

# Role of Artificial Intelligence to address Cyberbullying and Future Scope

Rakesh Kumar

Uttaranchal Institute of Management  
Uttaranchal University, Dehradun India  
Graphic Era University  
ayusharma.9356@gmail.com

Ayushi Sharma

Uttaranchal Institute of Management  
Uttaranchal University,  
Dehradun India  
aigrakeshkumar@gmail.com

Manleenjot kaur

Uttaranchal Institute of Management  
Uttaranchal University,  
Dehradun India  
manleenjotkaur@gmail.com

Kapil Joshi

Department of CSE,  
Uttaranchal Institute of Technology, Uttaranchal University,  
Dehradun, India.  
Kapileng0509@gmail.com

Sumeshwar Singh

Department of Computer Science and Engineering  
Graphic Era Hill University  
Dehradun Uttarakhand  
ssingh@gehu.ac.in

**Abstract**—Technology has become an important facet in every field. There is risk and challenge evolved through technology. Cyberbullying is an important concern in technological development. In the past ten years, cyberbullying has been recognized as a significant issue affecting young people. This paper summarizes recent research and addresses broader ideas. It addresses definitional issues like repetition and power imbalance, as well as different types of cyberbullying, age, and gender differences, overlap with traditional bullying, distinctions between cyberbullying and traditional bullying, the causes and effects of cyber victimization, coping mechanisms, and prevention and intervention options. This study emphasizes how Artificial intelligence can address cyberbullying issues in society. The study finds out problems regarding cyberbullying, role of AI in bullying and future recommendation. Society is victimized by these issues which are increasing day by day. The study emphasized a proposed model for addressing issues with the help of AI. The study emphasized futuristic challenges about cyberbullying.

**Keywords:** Artificial Intelligence, Cyberbullying, Computer, Machine Learning

## I. INTRODUCTION

Due to increased urbanization and globalization, smart cities are having difficulty maintaining development and a high standard of living. Cyberbullying (CB) is a type of bullying that occurs online and involves creating deliberate hostility towards a different group or individual on social media platforms. Bullying has both physical and psychological effects on its victims, and Cyberbullying can be used to lessen susceptibility in smart cities. The intention to injure oneself is a crucial role in the transition from hostility to CyberBullying, and current frameworks lack an intelligent automated mechanism to change and reliably recognise the CyberBullying contents. In a smart city, a brand-new automated classification model is created to recognise CyberBullying texts from the twitter engine.

The study created a text classification engine that pre-processes tweets, extracts features, and categories without data overfitting. It also created an automated classification model that fits with large redundant datasets. Finally, it created a novel Deep Decision Tree classifier that uses hidden layers of Deep Neural Network (DNN) to produce improved classification accuracy [1]. 25% of parents report experiencing cyberbullying, which shows that people are using social networking sites to harass others. Due of Form

spring high concentration of bullying-related content, we web crawled it to gather data for our own labelled dataset. In order to create a model for cyberbullying, we used data mining algorithms and prioritized recall above precision[2]. Children and young people who experience cyberbullying suffer severe consequences, with many cases leading to suicide.

Cyberbullying is when someone uses a computer, a phone, or another device to send or post messages that are meant to hurt or humiliate someone else. Bullying that takes place online and outside of a school, sports arena, or workplace is known as cyberbullying. We identify textual cyberbullying, one of the main types of cyberbullying, using a corpus of comments from YouTube videos. Comparatively better than multiclass classifiers using a collection of labels are classifiers that categorize a comment into a certain category or not [3].

A group of people engaging in cyberbullying scare others by taking advantage of modern technology. Flaming, harassment, denigration, impersonation, outing, boycott, and cyberstalking are just a few examples of the various forms that cyberbullying can take. Cyberbullying takes many different forms, including denigration, impersonation, outing, boycotting, and cyberstalking, but the latter is less harmful because it can be quickly identified. In addition to the victim and cyberbully, cyberbullying also involves the bully, victim, bystander, assistant, defense, reporter, accuser, and reinforcer [5].

According to two studies, cyberbullying is less common than traditional bullying, yet it is reported outside of schools more often than within. Bullying via phone calls and texts was most common, while bullying via instant messaging had a more detrimental effect. The two studies had a range of age and gender inequalities, and most of the cyberbullying was carried out by one or a small group of students, typically from the same year group. Internet use and being a cyber victim were intertwined, and many cyber victims were also traditional "bully-victims." As coping mechanisms, students suggested blocking or ignoring communications and talking to someone, yet many had told no one [15]. Because cyberbullying and low self-esteem are related, bullying prevention programmes should cover this topic and educators should step in to stop cyberbullying instances[16].

## II. THEORETICAL BACKGROUND

Few research teams are focused on the identification of cyberbullying, which is harm that is purposefully and repeatedly perpetrated through electronic text. This study used textual context in YouTube video comments to detect cyberbullying with a 66.7% accuracy rate[2]. Natural language processing (NLP), machine learning (ML), and deep learning are just a few examples of the artificial intelligence (AI) or algorithmic techniques that social media platforms are

increasingly using to help moderate or process cyberbullying situations (DL). To be proactive with moderation, AI is also utilized to crawl/screen content before it is reported to platforms. Social media sites are increasingly using artificial intelligence (AI) or algorithmic methods to help moderate or process cyberbullying situations, such as natural language processing (NLP), machine learning (ML), and deep learning (DL). In an effort to be proactive with moderation, AI is also employed to crawl and review content before it is submitted to platforms [4].

TABLE I. LITERATURE REVIEW

Technology	Year	Findings
AI	2013	In the past ten years, cyberbullying has been recognized as a serious issue affecting young people. This paper reviews recent findings and discusses general ideas in the field, including repetition and power imbalance, different types of cyberbullying, age and gender differences, overlap with traditional bullying and the chronology of events, the causes and effects of cyber victimization, coping mechanisms, and prevention/intervention options. The validity of various coping, intervention, and preventative techniques is also covered [10].
AI	2018	Exaggerated claims and conflicting findings plague the research on cyberbullying. It is crucial to reach some level of agreement on the definition of the phenomena as a scientific term and to quantify cyberbullying in a "bullying context" in order to create a meaningful and coherent body of knowledge. We propose that bullying in general, including verbal, physical, indirect, and relational bullying, should be divided into a subcategory or particular form called cyberbullying[11].
AI	2013	Teenagers now possess the same level of computational power as big enterprises, but some have chosen to abuse it. Cyberbullying, which is described as "willful and repeated harm done through the use of computers, mobile phones, and other electronic devices," is the word used to describe this issue. Because the attacks can be more severe, frequent, sneaky, and intimidating to stop, cyberbullying may be seen as more malicious than traditional or in-person bullying. Because they can include a wide variety of media in their attacks, such as sounds, manipulated images, text, video, slide shows, and polls, online bullies are sometimes much more nasty than their offline counterparts.[12].
AI	2012	This essay makes the case that cyberbullying is a low-prevalence problem that hasn't spread more widely over time or produced a lot of "new" bullies and victims. It is further stated that the subject of potential adverse impacts of cyberbullying has not received much serious scientific attention. It is based on two large samples of students, one from the USA and one from Norway. It is advised that schools focus the majority of their anti-bullying initiatives on preventing conventional bullying, along with an essential system-level strategy.
AI	2012	The need for a practical and academically valid definition of cyberbullying is critical since it is a reality of the digital age. This article adds to the body of knowledge regarding the definition of cyberbullying by emphasizing the fundamental components of repetition, power disparity, intention, and aggressiveness. Given the public character of the content in the internet context, it is recommended that the concept of recurrence needs to be redefined. An example definition is provided, and a clear differentiation between direct and indirect cyberbullying is given. Overall, the analysis sheds light on how the fundamental components of bullying have changed and should still hold true in our parallel cyber universe.

## III. PROBLEM DECOMPOSITION

By defining them as speaking on sensitive themes and identifying negativity and vulgarity, the issue of identifying damaging communications on social networking sites can be resolved. People frequently post comments and messages on social networking platforms, but when those posts touch on delicate subjects that could be private to an individual or a certain group of people, it may be termed cyberbullying [3].

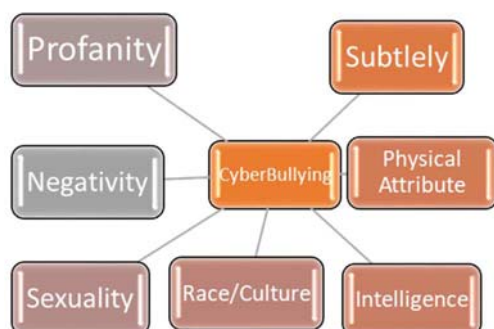


Fig. 1. Problem Decomposition: A comment involving a combination of negativity or profanity and topics that are personal and sensitive are those can be most hurtful, forming candidates for cyberbullying

## IV. CONCEPT OF CYBERBULLYING

One of the studies that was examined did not provide a definition but instead borrowed bullying concepts from the

literature, such as hostility and harassment. The word "cyberbullying" serves as a catch-all for a variety of cyberbullying tactics, although the scientific community is still divided over the nature and scope of the epidemic. Such polysemy produces some inconsistent research findings and makes it difficult to compare investigations. Systematic persecution clearly has the characteristics of harassment, which are not always present with cyberbullies. Cyberstalking, cyber harassment, cybernetic victimization, and cyber-victimization are just a few examples of the various definitions of cyberbullying.

The hostile nature of the conduct and the intention to inflict harm are essential components in the definition of cyberbullying. A recent trend is to categorize cyberbullying as merely "digital bullying." The usage of a broad definition (bullying) or the requirement to swap it out for one that is more particular are both up for debate. Below, we'll examine the arguments made in support of the details that characterize bullying and cyberbullying. Although there is disagreement about whether power disparity and repetitious behavior are characteristics of cyberbullying, they do appear in real-world bullying situations.

The purported incapacity of the intended victim to respond, or even a lack of technology abilities that would allow a better response, has been linked repeatedly to power imbalance. Digital networks give thousands of people access to the degrading content that is used in cyberbullying on a

broad scale. In contrast to bullying, whose audience is restricted to those present at the time of the attack, the more it is viewed or shared, the larger the audience. Researchers who study cyberbullying have not reached a consensus on the subject's concepts, terminologies, or the similarities and differences between cyberbullies and traditional bullies. There will probably be more scientific differences as this theoretical framework develops in this developing field[5].

## V. HOW AI CAN ADDRESS CYBERBULLYING

The authors of this research discuss how they can use AI to identify cyberbullying. The authors categorize messages as bullying or innocuous. To do this, they employ term frequency-inverse document frequency. To shield victims from hazardous traffic in 5G networks, the operator uses policy control. To detect cyberbullying in the World of Tanks game, an in-game chat data gathering system has been created. The likelihood of online bullying was shown to be substantially lower among very new players than among more seasoned players, indicating that it might be a taught behavior from other players [6].

The authors of this research discuss how they can use AI to identify cyberbullying. The authors categorize messages as bullying or innocuous. To do this, they employ term frequency-inverse document frequency. To shield victims from hazardous traffic in 5G networks, the operator uses policy control [7]. An illustration of the application of artificial intelligence to the creation of anti-cyberbullying solutions in the gaming sector is provided. Aggression has increased as a result of cyberbullying replacing in-person interactions [8]. This article discusses how the dignity theory may be applied to comprehend bullying practices. Here, the term "dignity" refers to the inherent value of every human being. Online safety can be used to analyze cyberbullying, which restricts the applicability of artificial intelligence-based solutions [9].

Building a system to identify, address, or prevent cyberbullying is a long-term objective. While it is challenging to identify a single scoring schema for cyberbullying that applies in all situations, the scoring schema offered is a useful metric to evaluate the usefulness of the information acquired and methods utilized in this study. To create an anti-cyberbullying system, researchers require access to vast data sets, and World of Tanks has been identified as an excellent target game for data collection. To merge player/vehicle information from Wargaming.net with link information scraped from WotReplays.com, a data gathering method was created.

A straightforward set of SQL statements and a specially created Windows client were used to categorize the data. Outcomes exceeded expectations in terms of success. The chat data and player statistics from wargaming.net public services showed that cyberbullying spreads among communities like an epidemic rather than necessarily coming from one specific set of players. To support this conclusion, more information and investigation are required. The investigation of time-of-death and cyberbullying showed that most of the cyberbullying behaviour happens shortly after death. Businesses have tried to stop intimidation by limiting all in-game dialogue or filtering out harsh words, but these measures haven't worked.

This suggested remedy is workable, useful, and practical. It's possible that insufficient experience precluded the full

utilisation of sentiment services. In-game chat has a specialised language structure and content, making it challenging for regular services to understand. There is no SQL that can determine whether a comment is directed at a specific target, but Microsoft's analysis services may be able to help in determining the target of a given chat session. The study opened up several avenues for further study, including non-English speaking users in data collection, using sentiment analysis tools, and compiling a list of phrases used by gamers[7].

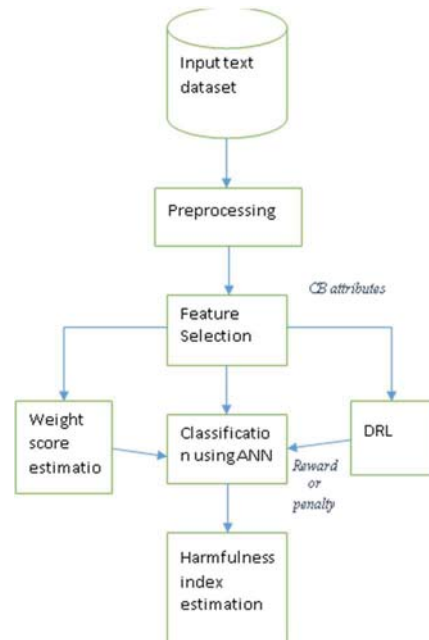


Fig. 2. AI Role in Cyberbullying

- AI is a branch of computer science that employs sophisticated mathematical formulas to simulate human thought. John McCarthy made the initial suggestion in 1956, and since then it has transformed several areas and industries, paving the way for the use of AI-based cybersecurity techniques.
- While Deep Learning (DL) is a subset of ML that can handle massive datasets and enhance cybersecurity, ML is a subset of AI that teaches machines how to make judgements.
- Cybersecurity specialists must address the threats and benefits of AI to assure cyber safety.
- To protect networks from current issues like botnets, which are used to conduct Distributed Denial of Service (DDoS) assaults, and IDPS, which produce false alerts and divert attention from genuine threats, cybersecurity experts need to discover new solutions.
- DDoS assaults heavily rely on botnets, which are also used to steal data and steal people's identities. Network and system administrators utilize IDPS technology to find intrusions and get email notifications.
- Network administrators must correctly configure IDPS tools to reach a better security level. Because network traffic is variable, setting up and configuring an IDPS takes time [16].



- AI systems can support cyberattack defence, but they can also create new vulnerabilities for people and companies. Frequent users frequently overlook security updates, resulting in background running unpatched software.
- Due to less-than-perfect compliance, efforts to avoid hostile uses of AI are unlikely to be successful, although risk can be decreased through actions including enhancing system security, responsibly sharing breakthroughs, and raising threat awareness.
- Cybercriminals and nation-state actors are increasingly utilizing AI, which enables them to take advantage of unknown weaknesses and exploit sensitive information to use it against a nation.

## VI. RESEARCH FINDING AND RECOMMENDATIONS

The most crucial point is that the intention of self-harm is what triggers the change from aggressiveness to CB. The social network's automatic behavior prompts moderators to analyses user content, but the existing frameworks lack an intelligent automated mechanism to change and detect CB contents quickly and accurately. Study emphasized following recommendation of AI can address cyberbullying: -

- Automated crime detection can lessen cyberbullying, a problem that AI can address. Artificial intelligence (AI) can address issues including fake message identification, cyberfraudster, sexual harassment, and more.
- Awareness about technology is still lacking among people. AI can implement awareness among people. Cybercrimes are increasing day by day due to lack of knowledge of technology.
- Government policy is important issues for AI in bullying detection like phone tapping, tracking of person etc. The AI implemented policy shall be in favor of detection of cyberbullying.
- Cyberbullying detection applications can be developed and disseminated with people for more awareness. It is important to involve AI in these application for smooth implement.

## VII. LIMITATION OF STUDY

This study is based on cyberbullying aspect in society and how AI decomposed these issues. The study emphasized concept of cyberbullying and how can be overcome with the help of artificial intelligence inventions. Study based on previous online study and analysis on the base of published literature. Further study may be conducted on application-based analysis which can be practically implemented against cyberbullying. Following are the limitations of this study.

- This is an analysis study based on already published literature. Analysis within cyberbullying aspect.
- Study conducted on the base of fact available at online platform. Further study may be conducted on an empirical approach.
- There are several issues regarding cyber victims. But study is concerned with cyberbullying aspect.

## VIII. CONCLUSION

The use of technology is crucial in combating cyberbullying. The study focused on the issue of cyberbullying and how AI can address it. Many techniques have been developed to stop cyberbullying; however, they are still unable to stop all kinds of problems. The study put forth a cyberbullying model that can be used to define the problems with cyberbullying. Nonetheless, numerous new criminal avenues have opened since the invention of the solution. In this aspect, AI is capable of quickly and precisely resolving various problems. Artificial intelligence can effectively managed cyberbullying issued in society.

## REFERENCES

- [1] Yuvaraj, Natarajan, et al. "Automatic detection of cyberbullying using multi-feature based artificial intelligence with deep decision tree classification." *Computers & Electrical Engineering* 92 (2021): 107186.
- [2] Reynolds, Kelly, April Kontostathis, and Lynne Edwards. "Using machine learning to detect cyberbullying." 2011 10th International Conference on Machine learning and applications and workshops. Vol. 2. IEEE, 2011.
- [3] Dinakar, Karthik, Roi Reichart, and Henry Lieberman. "Modeling the detection of textual cyberbullying." *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 5. No. 3. 2011.
- [4] Milosevic, Tijana, et al. "Effectiveness of Artificial Intelligence-Based Cyberbullying Interventions From Youth Perspective." *Social Media+ Society* 9.1 (2023): 20563051221147325.
- [5] Ali, Wan Noor Hamiza Wan, Masnizah Mohd, and Fariza Fauzi. "Cyberbullying detection: an overview." 2018 Cyber Resilience Conference (CRC). IEEE, 2018.
- [6] Yuvaraj, N., et al. "Nature-inspired-based approach for automated cyberbullying classification on multimedia social networking." *Mathematical Problems in Engineering* 2021 (2021): 1-12.
- [7] Murnion, Shane, et al. "Machine learning and semantic analysis of in-game chat for cyberbullying." *Computers & Security* 76 (2018): 197-213.
- [8] Rakhmatov, dilmurod. "methods and effectiveness of the use of artificial intelligence in the fight against cyberbullying." *Journal of Academic Research and Trends in Educational Sciences* 1.4 (2022): 122-129.
- [9] Milosevic, Tijana, Anne Collier, and James O'Higgins Norman. "Leveraging Dignity Theory to Understand Bullying, Cyberbullying, and Children's Rights." *International Journal of Bullying Prevention* (2022): 1-13.
- [10] Slonje, Robert, Peter K. Smith, and Ann Frisén. "The nature of cyberbullying, and strategies for prevention." *Computers in human behavior* 29.1 (2013): 26-32.
- [11] Olweus, Dan, and Susan P. Limber. "Some problems with cyberbullying research." *Current opinion in psychology* 19 (2018): 139-143.
- [12] Sabella, Russell A., Justin W. Patchin, and Sameer Hinduja. "Cyberbullying myths and realities." *Computers in Human behavior* 29.6 (2013): 2703-2711.
- [13] Olweus, Dan. "Cyberbullying: An overrated phenomenon?." *European journal of developmental psychology* 9.5 (2012): 520-538.
- [14] Langos, Colette. "Cyberbullying: The challenge to define." *Cyberpsychology, behavior, and social networking* 15.6 (2012): 285-289.
- [15] Smith, Peter K., et al. "Cyberbullying: Its nature and impact in secondary school pupils." *Journal of child psychology and psychiatry* 49.4 (2008): 376-385.
- [16] Calderon, Ricardo. "The benefits of artificial intelligence in cybersecurity." (2019)