International Journal of Engineering Technology & Advanced Learning
ISSN: 3107-4227, Vol. 02 (01), March 2026, pp. 38-45

Rapid Detection of Cyberbullying Using Artificial Intelligence Techniques: A Survey of Methods, Applications and Challenges

Sangeeta Binjhade[1], Sana Khan[1], Dr. Damodar Tiwari[1], Dr. Kailash Patidar[1]
[1]*Computer Science &Engineering, Bansal Institute of Science and Technology,* Bhopal, India

Abstract

The use of AI algorithms has become increasingly valuable in the battle against cyberbullying due to the widespread use of social media and other types of online communication. The ever-increasing level of user-generated content and the immediate character of online communication make it ineffective and unfeasible to monitor hand. Cyberbullying, which is a deliberate and recurrent harassment using the internet, is very dangerous psychologically and emotionally, especially among minors. The paper provides an overview of the concept of cyberbullying, its features, and prominent types, as well as discusses the nature of the problem of its automatic detection at length. Among the most recent AI methods covered are transformer-based architectures, DL models, and classical ML models for quickly and accurately identifying malicious content. The study also delves into data gathering, preprocessing, feature extraction, and assessment methods for rapid, real-time cyberbullying detection models. Also, there is the discussion of AI-based monitoring and alert systems, which might help with proactive intervention, parental monitoring, and a safer online space.

Keywords—Cyberbullying detection, Artificial intelligence (AI), Social media monitoring, and Real-time detection.

1. Introduction

The development of online communication systems, including social media, online forums, online games, and instant messaging, has radically altered the nature through which individuals communicate and exchange information [1]. Although these sites make it easier to connect and share information; they have also led to cyberbullying which has become a prevalent form of internet abuse that is extremely dangerous psychologically, socially and emotionally to the victims [2][3]. Anonymity, speed and the wide availability of web resources increase the impact of cyber bullying and the matter of cyber bullying is an emergent one to make sure that cyber bullying is identified and controlled at the right time.

Cyberbullying has become a serious social issue because technology is influencing nearly all aspects of everyday life. Systematic and intentional damage inflicted by cyber platforms is called cyberbullying, and it poses unique challenges that make it distinct compared to more traditional forms of bullying [4]. This has been on the rise especially among teenagers who are going through tricky social lives by using social media, messaging programs, and even in the world of online games [5]. The online environment provides secrecy and a large reach, where harasser can commit harassment without necessarily being reprimanded.

The conventional methods of cyberbullying detection based on manual moderation and rules are still inadequate due to a lack of scalability and the inability to adapt to changes in language patterns and other contextual factors [6][7]. The Artificial Intelligence (AI) techniques, specifically, the ML and DL model have become viable options in terms of automating cyberbullying[8][9]. These techniques can analyze data streams of large scale, learn complicated linguistic patterns and also detect with higher accuracy on a variety of platforms.

This survey provides a thorough analysis of AI-based methods for detecting cyberbullying as quickly as possible. It also critically examines the available methods, datasets, applications and evaluation strategies, and also identifies the important challenges including data imbalance, understanding the context, real-time limitations and ethical issues. The report concludes by outlining future research directions and outstanding research topics related to responsible, scalable, and effective cyberbullying detection systems.

1.1. Structured of the paper

This paper is organised as follows: Section II introduces the notions, features, and major types of cyberbullying, as well as the difficulties in detecting it. ML, DL, and transformer-based algorithms are some of the AI techniques discussed in Section III. Section IV explores high-speed, real-time AI-controlled systems for cyberbullying detection and their use. Section V contains a summary of the relevant literature, and Section VI leads to the study's conclusion and future research perspectives.

E-mail ID: binjhadesangu@gmail.com

Cyberbullying received significant attention from researchers, resulting in various definitions intended to delineate the phenomenon. Cyberbullying is often indicated by repeated threats, the dissemination of humiliating remarks on social media platforms or forums, or the transmission of menacing private communications [10]. According to the majority of studies, cyberbullying occurs when one person uses electronic methods to threaten another person, such as social media sites, emails, forums, and instant messaging.

1.2 Concepts and Characteristics

Cyberbullying is a kind of repetitive and premeditated aggressive conduct that is conducted with the help of digital technologies like social media, online gaming platforms, and messaging apps. Cyberbullying is an idea that focuses on how technology is abused to negatively affect people who are sometimes facilitated by anonymity and lack of responsibility. The main features of it are 24/7 availability, speedy and extensive spread of harmful material, endurance of digital documentation, and the inability to oversee and intervene. Cyberbullying is a complicated and multi-dimensional phenomenon since it manifests in several ways, including but not limited to: irritation, doxing, exclusion, trolling, cyberstalking, flaming, outing, denigration, and hate speech.

1.3 Key forms of the cyberbullying phenomenon.

The swift increase in the spread of online participation, people are starting to use social media and messaging services more, and they usually do not realize the risks[11]. Cyberbullying takes different shapes like harassment, denigration, exclusion, outing, flaming, cyberstalking, impersonation, and trickery, which all demand different analysis and monitoring strategies. Figure 1 shows the most common forms of cyberbullying on social media.

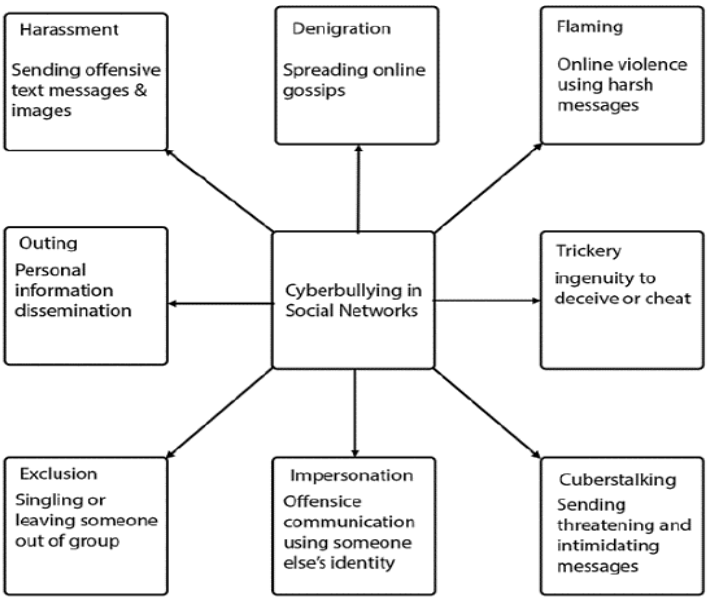

Figure 1: Cyberbullying on a social media platform

1. **Denigration (Denigration):** Cyberbullying is a kind of online harassment that " is feasible to accomplish by sharing the victim's private photo, video, or information on social media," and it includes insults, threats, and other forms of verbal and nonverbal abuse directed at the victim.
2. **Rejection (Exclusion):** A social way of excluding someone from a group. Such cyberbullying typically takes place in online forums or groups on social media sites.
3. **Cyberbullying (Harassment):** The internet version is very much like the old-fashioned one. The presence of aggressive and distressing behavior that is influenced by the victim's intellectual and physical characteristics constitutes this manifestation. As technology develops, cybercriminals may remain anonymous and persistent online.
4. **Disclosure of secrets (Outing):** The consequence of this is what the law calls " sharing private information about the victim on social media, which was never intended for public consumption" on social media. The purpose of posting secrets is to make fun of or humiliate the person, and this often happens without their knowledge or consent.
5. **Inflammation (Flaming):** One way to inflame things is to send and post threatening content to someone or some group online, which creates conflict situations. This kind of expression is most prevalent in online discussion groups and chat rooms that are specifically tailored to a certain topic.

1.4 Challenges in Cyberbullying Detection

There are several challenges faced in cyberbullying detection as follows:

1. **Cultural diversity for cyberbullying:** The language of a country is fundamental to its way of life. Cyberbullying has become a common problem across many nations, so using a dataset from one culturally varied nation and testing it over another would not yield a viable prediction model.
2. **Language challenge:** One of the most problematic aspects of cyberbullying detection is capturing context and evaluating the sentiment from various sentence patterns.
3. **Dataset challenge:** The sensitive nature of the personal information involved makes data retrieval from social media platforms a difficult task. Not to mention that social media platforms keep user data private [12]. These problems make it difficult to get high-quality data from social media, reducing the amount of useful data for enhancing education.
4. **Data representation challenge:** The nature of cyberbullying and the necessity of human interaction make it challenging to set up a system that can identify it effectively. Words used to harass or threaten someone on social media were directly detected in most of the exploratory studies [13]. Nevertheless, there are challenges associated with disentangling features based on content.
5. **Natural Language Processing (NLP) challenges:** Comprehending the text's meaning is the primary obstacle in natural language processing [14]. The important thing is to create the right language, connect its parts, set the scene, and get the semantic information out of the data.

2 Artificial Intelligence Techniques for Cyberbullying Detection

This section describes ML, DL, and transformer-based algorithms to detect cyberbullying, as well as the conventional detection pipeline. The focus is directed at the proper learning of features and proper classification of complex social media content.

2.2 Detection Framework and Process

The detection procedure involves gathering data, cleaning it up, extracting features, and training a model to accurately identify cyberbullying content. Performance is measured based on conventional classification measures.

2.2.1 Data Acquisition

Dataset preparation includes identifying, acquiring, and distributing data sources for training and testing. Social media APIs facilitate data collection: the Twitter API (Twitter4J, Tweepy) is most commonly used; the Facebook Graph API with strict controls; the Instagram Graph API for public content; the YouTube Data API for comments; and REST APIs for generic access[15].

2.2.2 Preprocessing Pipeline

The preprocessing of data is used to prepare data as the input to algorithms. They are: data cleaning (URL removal, duplicates removal, spam removal, missing-value removal), tokenization (text tokens are the result of word level and subwordlevel and character level breaking), stop words removal, normalization (lowercase conversion, contraction expansion, spelling standardization, emoji handling), stemming and lemmatization (removal of words to root forms), and data labeling (binary or multi-class labeling).

2.2.3 Feature Extraction

The reason behind this is that feature extraction greatly influences algorithm performance. There are standard techniques, that are: TFIDF (importance of words), Bag of Words (frequency counts), N-grams (word sequences), PoS tagging (grammatical roles), semantic features (meaning-based), topic modeling (LDA), PCA (dimensionality reduction) and sentiment analysis (emotional tone).

2.2.4 Model Training and Evaluation

Model learning is used to learn algorithms that classify instances of cyberbullying. The training process consists of: data splitting, hyperparameter optimization, cross-validation and class imbalance (SMOTE, under-sampling, weighted loss) [16]. Measures of evaluation are: accuracy (avoiding false negatives and false positives in general), precision (reducing FP), recall (reducing FN), F1-score (harmonic mean) and ROC-AUC.

2.3 Machine Learning Approaches.

A number of supervised ML classification approaches are compared after a step of preprocessing and feature selection. In particular, the comparison of the most common and simple to use models is chosen, and their training and prediction times are given attention. The following classifier was used in the study:

2.3.1 Multinomial NB Classifier

The Multinomial NB classifier is a popular tool in NLP due to its probabilistic nature [17]. For textual content such as emails or news articles, this algorithm predicts the tag using Bayes' theorem. It does this by determining the probability of a specific sample and then offering a tag with a high likelihood of accuracy.

2.3.2 Bagging Classifier

One method of machine learning is the bagging classifier, which uses an ensemble of models to make predictions [18][19]. To reduce model overfitting, each model is trained independently and then averaged.

2.3.3 Logistic Regression

A different type of ML method that is in the guided learning group is logistic regression. Its primary function is to use a set of independent factors to predict a set of pre-categorized dependent variables. It can break down data into binary choices, such as 0 or 1, Yes or No, True or False, and the likelihood that the information given belongs to the 1 category.

2.3.4 Support Vector Machine (SVM)

SVM is a method for classifying data that uses training data to find the best hyperplane for classifying the data. The orientation and placement of the hyperplane are determined by the support vectors, which are a set of extreme vectors or points used by SVMs to construct them. The decision boundaries are the hyperplane and the set of supporting class vectors on one side, and the hyperplane and the set of supporting class vectors on the other side.

2.4 Deep Learning Models.

A large number of DL-based models have found various uses in the identification of cyberbullying. A few well-liked models are deep belief networks, deep autoencoders, Boltzmann machines, and DNNs.

2.4.1 Convolutional Neural Network (CNN)

CNNs are widely used in image identification and categorization. The layers that make up these networks execute tasks like as feature extraction, dimensionality reduction, and classification. These layers include convolutional, recurrent linear unit, pooling, and fully connected ones. A few examples of CNN's numerous applications include healthcare, image and video analysis, AD, time-series forecasting, and NLP.

2.4.2 Recurrent Neural Network (RNN)

The acronym "RNN" refers to a type of NN that processes sequential text data. When faced with such a test [20]. There are innumerable issues that RNN may enhance, including the accuracy of spam classifiers, time-series data, sales forecasts, stock forecasts, and various others. Regarding alternative models, the input is provided as a string of text using text preprocessing methods (e.g., Word2Vec, TF-IDF, BagOfWords).

2.4.3 Long Short-Term Memory

An improved version of RNNs, LSTM can maintain the long-term dependencies in sequential data. LSTMs with their unique memory units and gating control allow for the storing of pertinent past information, even though a long sequence of events of which learning and prediction are easier and more accurate in tasks like text analysis, time-series forecasting, and cyberbullying detection.

2.5 Transformer-Based Models

Cyberbullying detection is one area where transformers—attention-based deep learning models—find extensive application in NLP. Using self-attention, they are able to extract complicated social media text with long-range dependencies [21]. The most successful method for detecting and minimizing cyberbullying has been advanced, fine-tuned pre-trainer transformer models.

2.5.1 BERT (Bidirectional Encoder Representations from Transformers)

Cyberbullying detection has been greatly improved using the BERT extended language model. It can deduce literary context from input texts [22]. The transformer model can emulate complex associations between words and situations. BERT can identify bullying in social media messages that contain indirect or subtle cues.

2.5.2 Convolutional Neural Networks with Attention (CNN-ATT)

CNN-ATT with attention have also recorded positive outcomes in cyberbullying detection. A CNN also learns the input text in order to extract the most significant features. The attention mechanism then weighs these features using a careful approach that tends to set the relevancy of the classification task. Cyberbullying has been detected using CNNs with Attention by capitalizing on patterns and textual forms in social media messages.

2.5.3 Long Short-Term Memory Networks with Attention (LSTM-ATT)

The readability of attention mechanisms and LSTMs' ability to detect long-term relationships in sequential data make this method potentially applicable to cyberbullying detection [23]. Ideal for text data modeling, RNNs (and LSTM variants) can handle sequential data of varying lengths. LSTM models that incorporate attention mechanisms improve their ability to identify and categorize cyberbullying instances by focusing on the crucial parts of the input text.

3 Rapid And Real-Time Cyberbullying Detection Systems

The meteoric rise of numerous online venues, including websites, social media apps, and others has profoundly altered the transfer of knowledge and communication. However, as every coin has two sides, this digital revolution has also given rise to a very serious issue of cyber bullying. Harassment, intimidation, or disturbance inflicted upon a person using electronic means of communication [24]. This can happen in many ways, such as through insulting comments, threats, or the spread of false information. The ability for cyberbullying to happen at any moment and in any location is what sets it apart from more conventional forms of bullying, and it may cause victims significant mental and emotional suffering over time.

3.1 Applications of AI-Based Cyberbullying Detection

The monitoring platform has many features designed to give parents more power and information about their kids' online activities that could be dangerous. Here are the main characteristics:

3.1.1 Management of Minors' Social Networks to Monitor

This tool lets parents sign up for the platform and choose which kids' social networks and social media accounts they want to keep an eye on. Parents can keep their kids' information up to date by editing it and can add or delete social networks that are being watched as needed [25]. Also, "activating" and "deactivating" networks lets choose whether to include them in analyses without taking them off the site for good.

3.1.2 Analysis and Details of Cyberbullying Alerts

Monitored parents can look through their kids' social networks to find signs of cyberbullying [26][27]. Parents can quickly find possible bullies by seeing how many times their child interacts with people on each social network and a list of users who are causing problems [28][29]. The inappropriate material that triggered each alert can be privately viewed by parents. In addition to the original post where the bullying occurred and the bully's profile, they can also view the uncensored photo or words.

3.1.3 Summary of Alerts and Parental Advice

The software provides parents with tailored recommendations that take into account the latest notifications regarding their child's social media activity [30]. Another, more comprehensible visual representation of a child's alert count over time is a line graph. As a result, you can easily check each child's status in the "Home" area.

4 Literature Review

In the summary of the literature below, the researchers examined social media cyberbullying detection, focusing on AI methods, NLP preprocessing, real-time monitoring, and multimodal content analysis. An overview of the key focus areas, methodologies, main findings, goals, and future research directions is provided in Table 1, which also highlights trends and gaps in the field.

Cirillo *et al.* (2025) Use Prompt-based ML, a novel Machine Learning method, to conduct a large-scale evaluation of twenty LLM generators. Find real-life examples of cyberbullying in online posts. On binary and multiclass classification tasks, they compare 24 ML and NLP models to LLMs using thousands of real posts from X, Facebook, and Reddit. In particular, the goal of this comparison analysis is to learn how well LLMs discriminate against cyberbullying compared to more conventional models, and to use that knowledge to choose appropriate models for spotting malicious material on social media [31].

Sayed, Elnashar and Omara (2025) presents a model that uses a grouping of ML classifiers and NLP approaches to identify instances of cyberbullying. A total of 39,870 tweets and comments were analyzed for this study. The tweets and posts were classified into five categories: cyberbullying based on religion, age, gender, ethnicity, and non-cyberbullying. After processing using NLP approaches, the proposed model aims to train ML classifiers. The following five machine learning classifiers have been used to implement it: RF, KNN, LR, SVM, and NB [32].

Bhardwaj *et al.* (2024) consider and evaluate existing methods, practices, and obstacles for identifying cyberbullying in group chat rooms. It researches the use of NLP, ML, and DL methods for detecting harmful behavior in real-time conversations, analyzing existing literature and methods. The review also mentions the challenges of working with unstructured text data, the use of sentiment analysis, and the necessity of context-concrete models to identify latent types of harassment. The paper underscores the need to come up with more strong, scalable, and adaptive detection systems to establish safer online environments [33].

Teng, Varathan and Crestani (2024) offers a lot of literature that is relevant to the classification of cyberbullying since the past up to date to review it fully. An analysis of 126 articles was conducted. This work covers cyberbullying across text and several media. The entire evaluation was structured around the machine learning workflow and consisted of four key parts: dataset analysis, pre-processing analysis, feature analysis, and technique analysis. The critical analysis provides the groundwork for addressing deficiencies and outlining

directions for future study to fill in the blanks. In addition, the review examined how these strategies could affect ethical considerations. A thorough synopsis of the present state of the art, architecture, and methodology in the field is the goal of this review study [34].

Mathur *et al.* (2023) introduce a system for Twitter that detects cyberbullying in real time using ML and NLP. Evaluate the system's performance using various ML approaches according to its training on a dataset of cyberbullying tweets. Following adjustment, Random Forest was determined to deliver the most favorable outcomes. By using Selenium to scrape tweets from a specific Twitter account and save the timestamps of previously checked tweets, real-time analysis was enabled. In order to further reduce the amount of spam tweets, an image captioning model was used to create descriptions for the account's photographs and compare them to user-written captions [35].

Khan and Qureshi (2022) employing ML and NLP methods to identify instances of cyberbullying in Urdu-language tweets. As far as they are aware, there is no publicly available standard Urdu dataset, so cyberbullying detection in Urdu literature has not been conducted. The authors of this study compiled a database of insulting Urdu remarks sent by Twitter users. The dataset's comments are grouped into five distinct types. Features can be extracted at the character and word levels using n-gram algorithms. Cyberbullying is detected by applying several supervised machine-learning approaches to the dataset [36].

Table 1: Literature on social media cyberbullying detection using AI methods

Authors	Focus Area	Key Findings	Approaches	Objectives	Future Work
Cirillo et al., (2025)	Large-scale evaluation of generative LLMs for cyberbullying detection	LLMs evaluated on binary and multiclass classification show strong performance compared to 24 traditional ML/NLP models; effective for harmful content detection	Prompt-based ML, LLMs, binary & multiclass classification, comparison with ML/NLP models	Assess LLMs' capability in detecting cyberbullying and select suitable models for social media content	Further refine prompt-based methods and adapt LLMs for cross-platform applications
Sayed, Elnashar, & Omara (2025)	Detection of cyberbullying on Twitter across 5 categories	Model effectively categorizes posts into religion, age, gender, ethnicity bullying, and non-cyberbullying	ML classifiers (RF, SVM, LR, NB, KNN), NLP preprocessing	Train ML classifiers on NLP-processed Twitter data to identify cyberbullying types	Extend dataset size, explore DL models for improved accuracy
Bhardwaj et al., (2024)	Cyberbullying detection in group chats	Emphasizes challenges of unstructured text, need for context-aware models; highlights NLP, ML, DL techniques	NLP, ML, DL, sentiment analysis, real-time analysis	Analyze current techniques and challenges to improve detection of harmful behaviors in group chats	Develop scalable, adaptive, and context-sensitive detection systems
Teng, Varathan, & Crestani (2024)	Comprehensive review of cyberbullying classification	Reviewed 126 papers; highlights text-based and multimodal cyberbullying; workflow includes dataset, preprocessing, features, and techniques; ethical concerns addressed	ML workflow: dataset analysis, preprocessing, feature extraction, model analysis	Provide comprehensive review of trends, architectures, and techniques in cyberbullying detection	Address limitations and gaps in current research; explore ethical frameworks
Mathur et al., (2023)	Real-time cyberbullying detection on Twitter	Random Forest achieved best results; integrated image captioning and real-time scraping via Selenium; system helps filter harmful content	NLP, ML (Random Forest, others), real-time tweet scraping, image captioning	Detect and prevent cyberbullying in real-time on Twitter	Expand to multimodal platforms and optimize real-time system efficiency
Khan & Qureshi (2022)	Cyberbullying detection in Urdu Twitter comments	Created first Urdu offensive comments dataset; ML models applied with n-gram features show effective detection	NLP, ML (various supervised techniques), n-gram features	Detect cyberbullying in Urdu text comments and evaluate ML techniques	Increase dataset size, explore deep learning and multilingual models

5 Conclusion and Future Work

Artificial intelligence tools are essential in ensuring safe online spaces, especially when it comes to the most vulnerable groups in society, such as minors who are likely to be the most victims of cyberbullying. This paper underscores the changing definition of cyberbullying, which involves harassment, denigration, exclusion, outing, flaming, and impersonation and the issues associated with cultural diversity, language, and data. Machine learning, deep learning, and transformer-based solutions are examples of AI-based approaches that have shown significant potential for real-time detection and mitigation of cyberbullying. Conventional MLs are simpler to use and have faster training, whereas deep learning and transformer models achieve high performance in recognizing complex contextual and sequential patterns in social media content. Quick detection systems that incorporate data

collection, preprocessing, feature extraction, and classification help identify and intervene in time. AI-based detection systems will be able to assist parents, educators, and social media in tracking and reacting to cyberbullying, which will minimize the psychological and social effects of the phenomenon on its victims.

The future studies can be based on the enhancement of multilingual and cross-cultural detection models, offering multimodal data (text, images, video), and explainable AI frameworks. The improvement of capabilities of real-time and adaptable learning processes will also lead to more resilient AI systems and will allow proactive intervention and more efficient elimination of cyberbullying in various online environments.

Reference

[1] T. K. H. Chan, C. M. K. Cheung, and Z. W. Y. Lee, "Cyberbullying on social networking sites: A literature review and future research directions," Inf. Manag., vol. 58, no. 2, p. 103411, Mar. 2021, doi: 10.1016/j.im.2020.103411.

[2] M. A. Johanis, A. R. A. Bakar, and F. Ismail, "Cyber-Bullying Trends Using Social Media Platform: An Analysis through Malaysian Perspectives," J. Phys. Conf. Ser., vol. 1529, no. 2, p. 022077, Apr. 2020, doi: 10.1088/1742-6596/1529/2/022077.

[3] A. R. Bilipelli, "Forecasting the Evolution of Cyber Attacks in FinTech Using Transformer-Based Time Series Models," Int. J. Res. Anal. Rev., vol. 12, no. 3, pp. 1–7, 2023.

[4] R. Ghosh, M. Malhotra, and N. Kumar, "Cyber Bullying in the Digital Age," 2025, pp. 151–180. doi: 10.4018/979-8-3373-0543-1.ch006.

[5] M. M. Nair, T. F. Fernandez, and A. K. Tyagi, "Cyberbullying in Digital Era: History, Trends, Limitations, Recommended Solutions for Future," in 2023 International Conference on Computer Communication and Informatics (ICCCI), IEEE, Jan. 2023, pp. 1–10. doi: 10.1109/ICCCI56745.2023.10128624.

[6] A. Desai, S. Kalaskar, O. Kumbhar, and R. Dhumal, "Cyber Bullying Detection on Social Media using Machine Learning," ITM Web Conf., 2021, doi: 10.1051/itmconf/20214003038.

[7] R. Dattangire, R. Vaidya, D. Biradar, and A. Joon, "Exploring the Tangible Impact of Artificial Intelligence and Machine Learning: Bridging the Gap between Hype and Reality," in 2024 1st International Conference on Advanced Computing and Emerging Technologies (ACET), IEEE, Aug. 2024, pp. 1–6. doi: 10.1109/ACET61898.2024.10730334.

[8] R. Gupta, A. K. Singh, Utkarsh, P. Mittal, and Radhika, "AI Based Cyberbullying Detection and Prevention," in 2024 3rd Edition of IEEE Delhi Section Flagship Conference (DELCON), IEEE, Nov. 2024, pp. 1–6. doi: 10.1109/DELCON64804.2024.10866680.

[9] Vilas Shewale, "Demystifying the MITRE ATT&CK Framework: A Practical Guide to Threat Modeling," J. Comput. Sci. Technol. Stud., vol. 7, no. 3, pp. 182–186, May 2025, doi: 10.32996/jcsts.2025.7.3.20.

[10] G. Ray, C. D. McDermott, and M. Nicho, "Cyberbullying on Social Media: Definitions, Prevalence, and Impact Challenges," J. Cybersecurity, vol. 10, no. 1, Jan. 2024, doi: 10.1093/cybsec/tyae026.

[11] D. Sultan et al., "A Review of Machine Learning Techniques in Cyberbullying Detection," Comput. Mater. Contin., vol. 74, no. 3, pp. 5625–5640, 2023, doi: 10.32604/cmc.2023.033682.

[12] E. Serritella, A. Guazzini, and E. Menesini, "Countering bullying and cyberbullying using technology-based solutions: a systematic review," Aggress. Violent Behav., vol. 85, p. 102102, Nov. 2025, doi: 10.1016/j.avb.2025.102102.

[13] R. Patel, "Security Challenges In Industrial Communication Networks: A Survey On Ethernet/Ip, Controlnet, And Devicenet," Int. J. Recent Technol. Sci. Manag., vol. 7, no. 8, 2022.

[14] D. Patel, "AI-Enhanced Natural Language Processing for Improving Web Page Classification Accuracy," vol. 4, no. 1, pp. 133–140, 2024, doi: 10.56472/25832646/JETA-V4I1P119.

[15] M. Studies, "A Comprehensive Review of Cyberbullying Detection Techniques and Datasets on Social Media Platforms," Int. J. Innov. Res. Technol., vol. 12, no. 6, pp. 6943–6959, 2025.

[16] F. Akter, M. U. F. Jahangir, M. F. Rabbi, and R. R. Chowdhury, "Cyberbullying Detection on Social Media Platforms Utilizing Different Machine Learning Approaches," Int. J. Comput. Appl., vol. 186, no. 61, pp. 40–50, Jan. 2025, doi: 10.5120/ijca2025924395.

[17] N. Mehendale, K. Rajpara, K. Shah, and C. Phadtare, "A Review on Cyberbullying Detection Using Machine Learning," SSRN Electron. J., 2022, doi: 10.2139/ssrn.4116153.

[18] N. Gs, A. Shenoyy, K. Chaturya, L. Jc, and J. Shree, "Detection of Cyberbullying Using NLP and Machine Learning in Social Networks for Bi-Language," Int. J. Sci. Res. Eng. Trends, vol. 10, no. 1, pp. 2395–566, 2024.

[19] R. Q. Majumder, "Machine Learning for Predictive Analytics: Trends and Future Directions," Int. J. Innov. Sci. Res. Technol., vol. 10, no. 04, pp. 3557–3564, 2025.

[20] B. Joshi, B. K. Joshi, S. Pant, A. Kumar, and H. K. Sharma, "An Efficient Method for Detecting Cyberbullying Using Supervised Machine Learning Techniques," Procedia Comput. Sci., vol. 258, pp. 1254–1261, 2025, doi: 10.1016/j.procs.2025.04.359.

[21] M. T. Hasan, M. A. E. Hossain, M. S. H. Mukta, A. Akter, M. Ahmed, and S. Islam, "A Review on Deep-Learning-Based Cyberbullying Detection," Futur. Internet, vol. 15, no. 5, p. 179, May 2023, doi: 10.3390/fi15050179.

[22] M. Behzadi, I. G. Harris, and A. Derakhshan, "Rapid Cyber-bullying detection method using Compact BERT Models," in 2021 IEEE 15th International Conference on Semantic Computing (ICSC), IEEE, Jan. 2021, pp. 199–202. doi: 10.1109/ICSC50631.2021.00042.

[23] Nirav Kumar Prajapati, "Federated Learning for Privacy-Preserving Cybersecurity: A Review on Secure Threat Detection," Int. J. Adv. Res. Sci. Commun. Technol., vol. 5, no. 4, pp. 520–528, Apr. 2025, doi: 10.48175/IJARSCT-25168.

[24] A. Singh, H. O. S. Mishra, and S. K. Jha, "Filtering Online Harassment: ML based Cyberbullying Detectionz," SAMRIDDHI A J. Phys. Sci. Eng. Technol., vol. 17, no. 01, pp. 1–6, 2025, doi: 10.18090/samriddhi.v17i01.01.

[25] T. Sathyanarayana Rao, D. Bansal, and S. Chandran, “Cyberbullying: A virtual offense with real consequences,” Indian J. Psychiatry, vol. 60, no. 1, p. 3, 2018, doi: 10.4103/psychiatry.IndianJPsychiatry_147_18.

[26] M. C. Bularca, S. Cristescu, A. Netedu, and C. Coman, “Analyzing the cyberbullying phenomenon on social media from the perspective of students,” Front. Psychol., vol. 15, no. December, Dec. 2024, doi: 10.3389/fpsyg.2024.1458079.

[27] E. A. Nina-Gutiérrez, J. E. Pacheco-Alanya, and J. C. Morales-Arevalo, “SocialBullyAlert: A Web Application for Cyberbullying Detection on Minors’ Social Media,” Int. J. Adv. Comput. Sci. Appl., vol. 15, no. 7, pp. 769–778, 2024, doi: 10.14569/IJACSA.2024.0150776.

[28] Mr. Pradeep Nayak, C H Rakesh, Chandana A S, Chandana P T, and Darshan S, “Review Paper on Cyberbulling,” Int. J. Adv. Res. Sci. Commun. Technol., pp. 495–503, Jan. 2023, doi: 10.48175/IJARSCT-8002.

[29] S. Narang and V. G. Kolla, “Next-Generation Cloud Security: A Review of the Constraints and Strategies in Serverless Computing,” Int. J. Res. Anal. Rev., vol. 12, no. 3, pp. 1–7, 2025, doi: 10.56975/ijrar.v12i3.319048.

[30] M. Raj, S. Singh, K. Solanki, and R. Selvanambi, “An Application to Detect Cyberbullying Using Machine Learning and Deep Learning Techniques,” SN Comput. Sci., vol. 3, no. 5, p. 401, Jul. 2022, doi: 10.1007/s42979-022-01308-5.

[31] S. Cirillo, D. Desiato, G. Polese, G. Solimando, V. Sugumaran, and S. Sundaramurthy, “Exploring the ability of emerging large language models to detect cyberbullying in social posts through new prompt-based classification approaches,” Inf. Process. Manag., vol. 62, no. 3, p. 104043, May 2025, doi: 10.1016/j.ipm.2024.104043.

[32] F. R. Sayed, E. H. Elnashar, and F. A. Omara, “Cyberbullying detection in social media using natural language processing,” Sci. African, vol. 28, p. e02713, Jun. 2025, doi: 10.1016/j.sciaf.2025.e02713.

[33] R. Bhardwaj, A. Billade, S. Chenna, A. Deshpande, and V. Arthamwar, “Cyberbullying Detection in Group Chat Applications- A Review,” in 2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS), IEEE, Dec. 2024, pp. 210–214. doi: 10.1109/ICICNIS64247.2024.10823118.

[34] T. H. Teng, K. D. Varathan, and F. Crestani, “A comprehensive review of cyberbullying-related content classification in online social media,” Expert Syst. Appl., vol. 244, p. 122644, Jun. 2024, doi: 10.1016/j.eswa.2023.122644.

[35] S. A. Mathur, S. Isarka, B. Dharmasivam, and J. C. D., “Analysis of Tweets for Cyberbullying Detection,” in 2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC), IEEE, May 2023, pp. 269–274. doi: 10.1109/ICSCCC58608.2023.10176416.

[36] S. Khan and A. Qureshi, “Cyberbullying Detection in Urdu Language Using Machine Learning,” in 2022 International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE), IEEE, Dec. 2022, pp. 1–6. doi: 10.1109/ETECTE55893.2022.10007379.