberbullying solutions could additionally restrict applicable people or groups or unintentionally over-censor acceptable behaviours suppressing acceptable amounts of online expression.

In order to more effectively identify cyberbullying while avoiding false positives that may limit rights and liberties of users, developers must maximize the sensitivity and specificity of their AI outcomes.

To find the right balance and ensure that AI-based cyberbullying prevention solutions are effective and socially responsible will require, significant and extensive user testing, iterative model improvement, and close collaboration with a broad range of stakeholder groups [47].

2) Striking a balance between user privacy and the effectiveness of AI: While a simple conceptual model of a cyberbullying detection and intervention system based on AI could be developed, the "effective" part of a system likely requires access to a more extensive data set including social networking activity, online communications and perhaps even personal data.

Even if it is done to increase online safety, users might think that their private data has been inappropriately used or excessively exposed, which is why the gathering and utilization of such sensitive data impacts serious privacy concerns.

As a result, developers must carefully balance the requirement for efficient AI-powered cyberbullying detection with users' fundamental right to privacy and control over their personal information. This could entail creating privacy-preserving strategies like federated learning or differential privacy in addition to strong data governance frameworks that give users the ability to control how their data is used. The future of AI-powered cyberbullying prevention may be one of greater efficacy, social responsibility, and user trust if these major issues are resolved [25]. The summary of findings for AI-powered Cyberbullying mitigation has been tabulated in Table I.

TABLE I. SUMMARY OF FINDINGS: AI-POWERED CYBERBULLYING MITIGATION

Aspect	Key Technologies/Methods	Primary Applications & Capabilities	Identified Challenges & Considerations
AI Modalities & Solutions			
Natural Language Processing (NLP)	- Transformer Models (e.g., BERT, GPT) - Recurrent Neural Networks (RNN, LSTM) - Sentiment & Contextual Analysis	- Analysis text to detect abusive language, threats, and insults. - Identifies patterns of harassment and denigration. - Aims to understand sarcasm and indirect attacks.	- Difficulty with linguistic nuance, cultural context, and sarcasm. - High risk of false positives/negatives without deep contextual understanding.
Machine Learning (ML) / Predictive Analytics	- Supervised & Deep Learning (GRU, Bi-LSTM) - Behavioural Pattern Analysis - Peer Influence Modelling (e.g., PI-Bully)	- Classifies cyberbullying based on word frequency, emojis, and sentiment. - Predicts at-risk users (potential victims or perpetrators). - Enables proactive, targeted interventions.	- Scarcity of large, labelled, and multi-modal (text, image, video) datasets. - Lack of a standardized definition of cyberbullying for data labelling.
Image & Video Analysis	- Computer Vision - Facial Expression & Body Language Recognition - Object & Scene Recognition	- Detects harmful visual content (embarrassing images/videos). - Identifies visual cues of aggression or distress. - Multimodal analysis combines visual and textual context.	- A relatively under-explored area compared to text analysis. - Requires complex models to integrate visual and contextual data.
Behavioural Analytics	- User Interaction Analysis - Social Network Analysis	- Tracks digital interactions to identify early signs of aggression and conflict. - Analyses user profiles, posting frequency, and social connections to flag risky behaviour.	- Raises significant privacy concerns due to continuous monitoring. - Requires balancing intrusion with effectiveness.
Core Challenges			
Ethical & Privacy	- Data Governance Frameworks - Transparency Protocols	- Ensures user data is protected, especially for minors. - Aims for accountable and explainable AI decision-making.	- Privacy Risk: Collection of sensitive personal data. - "Zero-Sum Game": Tension between AI efficacy (needing more data) and user privacy/morality (protecting data).
Algorithmic Bias & Fairness	- Bias Mitigation Strategies - Fairness Audits	- Seeks to create equitable systems that do not disproportionately target or harm marginalized groups.	- Biased training data can lead to discriminatory outcomes and reinforce existing societal inequalities.
Effectiveness & Accuracy	- Model Optimization - Iterative Testing	- Aims to maximize accurate detection of true cyberbullying incidents.	- False Positives: Risk of censoring innocent banter or legitimate speech. - False Negatives: Risk of missing sophisticated or novel forms of harassment.
Future Directions			
Technical Innovations	- Advanced NLP for subtlety - Real-time monitoring dashboards (e.g., AI Ally) - Personalized AI (e.g., CAPTAIN chatbot) - Cross-platform tools (e.g., Bullstop)	- Improve detection of nuanced attacks. - Enable immediate intervention. - Provide tailored support to victims. - Create a unified defence across social media, messaging apps, and forums.	- Requires significant R&D investment and interdisciplinary collaboration.
Policy & Collaboration	- Multi-stakeholder Frameworks	- Bridges the gap between researchers (evidence), policymakers (regulation), and industry (implementation).	- Creating effective, enforceable, and adaptable policies that keep pace with technological change.
Education & Awareness	- AI-Powered Chatbots & Serious Games - Data-Driven Educational Content	- Provides interactive, personalized learning about cyberbullying effects and resilience. - Empowers users with knowledge and coping strategies.	- Ensuring widespread adoption and cultural relevance of educational programs.

IV. FUTURE DIRECTIONS AND RECOMMENDATIONS

A. Innovations Required to Make AI More Effective in Cyberbullying Prevention

Considerable advancements in the field of artificial intelligence (AI) have been recorded over the past few years, but none of them are sufficient in preventing cyberbullying. The main weakness is in the area of natural language processing (NLP). Modern AI may notice the boldest attacks, but even it will always fail to realize the milder strategies- particular moments that foster bullying or understated sarcasm that is implemented with the expressive intention to harm. NLP improvement would thus play a critical role towards addressing such blind spots. It is empirically proved that, under such exposure to a large set of examples, the AI models are shown to have an increased ability not only to detect blatantly inappropriate expressions but also to identify the malicious patterns that are repeatedly used by harassers [48].

One of the most telling developments is the ability of AI systems to follow real-time online activity and act in a timely manner. Continuously tracking what happens, computational agents are now able to identify dangerous interactions and act before things go out of control. One such example is AI Ally dashboard that monitors Discord and other platforms. It allows users to identify the wrong exchange and to collect proof of the cyberbullying [49]. This is a tool that will then allow more effective counteractions and this will enhance defences against abuse. In a more practical sense, cyberbullying is alleviated

through AI personalisation: the use of individualized assistance that would be shaped to suit specific victims has proven to be significantly more efficient as compared to uniform, one-size-fits-all interventions. The case of CAPTAIN chatbot demonstrates this concept because the chatbot adjusts discourse on the fly and supports victims who fight against cyberbullying. Since these AI tools are environment-dependent and tuned to the requirements of respective users, such a tool poses a significantly greater resource than generic, standardized alternatives [50].

Colleagues and students, first, it can be noted that there is nothing exceptional with the urgency of developing effective technology to mitigate cyberbullying in a particular social-media ecosystem; instead, what is necessary is that the robust solution be platform-agnostic. Stated differently, the tool has to work harmoniously with a variety of social-networking platforms, instant messaging programs, and Internet message boards. A distinctive suggestion that has been made in this area is the Bullstop application, which will attach to already existing social-media sites and will build a shield against online bullying. Constantly scanning user feeds checking on objectionable content, Bullstop can use this to give warnings and can recommend ways that a bullying event can be defused before it escalates [48].

Ethical imperatives form a preset prerequisite in the academic field of enquiry of creating Artificial Intelligence capable of preventing cyberbullying. The transparent interface, privacy of the user, and maintaining ethical standards require careful implementation into a technological solution. The informed consent, the maintenance of clear and transparent data-management procedures, and the adoption of the mechanisms to help eliminate the misuse or the interruption of the system itself, are essential. AI systems can only be trusted to work effectively against cyberbullying when these ethical dimensions are comfortably tackled.

With the help of prioritizing such forward-thinking of ethical approaches, modern AI technologies will be able to improve their potential to maximize the possibilities of harmful online behaviour identification, alleviation, and elimination. Such development of AI would allow at once to increase the safety and supportiveness of online environments to all users in addition to enhancing the capacity of the system to identify and address potentially dangerous conduct.

B. Collaboration Between Researchers, Policymakers, and Industry Stakeholders

Researchers, stakeholders in the industry, and policymakers need to work together in order to reduce cyberbullying with the help of artificial intelligence in our modern world. Practical approach to the problem of online harassment detection and prevention is entirely in the domain of research organizations. In this case, the academic community also needs to develop a new architecture of algorithmic solutions, conduct empirical studies, and provide evidence-based recommendations on the responsible use of AI-powered technologies [48]. Legislators, however, play a pivotal role in the development of regulation governing the application of AI in the online environment. They would be responsible to oversee the actions of AI systems in terms of moral regulations, user privacy, and taking ownership of the possible negative consequences [49]. There are ethical dilemmas or an abuse issue that can occur without a proper policy framework of using AI assistance from AI-based tools.

The wide acceptance of AI technologies will be based on the stakeholders in the industry, especially those belonging to the tech companies' section, messaging, and social media platforms. They will be in charge of incorporating these technologies in their systems and making sure that they can be used. The efficiency of cyberbullying prevention using AI will rely on the implementation of the results of the studies and the compliance of the stakeholders of the industry to the policies [50]. These stakeholders can assist in achieving a safer online space by making sure AI solutions are practical, scalable to a reasonable degree and meet the legal requirements.

This cooperation results in a strategy to stop cyberbullying made up of the various areas. Policymakers make sure that ethical and legal issues are considered, scientists deliver the groundbreaking information and researches, and industry authorities implement these solutions. In collaboration, they can create the system which will help to popularise AI technologies and make them both ethically right and productive.

C. Education and Awareness Programs for Users and Online Communities

Cyberbullying needs sufficient education and awareness programs to provide users and online communities with awareness of cyberbullying, and related knowledge to approach and respond to it. Besides creating awareness, such programs should be crafted to impart the needed skills in users on how to handle bullying in online environments. Such efforts have a potential to attract a substantial number of people and leave a lasting impact on the reduction of cyberbullying behaviour through the use of state-of-the-art strategies, such as artificial intelligence (AI).

One of the AI tools which is rapidly becoming an effective means of avoiding cyberbullying is the CAPTAIN (Cyberbullying Awareness and Prevention Through Artificial Intelligence) chatbot.

The CAPTAIN chatbot provides users with personalised knowledge on the negative effects of cyberbullying and is available on-demand when the user requires assistance. They can also practise the responses and perfect it as through simulations of actual cyber bullying instances, the users can become more prepared and confident to handle should it arise online [50]. Having conversational AI in the educational program, the users will be offered with a personalised experience in which they will be able to receive preliminary measures that will match their unique needs and, therefore, will shift towards higher effectiveness.

The application of AI-enabled serious games is another good strategy since they present an entertaining approach to learning about the negative effects of cyberbullying. It is possible to achieve high-levels of user engagement (an extended interaction) through the use of serious games because, by incorporating game mechanics to educate people on relevant social topics in an entertaining, interactive form; they advance engagement with greater results. Such games can also employ AI to customise the experience of each player by either varying

the content or changing the level of difficulty depending on the actions of the player. Such interactive learning also allows users to understand the practical effects of cyber bullying as well as promoting empathies and respect in online communication [51]. Moreover, the games give room to the players to practice different scenarios and create an emotional intelligence they would require in events of being cyberbullied.

AI can also have a high level of utility in the creation of specific educational materials through the analysis of user activity and preferences. This will be a data-driven measure, which makes the content that is going to be used valid and suitable to the requirements of people with different users, which allows creating the educational content that will be appealing to this or that demographic group. As an example, a gamified learning content might be of greater value to younger users, and case studies and detailed discussions might be of greater value to senior users [52]. Optimizing the educational contents based on the user data, AI guarantees that awareness campaigns will reach a broad audience, not confined to users of a particular age group and cultural background.

Combined with these AI-based tools, outreach to social networks is a crucial part of achieving education and awareness campaigns. Campaigns can be promoted through the social media platforms, online forums and other virtual platforms, educational materials can be disseminated and positive online behaviour can be empowered. Peer-to-peer support is also another very important element of these networks since users may provide each other with tips and techniques to be successfully introduced to fight cyber bullying. Ensuring the community-based approach to education, awareness campaigns could be more open and effective and allow users to participate in some proactive steps in cyber bullying prevention.

Finally, AI can be implemented in an educational and awareness program, which offers new solutions to cyberbullying. Chatbots and serious games are examples of the AI-guided technologies that provide scalable, personalised and engaging solutions that not only created awareness but also grant users confidence to act. Focusing on the needs of users and tailoring information to fit their needs, AI could contribute to making the online world safer and more respectful so that individuals would recognise the harmful effects of cyberbullying and possess the skills and knowledge needed to combat it.

D. Ethical Frameworks and Guidelines for Responsible AI Deployment

Due to the rise in AI technologies deployment, it is crucial to make sure that these systems are developed and implemented responsibly. Ethical principles play a pivotal role in eliminating the abuse of AI in delicate fields such as the elimination of cyberbullying. Such frameworks could act as a guideline to the legislators as well as the developers in coming up with systems that put emphasis on privacy, justice, accountability and transparency.

AI systems, especially the ones preventing cyberbullying should adhere to the principles of ethics in order to safeguard the users. Fairness is one of the keys of these structures. The AI algorithms should also be open and are not supposed to accidentally affirm built-in prejudices. As demonstrated by study [53], ethical AI laws must be able to tackle the issue of fairness by ensuring that AI systems avoid discrimination against certain people based on race, gender, as well as, the socioeconomic status. Based on their observation, AI models trained with biassed data tend to strengthen such biases, which may harm the under-represented population. According to them, the frameworks of ethical AI must have provisions to reduce biases and have ethically fair returns to all users.

User privacy is another basic value of successful AI implementation. The increased interest in data privacy, including the social media, has evinced an issue of handling personal data with care by AI systems. [54] highlighted the importance of data collection and processing transparency and considered the ethical issues of AI used in moderating social media. They suggest that the AI-based platforms can provide more control over the user data and guarantee privacy without interfering with the system in terms of identifying harmful behaviour, i.e., cyberbullying. It [54] states that their paradigm promotes a compromise between protection of user rights and effectiveness of an AI system.

Moreover, responsibility is crucial even when it comes to the creation and implementation of AI systems. The systems developed by AI frameworks should have systems that demand responsibility of the makers of the systems. Developers should make their systems explicable to both users and external auditors in order to be able to understand the decision-making process, according to a study conducted on ethical AI of online platforms by [55]. They stress that since there is a role of AI system developers in the designing of the technology and its consequences, they should also be blamed.

We also need ethical rules which would make sure that we implement AI system in a manner which cannot be misused especially among the already vulnerable groups. As an example, to fight the issue of cyberbullying, one would have to design AI that would be able to detect the harmful behaviour without violating the freedom of expression of users. Guidelines have to consider the need of ensuring that the freedom of expression of the users is not infringed upon and the need to regulate the contents. This is of particular concern to social media sites which, are habitually accused of censoring content in a manner that is discriminative to certain groups of users.

Such structures should be adaptable and versatile due to the development of AI technologies. In the development of AI, as well as enhancing these ethical frameworks, systematic work by the researchers, legislators, and business executives is necessary as stakeholders of the current development of AI. This will ensure that AI remains a good influence that brings about a fairer and a more just digital society.

V. CONCLUSION

The multifaceted nature of cyberbullying, with its considerable social, psychological, and physical effects, particularly on younger social media users, calls for more effective intervention strategies than what global efforts presently can provide. While these measures are quite essential and provide a good foundation, their inadequacy calls for new solutions. With AI, the possibilities are endless to fight

cyberbullying via AI-powered solutions to directly take action in real-time and forecasting models to determine risk. Breakthrough advances in AI-related technologies, such as Natural Language Processing, Machine Learning, Image & Video Analysis, and Behavioural Analytics are providing mind-blowing tools that can be used to comprehend and address this pervasive problem.

However, if the transformative potential of AI is to be truly unlocked in cyberbullying prevention, key areas will require targeted development and innovation. These areas include fostering strong collaboration between researchers, policymakers, and industry stakeholders to ensure that best possible solutions are designed and implemented ethically. Meanwhile, in parallel with the collaboration, there will need to be extensive education and awareness initiatives targeting users and online communities to empower them with knowledge on the identification and administration of cyberbullying. Finally, establishing ethical frameworks and guidelines for the responsible use of AI must take priority to address challenging questions around privacy, bias, and accountability; ultimately paving the way for a safer and more positive online environment for all.

REFERENCES

[1] R. Slonje and P. K. Smith, "Cyberbullying: Another main type of bullying?" Scandinavian Journal of Psychology, vol. 49, no. 2, pp. 147–154, 2007.

[2] S. Bauman, "Types of cyberbullying," in Cyberbullying: What Counselors Need to Know, 2015, pp. 53–58.

[3] S. Mahbub, E. Pardede, and A. S. M. Kayes, "Detection of Harassment Type of Cyberbullying," Security and Communication Networks, vol. 2021, pp. 1–12, 2021.

[4] F. A. Esquivel, I. L. De La Garza López, and A. D. Benavides, "Emotional impact of bullying and cyber bullying," Revista Caribeña De Ciencias Sociales, vol. 12, no. 1, pp. 367–383, 2023.

[5] G. Niu, J. He, S. Lin, X. Sun, and C. Longobardi, "Cyberbullying victimization and adolescent depression," International Journal of Environmental Research and Public Health, vol. 17, no. 12, p. 4368, 2020.

[6] S. Hinduja and J. W. Patchin, "Bullying, cyberbullying, and suicide," Archives of Suicide Research, vol. 14, no. 3, pp. 206–221, 2010.

[7] O. M. E. Ramadan et al., "Digital Dilemma of Cyberbullying Victimization among High School Students," Children, vol. 11, no. 6, p. 634, 2024.

[8] H. Güllü, E. Karahan, and A. O. Akçay, "A comprehensive investigation of cyberbullying and cyber victimization," Education and Information Technologies, vol. 28, no. 10, pp. 12633–12650, 2023.

[9] A. Rijal, P. Thapa, A. Sapkota, and S. Rijal, "Social Media Use and its impact on mental health," Modern Issues of Medicine and Management, vol. 28, no. 2, 2024.

[10] L. Rüther, J. Jahn, and T. Marksteiner, "#influenced! The impact of social media influencing on self-esteem," Frontiers in Psychology, vol. 14, 2023.

[11] E. A. Vogel, J. P. Rose, L. R. Roberts, and K. Eckles, "Social comparison, social media, and self-esteem," Psychology of Popular Media Culture, vol. 3, no. 4, pp. 206–222, 2014.

[12] J. E. Copp, E. A. Mumford, and B. G. Taylor, "Online sexual harassment and cyberbullying in a nationally representative sample of teens," Journal of Adolescence, vol. 93, pp. 202–211, 2021.

[13] M. Smith, M. Nolan, and J. Gaffey, "Online safety and social media regulation in Australia," Griffith Law Review, pp. 1–17, 2024.

[14] A. Janković and L. Stošić, "Cyberbullying legislation: The role of cyberbullying law," Pravo - Teorija I Praksa, vol. 39, no. 4, pp. 97–108, 2022.

[15] N. Trompeter et al., "Cyberbullying prevalence in Australian adolescents: Time trends 2015–2020," Journal of School Violence, vol. 21, no. 3, pp. 252–265, 2022.

[16] K. Verma, B. Davis, and T. Milosevic, "Examining the effectiveness of artificial intelligence-based cyberbullying moderation on online platforms," AoIR Selected Papers of Internet Research, 2023.

[17] T. N. Prabhu, "U.S. Patent No. 9,686,217," Washington, DC: U.S. Patent and Trademark Office, 2017.

[18] M. Paciello et al., "The role of traditional and online moral disengagement on cyberbullying," Computers in Human Behavior, vol. 103, pp. 190–198, 2019.

[19] R. Kumar et al., "Role of Artificial Intelligence to address Cyberbullying and Future Scope," in Proc. CISES, pp. 974–977, 2023.

[20] J. De Angelis and G. Perasso, "Cyberbullying detection through Machine learning," International Journal of Management and Humanities, vol. 4, no. 11, pp. 57–69, 2020.

[21] O. K. E. Hussien, A. E. Aboutabl, and R. M. Y. Haggag, "Comparative performance of data mining Techniques for cyberbullying detection," International Journal on Recent and Innovation Trends in Computing and Communication, vol. 11, no. 11s, pp. 392–400, 2023.

[22] S. Salawu, Y. He, and J. Lumsden, "Approaches to Automated Detection of Cyberbullying: A survey," IEEE Transactions on Affective Computing, vol. 11, no. 1, pp. 3–24, 2017.

[23] M. A. Al-Garadi et al., "Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms," IEEE Access, vol. 7, pp. 70701–70718, 2019.

[24] L. Mirtskhulava, "Predictive Modelling for Mitigating Cyber-Violence by Leveraging AI," International Journal of Simulation Systems Science & Technology, vol. 25, no. 1, 2024.

[25] S. Gowthami et al., "Cyber Bullying Detection Using Deep Learning and Natural Language Processing," in Proc. ICICNIS, 2024, p. 1559.

[26] R. Rishi, M. Irfan, and G. Balamurugan, "NLP Techniques Cyberbullying Text Analysis on Twitter," in Proc. ICCSP, 2024.

[27] Y. K. Hsien, Z. A. A. Salam, and V. Kasinathan, "Cyber Bullying Detection using Natural Language Processing (NLP) and Text Analytics," in Proc. IEEE ICDCECE, 2022, p. 1.

[28] M. Al-Hashedi, L.-K. Soon, and H.-N. Goh, "Cyberbullying Detection Using Deep Learning and Word Embeddings," in Proc. ACM, 2019, p. 17.

[29] T. Milosevic, K. V. Royen, and B. Davis, "Artificial Intelligence to Address Cyberbullying, Harassment and Abuse," International Journal of Bullying Prevention, vol. 4, no. 1, pp. 1–10, 2022.

[30] A. Toktarova, D. Sultan, and Z. Azhibekova, "Review of Machine Learning Models in Cyberbullying Detection Problem," in Proc. SIST, 2024, p. 233.

[31] L. Cheng et al., "PI-Bully: Personalized Cyberbullying Detection with Peer Influence," in Proc. IJCAI, pp. 5829–5835, 2019.

[32] P. Vivekananth and N. Sharma, "Detecting Cyberbullying in Social Media: An NLP-Based Classification Framework," Indian Journal of Science and Technology, vol. 18, no. 5, p. 380, 2025.

[33] S. W. Azumah et al., "Cyberbullying in text content detection: an analytical review," International Journal of Computers and Applications, vol. 45, no. 9, p. 579, 2023.

[34] M. Kulkarni et al., "Cyberbully and Online Harassment: Issues Associated with Digital Wellbeing," arXiv preprint, arXiv:2404.18989, 2024.

[35] A. Ioannou et al., "From risk factors to detection and intervention," Behaviour and Information Technology, vol. 37, no. 3, p. 258, 2018.

[36] H.-Y. Chen and C. Li, "HENIN: Learning heterogeneous neural interaction networks for explainable cyberbullying detection," arXiv preprint, arXiv:2010.04576, 2020.

[37] B. G. Bokolo and Q. Liu, "Cyberbullying detection on social media using machine learning," in IEEE INFOCOM Workshops, 2023.

[38] R. K. Sharma, "Ethics in AI: Balancing innovation and responsibility," International Journal of Science and Research Archive, vol. 14, no. 1, p. 544, 2025.

[39] R. Bala, "Challenges and ethical issues in data privacy," International Journal of Information Retrieval Research, vol. 12, no. 2, p. 1, 2022.

[40] A. Fabris et al., "Fairness and Bias in Algorithmic Hiring: A Multidisciplinary Survey," ACM Transactions on Intelligent Systems and Technology, 2024.

[41] D. Ritika et al., "Predicting Cyberbullying Behavior in Social Media," in Proc. ICMI, 2024.

[42] C. Barrett et al., "Identifying and Mitigating the Security Risks of Generative AI," Foundations and Trends® in Privacy and Security, vol. 6, no. 1, p. 1, 2023.

[43] S. H. A. Harbi et al., "Responsible Design Patterns for Machine Learning Pipelines," arXiv preprint, arXiv:2306.01788, 2023.

[44] L. Lobschat et al., "Corporate digital responsibility," Journal of Business Research, vol. 122, pp. 875–884, 2019.

[45] Y. Wang, "When artificial intelligence meets educational leaders' data-informed decision-making," Studies in Educational Evaluation, vol. 69, p. 100872, 2020.

[46] B. Klímová et al., "Ethical issues of the use of AI-driven mobile apps for education," Frontiers in Public Health, vol. 10, 2023.

[47] D. U. Patton et al., "Meet Them Where They Are," The MIT Press eBooks, 2023, pp. 43–55.

[48] T. Ige and S. Adewale, "AI powered anti-cyber bullying system," arXiv preprint, arXiv:2207.11897, 2022.

[49] R. Biswas et al., "Securing social spaces," arXiv preprint, arXiv:2404.03686, 2024.

[50] A. T. Lian, A. C. Reyes, and X. Hu, "CAPTAIN: An AI-based chatbot for cyberbullying prevention," in Artificial Intelligence in HCI, 2023, pp. 98–107.

[51] J. Pérez et al., "The uses of chatbots in the context of children and teenagers bullying," Cogent Psychology, vol. 11, no. 1, 2023.

[52] R. M. Kowalski and T. H. Witte, "Effectiveness of Artificial Intelligence-Based Cyberbullying Prevention Programs," Social Media + Society, vol. 8, no. 1, 2022.

[53] J. Doe and J. Smith, "Ethical considerations in the deployment of AI," in Proc. Int. Conf. Artificial Intelligence Ethics, 2023, pp. 45–56.

[54] A. Johnson, E. Brown, and L. Zhang, "The global landscape of AI ethics guidelines," in Proc. IEEE International Conference on AI and Ethics, 2022, pp. 12–23.

[55] M. Brown and S. Lee, "Towards a conceptual framework for ethical AI development," in Proc. Int. Conf. Responsible AI, 2024, pp. 89–100.

[56] Tan, J. M., Wider, W., Rasli, A., Jiang, L., Tanucan, J. C. M., & Udang, L. N. (2024). Exploring positive impact of social media on employee mental health: A Delphi method. *Online Journal of Communication and Media Technologies*, *14*(3), e202436.

[57] Ting, T. T., Chin, W., Lim, L. X., Yuen, S. M., Chaw, J. K., Husin, W. N. A. A. W., ... & Siddiqui, Y. A. (2024). Exploring the Factors Affecting Mental Health and Digital Cultural Dependency among University Students. Pakistan Journal of Life and Social Sciences, 22(2), 16062-16080.