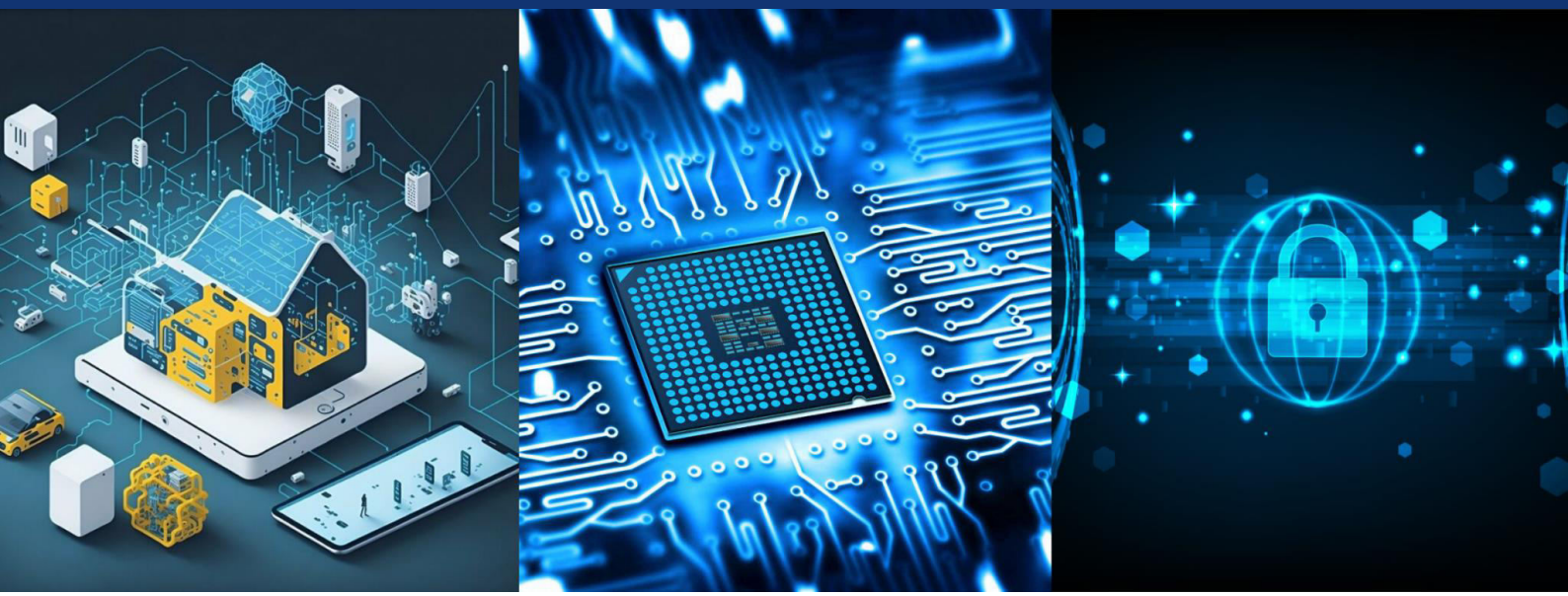
IJIRCCE

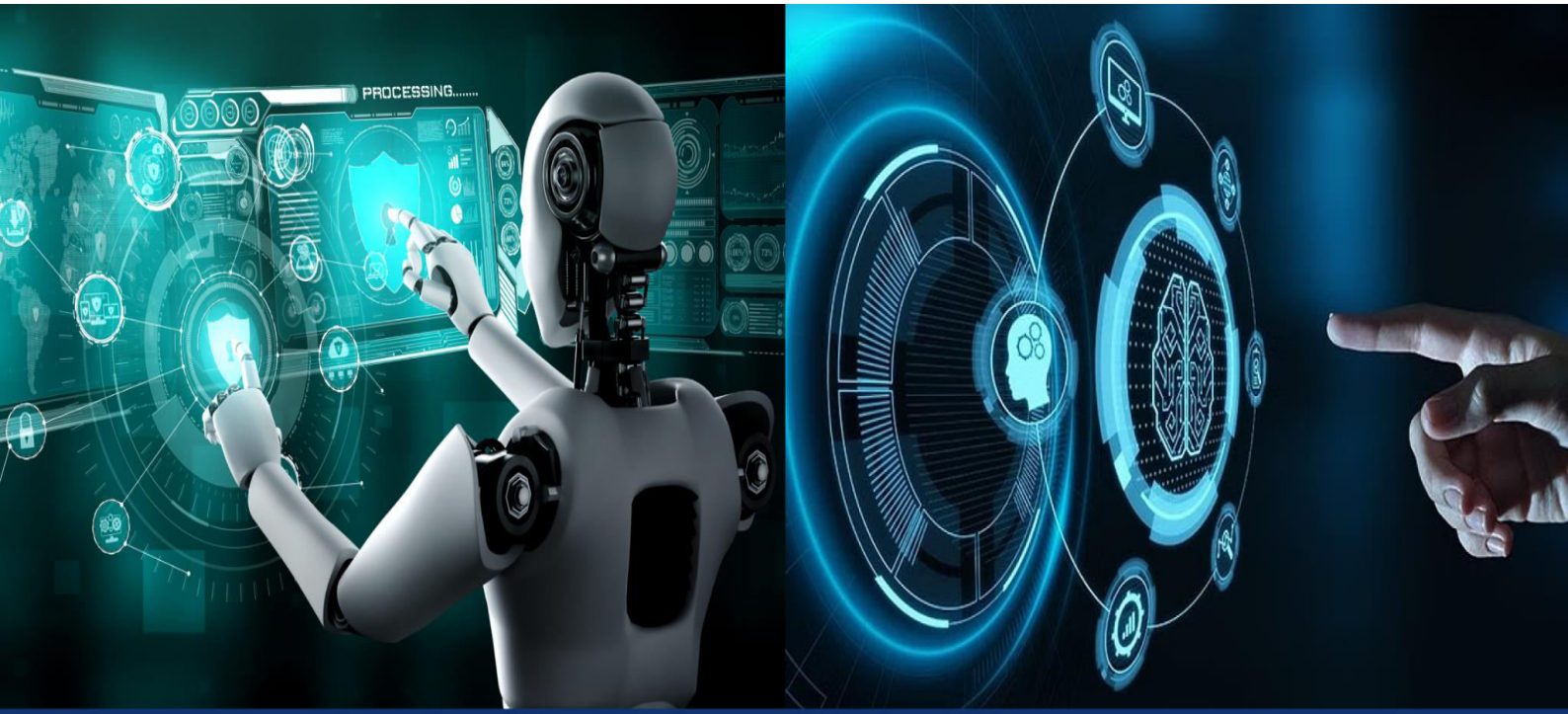
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 14, Issue 4, April 2026

© 2026 IJIRCCE | Volume 14, Issue 4, April 2026 DOI: 10.15680/IJIRCCE.2026.1404082

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

To Implement Cyberbullying Detection on Social Networks Using Machine Learning

Dr.Naveen Bilandi. P[1], Noorjahan. D[2], Prithika.V[2], Ruba Dharshini. D.M[2]

Associate Professor, Department of Computer Science and Engineering, P.S.V College of Engineering and Technology, Mittapalli, Krishnagiri, India[1]

UG Scholar, Department of Computer Science and Engineering, P.S.V College of Engineering and Technology, Mittapalli, Krishnagiri, India[2]

ABSTRACT: Cyberbullying has emerged as a serious social issue with the rapid growth of social media and online communication platforms. It negatively impacts users' mental health, emotional well-being, and overall online safety. Traditional cyberbullying detection systems mainly rely on keyword matching or basic sentiment analysis, which often fail to understand context, sarcasm, and non-textual content such as images and GIFs. The proposed system considers five major feature categories, including linguistic, contextual, behavioural, visual, and metadata features, to accurately identify cyberbullying content. It supports detection from text posts, images, and GIFs, enabling comprehensive analysis of modern social media interactions. To achieve this, the system will be developed using the Python programming language and the Django framework, ensuring a robust and secure backend architecture

KEYWORDS: Cyberbullying Detection, Django Web Application, Machine Learning, Deep Learning, Natural Language Processing, Text Classification, Offensive Language Detection, Convolutional Neural Networks, Image Classification, GIF Analysis, Optical Character Recognition, Easy OCR, Frame Extraction, Pattern Recognition, Multi-format Content Analysis, Real-time.

Domain: MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE.

I. INTRODUCTION

Cyberbullying is a growing problem in social networking platforms that affects the safety and well-being of users. It refers to the use of digital communication to harass, threaten, or humiliate individuals. With the increasing use of social media, the amount of user-generated content has also increased. This makes manual monitoring difficult and time-consuming. To overcome this problem, automated systems using machine learning are required. These systems can analyze large amounts of data and detect harmful content efficiently. In this project, a cyberbullying detection system is developed using machine learning techniques. The system analyzes both text and images to identify abusive content. This helps in improving the accuracy of detection. The main aim is to provide a safe and secure online environment.

II. LITERATURE REVIEW

"Cyberbullying Detection using Pre-Trained BERT Model" This work uses a pre-trained BERT model to detect cyberbullying in textual data. BERT is a transformer-based model that understands the context of words by analyzing the entire sentence. This helps in identifying the actual meaning of user comments, including complex language patterns. The model improves classification accuracy compared to traditional machine learning methods. However, it requires high computational power and memory, which makes it less suitable for real-time applications.

"ProTect: a hybrid deep learning model" The ProTect model proposes a hybrid approach by combining Random Forest and Convolutional Neural Networks for cyberbullying detection. Random Forest is used for feature selection, while CNN extracts deep features from the data. This combination improves detection accuracy and performance. The model can handle both simple and complex patterns effectively. However, the architecture becomes complex and requires more time for training and implementation.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

"The TLA-NET Approach via Trustable LSTM-Autoencoder Networks" The TLA-NET approach uses LSTM combined with an autoencoder for detecting cyberbullying. LSTM helps in understanding sequential data and capturing context in sentences. The autoencoder reduces noise and extracts important features from the data. This improves the reliability and performance of the model. However, the system requires a large dataset and has high computational complexity, which increases training time.

"Multi-Guard: Cross-Modal Attention for Meme-Based Bullying" The Multi-Guard model focuses on detecting cyberbullying in memes by analyzing both text and images. It uses a cross-modal attention mechanism to understand the relationship between different data types. This approach improves detection accuracy, especially for multimedia content. It is useful for modern social media platforms where memes are widely used. However, the model requires large datasets and high processing power, making it complex to implement.

III. METHODOLOGY

A. EXISTING SYSTEM

he existing systems for cyberbullying detection mainly rely on analyzing textual data using basic machine learning and deep learning techniques. These systems classify user-generated content such as comments and messages into bullying or non-bullying categories. Most approaches use algorithms like Support Vector Machines, Naïve Bayes, and LSTM models for detection. However, they primarily focus only on text-based analysis and do not consider multimedia content such as images and memes. In modern social media platforms, users often express harmful content through images and combined formats, which these systems fail to detect effectively. As a result, the performance of existing systems is limited in real-time applications.

B. DISADVANTAGES

1. Existing systems mainly focus only on textual data for cyberbullying detection.They fail to analyze images, memes, and other multimedia content effectively.
2. These systems are not able to understand context, sarcasm, or hidden meanings properly. This leads to incorrect classification and reduces overall accuracy.
3. Many models require large datasets and high computational resources.This makes them less efficient for real-time applications.
4. The systems are not scalable for handling large volumes of social media data.They also depend on manual monitoring in some cases.

C. PROPOSED SYSTEM

The proposed system introduces a machine learning-based approach for detecting cyberbullying using a multimodal technique. It analyzes both textual and visual data from social media platforms to improve detection accuracy. Natural Language Processing techniques are used for processing and understanding text data. For image analysis, deep learning models such as Convolutional Neural Networks are used to extract important features. The system combines the outputs from both text and image models using a fusion technique to make a final decision. It classifies the content as bullying or non-bullying and generates alerts for harmful content. The system is designed to support real-time detection and improve overall performance.

D. ADVANTAGES

1. The proposed system uses a multimodal approach by analyzing both text and images.This
2. improves the accuracy of cyberbullying detection significantly.
3. It supports real-time detection of harmful content on social media platforms.This helps in taking quick action and improving user safety.
4. The system reduces manual effort by automatically identifying abusive content.It saves time and increases efficiency in monitoring.
5. It is scalable and can handle large amounts of data effectively.The model performance improves continuously with training data.

E. DESIGN OF THE SYSTEM

The system is designed using a modular architecture to efficiently detect cyberbullying from social media content. It begins with the user interface, where users can register, log in, and post content in the form of text, images, or GIFs.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The input data is sent to the backend server, which acts as an API gateway and handles all user requests. The data is then forwarded to the data collection module, where it is stored for further processing and analysis.

In the next stage, the preprocessing module cleans and prepares the data by removing noise, special characters, and unnecessary information. Text data is processed using techniques such as tokenization and stopword removal, while images and GIFs are normalized and converted into suitable formats. The processed data is then passed to the feature extraction module, where important features are extracted using Natural Language Processing techniques for text and deep learning models for images.

The extracted features are given to the classification module, where machine learning and deep learning algorithms are applied to detect cyberbullying. The system uses a fusion mechanism to combine the outputs from both text and image analysis for better accuracy. Based on the final result, the alert and moderation module identifies harmful content and generates alerts. The system also includes a reporting and analytics module to provide insights into user

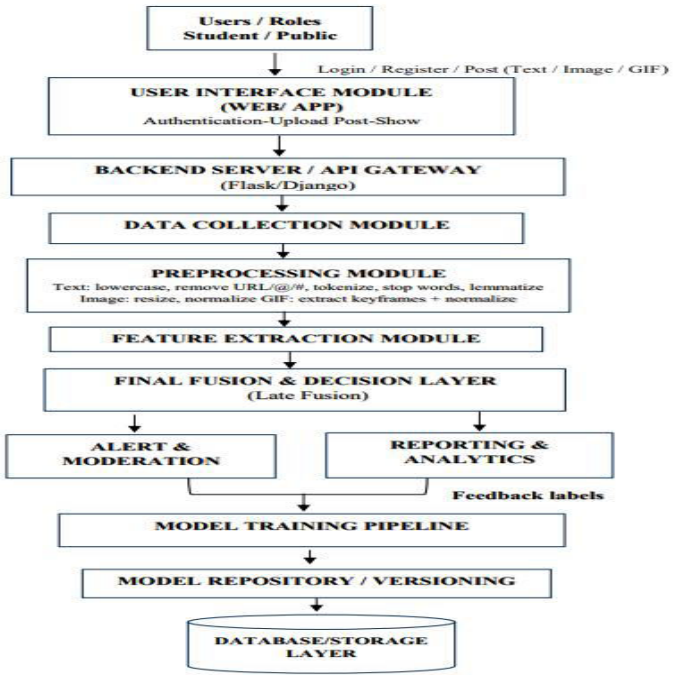


Fig.No.1 System Architecture

behavior. Finally, all data and trained models are stored in the database, and the model training pipeline updates the system for improved performance over time.

Fig. 1 shows the architecture of the proposed cyberbullying detection system, which processes user-generated content such as text and images using machine learning techniques to identify and classify harmful content. The system includes modules for data collection, preprocessing, feature extraction, classification, and alert generation to ensure accurate and efficient detection. The extracted features are analyzed using machine learning and deep learning models to improve prediction accuracy. The final output is used to generate alerts and maintain a safer online environment.

IV. IMPLEMENTATION

MODULE DESCRIPTION

1. USER INTERFACE MODULE

This work uses a pre-trained BERT model to detect cyberbullying in textual data. BERT is a transformer-based model that understands the context of words by analyzing the entire sentence. This helps in identifying the actual meaning of

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

user comments, including complex language patterns. The model improves classification accuracy compared to traditional machine learning methods. However, it requires high computational power and memory, which makes it less suitable for real-time applications.

2. DATA COLLECTION MODULE

The ProTect model proposes a hybrid approach by combining Random Forest and Convolutional Neural Networks for cyberbullying detection. Random Forest is used for feature selection, while CNN extracts deep features from the data. This combination improves detection accuracy and performance. The model can handle both simple and complex patterns effectively. However, the architecture becomes complex and requires more time for training and implementation.

3. PREPROCESSING MODULE

The TLA-NET approach uses LSTM combined with an autoencoder for detecting cyberbullying. LSTM helps in understanding sequential data and capturing context in sentences. The autoencoder reduces noise and extracts important features from the data. This improves the reliability and performance of the model. However, the system requires a large dataset and has high computational complexity, which increases training time.

4. FEATURE EXTRACTION MODULE (5 CATEGORIES)

The Multi-Guard model focuses on detecting cyberbullying in memes by analyzing both text and images. It uses a cross-modal attention mechanism to understand the relationship between different data types. This approach improves detection accuracy, especially for multimedia content. It is useful for modern social media platforms where memes are widely used. However, the model requires large datasets and high processing power, making it complex to implement

5. MODEL INFERENCE MODULE

The Model Inference Module is responsible for making predictions based on the trained machine learning model. It takes the processed input data and applies the trained model to classify the content as bullying or non-bullying. This module ensures fast and accurate decision-making during real-time usage. It plays a key role in applying the learned patterns from training to new data. However, its performance depends on the quality of the trained model and input data.

6. ALERT & MODERATION MODULE

The Alert and Moderation Module is used to identify harmful content and take appropriate actions. When cyberbullying content is detected, the system generates alerts to notify users or administrators. It can also flag or block harmful posts to prevent further spread. This module helps in maintaining a safe online environment. However, incorrect predictions may lead to false alerts or missed detections.

7. REPORTING & ANALYTICS MODULE

The Reporting and Analytics Module provides insights based on the detected cyberbullying data. It generates reports that help in understanding user behavior and the frequency of harmful content. This information can be used to improve system performance and decision-making. It also helps administrators monitor overall activity. However, it requires proper data management and storage to function effectively.

V. RESULT

proposed cyberbullying detection system was successfully implemented using a web application interface. The system is capable of analyzing different types of user inputs such as text, images, and GIFs to detect harmful content. It provides accurate classification results along with confidence levels. The results show that the system effectively identifies cyberbullying content and ensures better moderation. The user-friendly interface allows easy interaction and quick analysis of uploaded data. Overall, the system demonstrates efficient performance in detecting and classifying harmful content.

© 2026 IJIRCCE | Volume 14, Issue 4, April 2026 DOI: 10.15680/IJIRCCE.2026.1404082

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Fig No.2 Image Cyberbullying Detector

Fig No.3 Text Cyberbullying Detection

© 2026 IJIRCCE | Volume 14, Issue 4, April 2026 DOI: 10.15680/IJIRCCE.2026.1404082

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

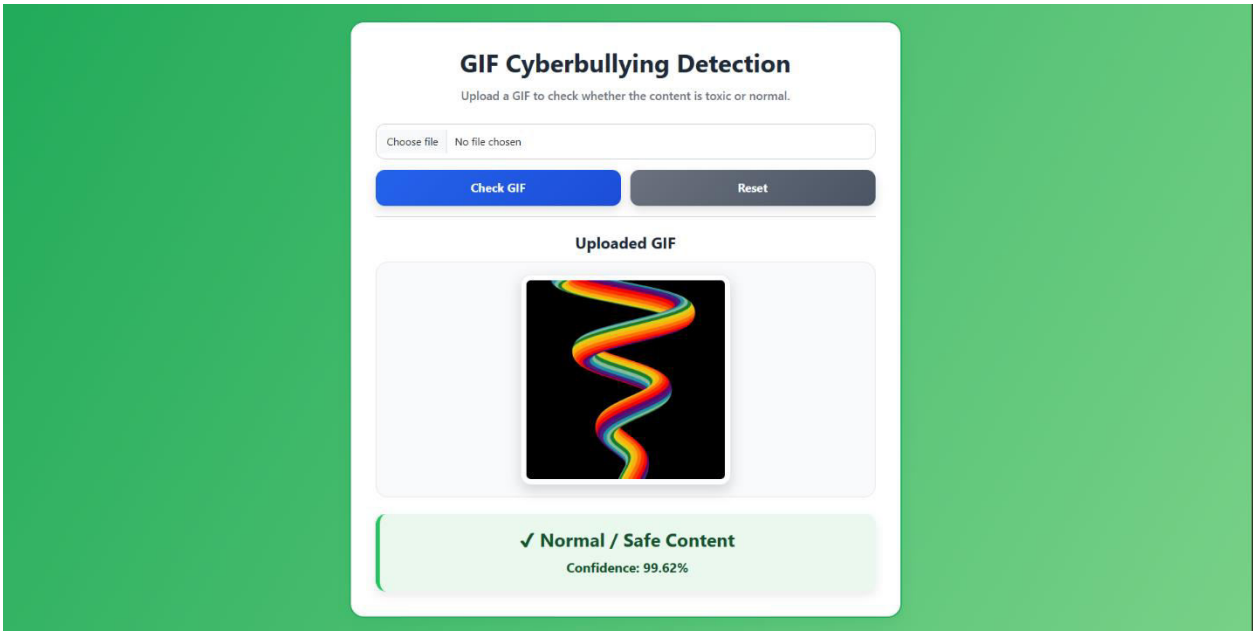


Fig No.4 GIF Cyberbullying Detection

Fig. 2 shows the image cyberbullying detection interface of the system. The user uploads an image, and the system analyzes it using deep learning techniques. The uploaded image is processed, and the model predicts whether the content is harmful or safe. The interface displays the selected image along with the prediction result. This module helps in detecting offensive visual content effectively. Fig. 3 shows the GIF cyberbullying detection module. The user uploads a GIF file, and the system extracts frames and analyzes them using trained models. The result is displayed along with a confidence score indicating whether the content is safe or harmful. This module is useful for detecting bullying content in animated media formats. Fig. 4 shows the overall web application interface of the cyberbullying detection system. It provides options for text, image, and GIF detection through a simple and user-friendly design. Users can easily upload content and view results instantly. The interface ensures smooth navigation and improves user experience.

VI. CONCLUSION

The proposed cyberbullying detection system provides an effective solution for identifying harmful content on social media platforms. It uses machine learning and deep learning techniques to analyze both textual and visual data efficiently. By adopting a multimodal approach, the system improves detection accuracy compared to traditional methods. It can classify content as bullying or non-bullying and generate alerts for harmful behavior. The system reduces manual effort and supports real-time monitoring of user-generated content. The results demonstrate that the system performs reliably and helps in controlling cyberbullying. Overall, it contributes to creating a safer and more secure online environment for users.

VII. FUTURE ENHANCEMENT

The system can be further enhanced by integrating advanced deep learning models such as transformer-based architectures for better performance. It can be extended to support multiple languages to increase usability across different regions. Real-time integration with popular social media platforms can improve its practical implementation. Additional features like user behavior analysis and sentiment tracking can enhance prediction accuracy. Increasing the dataset size and improving data quality can further strengthen the model. Future improvements can also focus on reducing computational complexity and improving system speed. The system can also be enhanced with automatic content filtering and blocking mechanisms to prevent the spread of harmful content. These enhancements will make the system more robust, scalable, and efficient.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

REFERENCES

[1] Yadav, J., Kumar, D., & Chauhan, D. "Cyberbullying Detection using Pre-Trained BERT Model," 2020.
[2] Chen, S., Wang, J., & He, K. "Chinese Cyberbullying Detection Using XLNet and Deep Bi-LSTM Hybrid Model,"2024.
[3] Cuzzocrea, A., Akter, M. S., Shahriar, H., & García Bringas, P. "Cyberbullying Detection, Prevention, and Analysis on Social Media via Trustable LSTM-Autoencoder Networks over Synthetic Data," 2025.
[4] Hasan, M. T., Hossain, M. A. E., Mukta, M. S. H., Akter, A., Ahmed, M., & Islam, S. A Review on Deep-Learning-Based Cyberbullying Detection," 2023.
[5] Nitya Harshitha, T., Prabu, M., Suganya, E., Sountharrajan, S., Bavirisetti, D. P., Gadde, N., & Uppu, L. S. "ProTect: a hybrid deep learning model for proactive detection of cyberbullying on social media," 2024.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details