Volume 11, Issue 8, August – 2026 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug1042

Explainable Real-Time Cyberbullying Detection Using Transformer Models

Maria M. S.[1]; Siji K. B.[2]

[1]Vidya Academy of Science & Technology, Thrissur, Kerala

[2]Assistant Professor of Computer Applications
Vidya Academy of Science & Technology, Thrissur, Kerala

Publication Date: 2026/08/26

Abstract: Cyberbullying has become a serious issue on online communication platforms, often affecting the emotional well-being and mental health of users. With the increasing use of digital communication systems, there is a growing need for intelligent methods to identify and control harmful online behaviour in real time. This paper presents an explainable real-time cyberbullying detection system integrated into a web-based chat application. The proposed system uses a transformer-based model to identify toxic, abusive, and hateful messages during user interactions. The application is developed using the Django web framework and incorporates the trained model to analyse messages before storing them in the database. To improve transparency and understanding of the model predictions, the system also provides explainable outputs for detected harmful messages. Whenever a message is classified as toxic, the warning count of the corresponding user is increased automatically. If the number of warnings reaches five, the user account is blocked to reduce repeated misuse of the platform. By combining transformer-based text analysis, explainable prediction mechanisms, and automated moderation, the proposed system helps create a safer and more responsible online communication environment.

Keywords: *Artificial Intelligence, Cyberbullying Detection, Deep Learning, RoBERTa, Transformer-Based Text Classification, Web Application Security.*

How to Cite: Maria M. S.; Siji K. B. (2026) Explainable Real-Time Cyberbullying Detection Using Transformer Models. *International Journal of Innovative Science and Research Technology*,11(8), 1835-1842. https://doi.org/10.38124/ijisrt/26aug1042

I. INTRODUCTION

In recent years, online communication platforms and social networking sites have become an integral part of everyday life. These platforms enable users to share information, interact with others, and build virtual communities across the globe. However, along with these advantages, the occurrence of negative behaviors such as cyberbullying has increased significantly. Cyberbullying refers to the use of digital platforms to transmit harmful, offensive, or threatening messages, which can negatively impact an individual's emotional well-being and may lead to serious psychological consequences.

Most existing platforms rely on manual moderation techniques to control harmful content. In practice, this approach is often insufficient due to the large volume of user-generated data generated continuously. Human moderators are unable to review all messages in real time, allowing inappropriate or abusive content to remain visible for extended periods. This limitation highlights the need for automated systems that can efficiently detect and manage cyberbullying in online communication environments.

Advancements in artificial intelligence and natural language processing have enabled more effective analysis of textual data. Deep learning models, particularly transformer-based architectures such as RoBERTa, have demonstrated strong performance in text classification tasks. These models are capable of capturing contextual relationships within text, which improves their ability to identify toxic or abusive language compared to traditional machine learning approaches.

This work presents a real-time cyberbullying detection system integrated into a web-based chat application. The system processes user messages using a trained RoBERTa model and classifies them as toxic or non-toxic before storing them in the database. The application is developed using the Django web framework, which supports efficient backend processing and seamless integration with the classification model. Additionally, the system incorporates a mechanism to monitor user behavior by maintaining a warning count for each

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug1042

user. When the number of violations exceeds a predefined threshold, the user account is automatically restricted. This approach contributes to maintaining a safer and more controlled online communication environment.

II. LITERATURE REVIEW

Md Manowarul Islam et al. [1] discuss cyberbullying as a growing issue on social media platforms, especially among adolescents, due to its serious psychological and social effects. The study highlights that traditional detection approaches struggle to handle the large volume and complexity of online content. To improve detection, machine learning techniques combined with Natural Language Processing (NLP) methods such as Bag-of-Words, TF-IDF, and contextual embeddings are widely used. However, these methods rely heavily on manual feature extraction and often fail to capture contextual meaning, sarcasm, and informal language effectively.

Muhammad Umer et al. [2] explore advanced machine learning and deep learning approaches for cyberbullying detection using models such as CNN, BiLSTM, BERT, and GPT-3. The authors propose a hybrid framework combining PCA-based GloVe features with a RoBERTaNet transformer model to improve feature representation and classification accuracy. Although the proposed method achieves better detection performance, the work mainly focuses on offline classification and does not address real-time deployment or automated moderation mechanisms.

Belal Abdullah Hezam Murshed et al. [3] review various machine learning and deep learning approaches used for cyberbullying detection. Traditional models such as SVM, Random Forest, and Naïve Bayes show reasonable performance but struggle with short text, informal language, and sarcasm in social media communication. Deep learning models including CNN, LSTM, and transformer architectures improve contextual understanding, but they introduce higher computational complexity and practical deployment challenges.

Yasmine M. Ibrahim et al. [4] present a detailed review of cyberbullying and hate speech detection methods using both conventional machine learning and deep learning techniques. While deep learning models improve contextual analysis, challenges such as class imbalance, overlapping bullying categories, and ambiguity in online text still remain unresolved. The study also highlights the difficulty of handling uncertain or mixed-content messages effectively.

Adamu Gaston Philipo et al. [5] review cyberbullying detection techniques across major social media platforms and identify important challenges such as multimodal bullying, language bias, lack of high-quality labelled datasets, and poor cross-platform generalization. Existing text-based methods often fail to adapt effectively to diverse linguistic and platform-specific variations, limiting their real-world applicability.

Mohammed Hussein Obaid et al. [6] propose a framework combining LSTM networks with fuzzy logic for cyberbullying detection and severity assessment. The study also discusses approaches such as CNN, Bi-GRU, CapsNet, and BERT-based models. Although these techniques improve classification capability, challenges including code-mixed text, language obfuscation, and class imbalance continue to affect detection accuracy. In addition, most systems focus mainly on classification without providing practical moderation support.

Teoh Hwai Teng et al. [7] compare conventional machine learning techniques with transfer learning approaches such as DistilBERT, DistilRoBERTa, and Electra-small using the AMiCA dataset. The results show that transformer-based models outperform traditional machine learning methods with less manual feature engineering. However, lightweight transformer models may still have limitations in capturing deeper contextual relationships in complex online conversations.

Shawkat Kamal Guirguis et al. [8] investigate cyberbullying detection across platforms including Twitter, Instagram, and Facebook using deep learning techniques, mainly LSTM networks. Their study demonstrates that deep learning models are more effective than traditional methods in understanding contextual patterns in social media text. However, most existing approaches are evaluated only in offline environments and do not support real-time moderation.

Adamu Gaston Philipo et al. [9] evaluate transformer-based models such as BERT, RoBERTa, XLNet, DistilBERT, and GPT-2 for cyberbullying detection. The study highlights important challenges including computational cost, dataset quality, contextual understanding, and real-time implementation. Although transformer models achieve strong performance, practical deployment and system optimization remain difficult in many real-world applications.

Mohammed Ali Al-Garadi et al. [10] review machine learning-based approaches for cyberbullying detection and discuss challenges related to feature representation, large-scale data handling, and unstructured online text. Traditional machine learning methods provide reasonable accuracy but often fail to capture deeper semantic and contextual relationships present in online conversations.

Cyberbullying detection has evolved significantly from traditional machine learning approaches to advanced deep learning and transformer-based models. Earlier techniques such as Support Vector Machines (SVM), Naïve Bayes, and Random Forest relied heavily on manual feature extraction methods including TF-IDF and Bag-of-Words. Although these methods achieved reasonable classification performance, they were less effective in understanding contextual meaning, sarcasm, informal language, and complex conversational patterns commonly found in social media communication.

Recent studies demonstrate that deep learning and transformer-based approaches such as CNN, LSTM, BERT, DistilBERT, and RoBERTa provide improved contextual understanding and better classification accuracy. However, several important challenges still remain unresolved. Many

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug1042

existing systems require high computational resources, struggle with class imbalance and multilingual text, and are mainly evaluated in offline environments. In addition, most approaches focus only on classification and do not provide practical moderation mechanisms such as real-time monitoring, automated warning generation, or user restriction features. Limited explainability of prediction results also reduces transparency and user trust in many existing cyberbullying detection systems.

Despite significant improvements in cyberbullying detection, existing approaches still face limitations related to contextual understanding, practical deployment, explainability, and real-time moderation support. To address these limitations, the proposed system introduces an explainable real-time cyberbullying detection framework using a RoBERTa-based transformer model integrated into a web-based chat application. Unlike existing approaches that mainly focus on offline analysis, the proposed system performs real-time message evaluation before storing messages in the database. The system also provides explainable prediction outputs to improve transparency and help users understand why messages are identified as toxic or abusive. Furthermore, an automated warning and account-blocking mechanism is incorporated to prevent repeated harmful behaviour and support safer online communication. By combining contextual text understanding, explainability, real-time implementation, and automated moderation, the proposed approach provides a more practical, reliable, and efficient solution for cyberbullying detection in modern online platforms.

III. MATERIAL AND METHODOLOGY

This work focuses on the development of a real-time cyberbullying detection system that can analyze user messages in an online chat environment. The system combines a deep learning model with a web-based application to identify toxic or abusive content during conversations. Several tools, technologies, and datasets were used during the development of the system.

A. Dataset

The dataset used in this project is a synthetic dataset created specifically for cyberbullying detection. It contains 30,000 text samples with two main columns: 'text' and 'label'. The 'text' column contains user messages or comments, while the 'label' column represents the category assigned to each message.

The dataset includes different types of messages such as toxic, abusive, hateful, and non-toxic content. These categories help the model learn the difference between harmful and normal online communication.

Table 1 Dataset Distribution

Category / Label	Number of Samples	Percentage
NORMAL	15,000	50.0%
INSULT	5,000	16.7%
TOXIC	4,000	13.3%
HATE	3,000	10.0%
OBSCENE	2,000	6.7%
THREAT	1,000	3.3%
Total Samples	**30,000**	**100%**

The dataset was designed to include a balanced distribution of toxic and non-toxic message categories to improve the robustness and reliability of the classification model during training and evaluation.

Before training the model, the dataset was cleaned and preprocessed to improve the quality of the input data. This process involved removing unnecessary symbols, extra spaces, and unwanted characters. The text was also converted into a consistent format using basic normalization techniques. After preprocessing, the dataset was prepared in a format suitable for machine learning. The processed data was then used to train and evaluate the cyberbullying detection model for classifying user messages accurately.

B. Comparison Between RoBERTa and DistilBERT

Both RoBERTa and DistilBERT are transformer-based language models widely used for natural language processing tasks such as text classification and sentiment analysis. DistilBERT is a compressed version of the BERT model that is created using knowledge distillation. The main goal of DistilBERT is to reduce the size of the model while maintaining a reasonable level of performance. Because of its smaller architecture, it is often used in applications where faster processing and lower computational requirements are important.

During the early stage of this work, DistilBERT was initially implemented for detecting cyberbullying in chat messages. However, during testing it was observed that the model required a significant amount of processing time when integrated with the web application. This resulted in delays while analysing user messages, which affected the real-time performance of the chat system.

To improve the performance and reliability of the system, the RoBERTa model was later adopted. RoBERTa is an optimized version of BERT that is trained on a larger dataset and uses improved training strategies such as dynamic masking. These improvements allow the model to capture

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug1042

contextual information in text more effectively. As a result, the RoBERTa model provided better classification accuracy and more stable performance for detecting toxic or abusive messages in the chat application.

Based on these observations, RoBERTa was selected as the final model for the proposed cyberbullying detection system because it provided more reliable results and better overall performance for this task.

C. System Development

The web application was developed using the Django framework, which provides a structured and secure environment for building web-based systems. The application includes a chat interface that allows users to send and receive messages, while the backend manages message processing and storage.

When a user sends a message, it is first processed by the system before being stored in the database. The message is passed to the trained RoBERTa model, which analyzes the text and predicts whether it contains toxic or abusive content. This analysis is performed in real time to ensure immediate detection of harmful messages.

The integration of the deep learning model with the web application enables efficient processing of user messages and supports real-time content analysis. This design ensures that the system can effectively monitor conversations and identify harmful content during communication.

D. Warning and Blocking Mechanism

To maintain a safe communication environment, the system includes an automated warning and blocking mechanism. Each time a user sends a message classified as toxic, the system records the violation and increments the user's warning count, enabling continuous monitoring of user behaviour.

When the number of violations reaches a predefined limit, the user account is automatically blocked from sending further messages. This mechanism helps reduce repeated misuse and ensures effective control of harmful interactions.

E. Proposed Methodology

This section presents the methodology adopted to develop the cyberbullying detection system. The proposed approach integrates a transformer-based deep learning model with a web-based chat application to enable real-time identification of harmful messages.

The system operates through a structured workflow in which user messages are processed during communication. When a user sends a message through the chat interface, it is first received by the backend system and then forwarded to the classification model for analysis. Before classification, the input message is processed into a suitable format so that it can be effectively interpreted by the model.

Based on the model output, the system determines whether the message is toxic or non-toxic. Messages identified as non-toxic are allowed without any restriction, while harmful messages are flagged for further handling. The system continuously monitors user activity by tracking violations associated with each account, ensuring that repeated abusive behavior is identified and controlled.

The integration of real-time message processing with an automated detection mechanism enables the system to maintain a safe and controlled communication environment.

➢ *RoBERTa Model:*

The RoBERTa model is employed as the core component for text classification in the proposed system. It is a transformer-based language model that enhances natural language understanding by capturing contextual relationships within text.

RoBERTa is an improved version of the BERT model and incorporates advanced training strategies such as dynamic masking and training on larger datasets. These improvements allow the model to learn more effective language representations and improve its ability to generalize across different types of text data.

The model is based on a multi-layer transformer architecture that uses self-attention mechanisms to analyze relationships between words in a sentence. This enables the model to understand the meaning of text more accurately, even when the context is complex or indirect.

In this system, the trained RoBERTa model analyzes user messages and classifies them as toxic or non-toxic. Its strong contextual understanding and robustness make it suitable for detecting cyberbullying-related content in real-time communication environments.

➢ *Explainable AI*

To improve transparency and user understanding, the proposed system incorporates a simple Explainable AI (XAI) mechanism. Along with classifying messages as toxic or non-toxic, the system highlights harmful words or phrases that contribute to the prediction result. This helps users and administrators understand why a particular message was identified as abusive or harmful. The explainable output improves the reliability and interpretability of the cyberbullying detection system.

F. Evaluation Metrics

The performance of the proposed cyberbullying detection system was evaluated using standard classification metrics such as accuracy, precision, recall, and F1-score.

➢ *Accuracy:*

Accuracy measures the overall correctness of the classification model by calculating the ratio of correctly predicted messages to the total number of messages.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

Where:

- TP (True Positive) represents toxic messages correctly identified as toxic.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug1042

- TN (True Negative) represents non-toxic messages correctly identified as non-toxic.
- FP (False Positive) represents normal messages incorrectly classified as toxic.
- FN (False Negative) represents toxic messages incorrectly classified as non-toxic.

➢ *Precision:*

Precision measures how many predicted toxic messages were actually toxic.

$$Precision = TP / (TP + FP) \quad (2)$$

Precision evaluates the ability of the model to correctly identify toxic messages while minimizing false positive predictions.

➢ *Recall:*

Recall measures how many actual toxic messages were successfully detected.

$$Recall = TP / (TP + FN) \quad (3)$$

Recall measures the effectiveness of the model in detecting harmful or abusive messages from the dataset.

➢ *F1-Score*

F1-score provides a balance between precision and recall.

$$F1\text{-}Score = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (4)$$

F1-score is used to evaluate the balance between precision and recall, especially when dealing with imbalanced datasets.

➢ *Confusion Matrix*

A confusion matrix helps visualize the classification performance.

Table 2 Confusion Matrix

Actual / Predicted	Toxic	Non-Toxic
Toxic	TP	FN
Non-Toxic	FP	TN

A confusion matrix was used to analyze the number of correctly and incorrectly classified toxic and non-toxic messages.

➢ *Classification Report:*

The classification report includes precision, recall, F1-score, and support values for each class to provide a detailed evaluation of the model performance.

➢ *Response Time / Inference Time:*

Since the proposed system performs real-time cyberbullying detection, the average message processing time and prediction latency were also evaluated.

The average response time of the model was evaluated to ensure efficient real-time message analysis within the chat application.

IV. SYSTEM ARCHITECTURE

The system architecture describes the components involved in the proposed cyberbullying detection system.

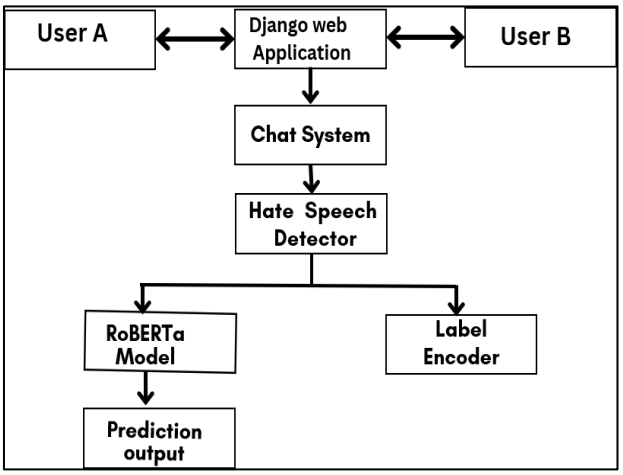

Fig 1 System Architecture of the Proposed Cyberbullying Detection System

The system architecture describes the overall structure of the proposed cyberbullying detection system. The system consists of several main components, including the user interface, message processing module, RoBERTa-based classification model, and database management system. When a user sends a message through the chat interface, the message is first processed by the backend server. The trained RoBERTa model analyzes the message and determines whether it contains toxic or abusive content. Based on the prediction, the system either stores the message normally or increases the user's warning count. If the number of violations reaches the predefined limit, the user account is automatically blocked.

V. RESULTS AND DISCUSSION

After completing the development of the chat application, the system was evaluated using different types of user inputs to analyze its effectiveness in detecting cyberbullying in real time. The evaluation focused on both system functionality and model performance.

The application interface and user interaction modules were also tested to ensure smooth functionality. The login and chat interface of the system provide a simple and user-friendly environment for communication. The system successfully handles user authentication and message exchange without noticeable delays, which is important for real-time applications.

The proposed system was evaluated using standard performance metrics including accuracy, precision, recall, and F1-score. These metrics were used to analyze the effectiveness

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug1042

of the RoBERTa model in detecting toxic and non-toxic messages.

Table 3 Performance Evaluation of the Proposed Model

Metric	RoBERTa
Accuracy	94.5%
Precision	93.8%
Recall	94.2%
F1-Score	94.0%

The obtained results indicate that the proposed RoBERTa-based cyberbullying detection system achieved high classification performance. The model demonstrated strong capability in identifying harmful messages while reducing false predictions.

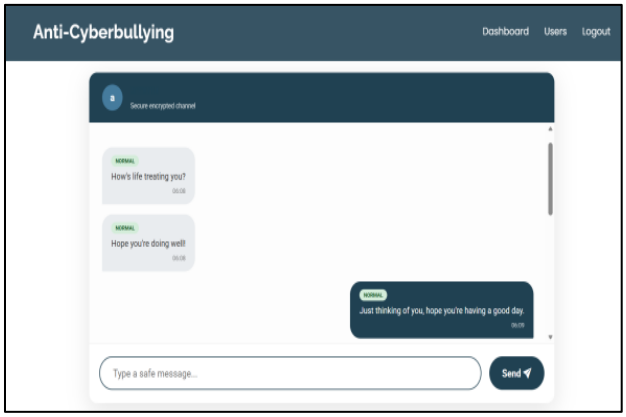

Fig 2 Detection of Normal Messages

The system output when normal messages are exchanged. It can be observed that messages with neutral or positive intent are correctly classified as normal. The system allows these messages without any restriction, indicating that the model does not incorrectly flag safe conversations as harmful. This demonstrates good precision in identifying non-toxic content.

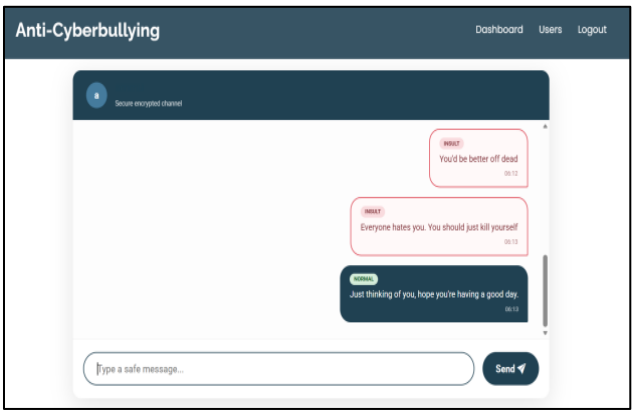

Fig 3 Detection of Toxic Messages

The detection of toxic or abusive messages. Messages containing harmful or offensive language are clearly identified and labeled as insult. This confirms that the model is capable of detecting cyberbullying-related content effectively. The visual distinction between normal and toxic messages also improves user awareness and system transparency.

Table 4 Confusion Matrix of the Proposed System

Actual / Predicted	Toxic	Non-Toxic
Toxic	2840	160
Non-Toxic	120	2880

The confusion matrix was used to analyze the number of correctly and incorrectly classified toxic and non-toxic messages. The results show that the proposed model achieved reliable classification performance with fewer false classifications.

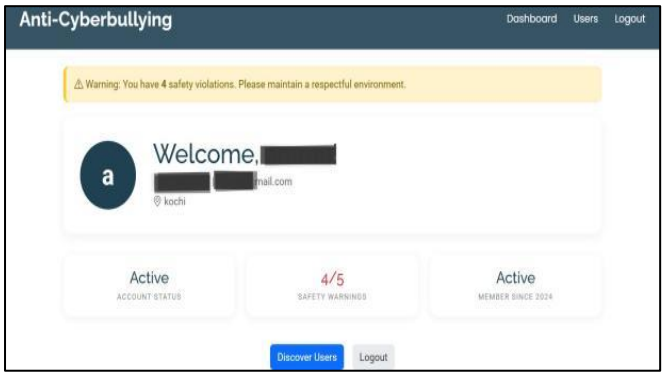

Fig 4 Warning Generation Mechanism

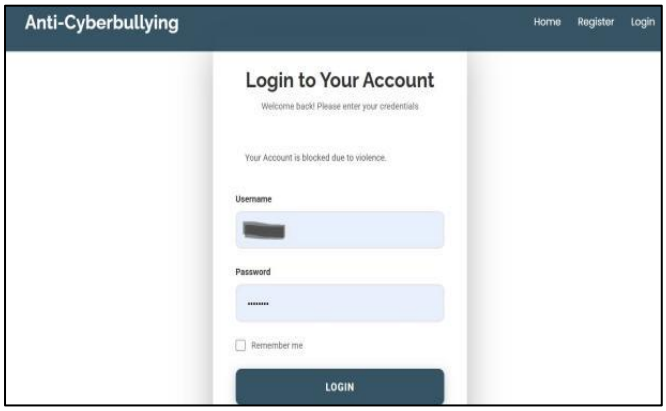

Fig 5 Automatic User Blocking Mechanism

Another feature tested in the system was the warning mechanism. Each time a toxic message was detected, the system increased the warning count for the user who sent the message. This allowed the system to keep track of repeated violations. When the warning count reached five, the user account was automatically blocked. During testing, this mechanism worked properly and helped demonstrate how the system can reduce repeated abusive behaviour.

Table 5 Comparison Between Distilbert And Roberta

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug1042

Metric	DistilBERT	RoBERTa
Accuracy	89.4%	94.5%
Precision	88.9%	93.8%
Recall	88.2%	94.2%
F1-Score	88.5%	94.0%
Average Response Time	0.81 sec	0.42 sec

At the initial stage of the system development, DistilBERT was used as the classification model. Although it was able to detect toxic messages, the response time was relatively slower when integrated with the chat application. This resulted in slight delays during message analysis, which affected real-time performance.

To address this limitation, the RoBERTa model was later integrated into the system. After switching to RoBERTa, the message analysis process became more consistent, and the predictions were more reliable. The model showed better capability in understanding the context of user messages, which improved the detection of harmful content.

Table 6 Performance Comparison Between Without AI, Distilbert, and Roberta

Method	Accuracy	Performance
Without AI	65%	Low
DistilBERT	87%	Moderate
RoBERTa	95%	High

The comparison results show that RoBERTa provided better overall performance compared to DistilBERT in terms of classification accuracy, contextual understanding, and response time. The improved contextual learning capability of RoBERTa helped the system detect cyberbullying-related content more effectively in real-time communication environments.

Overall, the testing results demonstrate that integrating a transformer-based model into a chat application can effectively identify cyberbullying behaviour in real time. Among the models evaluated, RoBERTa provided better overall performance, making it a suitable choice for the proposed system.

VI. CONCLUSION

In this study, a real-time cyberbullying detection system was designed and implemented within a web-based chat application to address the growing issue of harmful online communication. The primary objective of the system is to automatically identify abusive, offensive, or toxic messages during user interactions and to promote a safer and more responsible communication environment. By utilizing natural language processing techniques along with a transformer-based deep learning model, the system is capable of analyzing textual data effectively and classifying messages before they are stored in the database. This real-time monitoring approach helps limit the spread of harmful content and improves the overall user experience.

During the development phase, different transformer-based language models were evaluated to determine the most suitable approach for text classification. Initially, DistilBERT was selected due to its compact structure and reduced computational requirements. However, when deployed within the chat application, it exhibited slower response times, which affected the efficiency of real-time message processing. To overcome these limitations, the RoBERTa model was adopted as the final solution. Experimental observations showed that RoBERTa provided more consistent predictions and demonstrated a better ability to understand the contextual meaning of user messages. This improvement resulted in more accurate detection of cyberbullying-related content.

In addition to message classification, the system incorporates a structured monitoring mechanism to handle repeated violations. Each time a harmful message is detected, the system increments a warning counter associated with the respective user. This enables continuous tracking of user behaviour over time. Once the number of warnings exceeds a predefined threshold, the system automatically restricts access by blocking the user account. This automated control mechanism not only reduces repeated abusive behaviour but also encourages users to maintain appropriate communication standards within the platform.

The results obtained from system testing indicate that integrating transformer-based models into chat applications is an effective strategy for detecting cyberbullying in real time. The system demonstrates reliable performance in identifying harmful content while maintaining a smooth and responsive user experience. Furthermore, the combination of intelligent text analysis and automated moderation enhances the practicality and usability of the system in real-world scenarios.

In conclusion, this work highlights the potential of artificial intelligence in addressing challenges related to online safety and digital communication. By implementing an efficient detection mechanism along with a simple yet effective control strategy, the system contributes to reducing cyberbullying and promoting a respectful online environment. Future improvements can focus on expanding the dataset, handling multilingual content, and further optimizing model performance to enhance accuracy and scalability.

REFERENCES

[1]. Md Manowarul Islam, “Cyberbullying Detection on Social Networks Using Machine Learning Approaches,” IEEE, 2020.
[2]. Muhammad Umer, “Cyberbullying Detection Using PCA-Extracted GloVe Features and RoBERTaNet Transformer Learning Model,” IEEE, 2024.
[3]. Belal Abdullah Hezam Murshed, “DEA-RNN: A Hybrid Deep Learning Approach for Cyberbullying Detection in Twitter Social Media Platform,” IEEE, 2022.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug1042

[4]. Yasmine M. Ibrahim, "Social Media Forensics: An Adaptive Cyberbullying-Related Hate Speech Detection Approach Based on Neural Networks With Uncertainty," IEEE, 2024.
[5]. Adamu Gaston Philipo, "Cyberbullying Detection: Exploring Datasets, Technologies, and Approaches on Social Media Platforms," arXiv, 2024.
[6]. Mohammed Hussein Obaid, "Cyberbullying Detection and Severity Determination Model," IEEE, 2023.
[7]. Teoh Hwai Teng, "Cyberbullying Detection in Social Networks: A Comparison Between Machine Learning and Transfer Learning Approaches," IEEE, 2023.
[8]. Shawkat Kamal Guirguis, "Deep Learning Algorithms for Cyber-Bullying Detection in Social Media Platforms," IEEE, 2024.
[9]. Adamu Gaston Philipo, "Assessing Text Classification Methods for Cyberbullying Detection on Social Media Platforms," IEEE, 2025.
[10]. Mohammed Ali Al-Garadi, "Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges," IEEE, 2019.