

AI-DRIVEN LEGAL AND EMOTIONAL SUPPORT SYSTEM FOR CYBER HARASSMENT VICTIMS

Machine Learning

Mrs. M. Ramya	Assistant Professor, Dept. of Artificial Intelligence & Data Science, Jansons Institute of Technology, Coimbatore, India.
Maha Mugesh Ananth. P	UG Student, Dept. of Artificial Intelligence & Data Science, Jansons Institute of Technology, Coimbatore, India.
Vamsi Mohan. G	UG Student, Dept. of Artificial Intelligence & Data Science, Jansons Institute of Technology, Coimbatore, India.
Sriram. S	UG Student, Dept. of Artificial Intelligence & Data Science, Jansons Institute of Technology, Coimbatore, India.
Sujith. M	UG Student, Dept. of Artificial Intelligence & Data Science, Jansons Institute of Technology, Coimbatore, India.

ABSTRACT

Cyber harassment has increased rapidly with the growth of online platforms and social media, causing significant emotional and legal challenges for victims. Many victims lack awareness of applicable cyber laws and reporting procedures, while fear of exposure further discourages timely action. Existing platforms primarily provide reporting mechanisms without offering personalised legal guidance or emotional reassurance. This paper presents AI-Driven Legal and Emotional Support System For Cyber Harassment Victims. The system accepts incident descriptions via text or voice, classifies the harassment type using a BERT-based machine learning model, and maps the incident to applicable IPC and IT Act provisions. It auto-generates an FIR draft, provides step-by-step legal guidance, and delivers empathetic emotional support through the Gemini API — all within a privacy-preserving, anonymous interaction architecture. Experimental evaluation on real user inputs demonstrates 79.7% classification confidence, correct identification of IPC 503, 506, 507 and IT Act 66 and 66D provisions, and contextually appropriate emotional support generation.

KEYWORDS

Cyber Harassment, NLP, Legal Guidance, Emotional Support, BERT, Machine Learning, FIR Assistance, Privacy-Preserving AI.

I. INTRODUCTION

The rapid expansion of social media platforms, messaging applications, and online communication tools has significantly transformed how individuals interact in the digital world. While these technologies provide tremendous benefits in terms of connectivity and information sharing, they have simultaneously created new avenues for abuse and harassment. Cyber harassment — encompassing online threats, cyberstalking, impersonation, and repeated unwanted contact — has emerged as a serious societal concern affecting millions of individuals globally.

According to the World Health Organization (WHO), approximately 280 million people worldwide suffer from depression, with a considerable proportion of mental health deterioration among younger demographics linked to online abuse and cyberbullying. In India, the National Crime Records Bureau (NCRB) reports thousands of cybercrime complaints annually, yet the actual number of incidents is believed to be far higher due to widespread underreporting. Victims frequently experience fear, anxiety, and confusion, and many hesitate to take legal action due to a lack of awareness about applicable laws and complaint procedures.

This paper presents Now I Will Speak AI, an AI-driven platform that bridges this gap by integrating DistilBERT-based harassment classification, jurisdiction-aware legal mapping, automated FIR draft generation, Gemini API-powered empathetic response, and geolocation-based complaint portal redirection — all within a privacy-preserving web application built on React and Flask.

The remainder of this paper outlines the related work, system architecture, proposed methodology, implementation details, and experimental evaluation of the proposed framework, highlighting its practical applicability for cyber harassment victims.

II. System Design and Methodology

A. System Architecture

The Now I Will Speak AI system is a multi-layer intelligent web application. The frontend is built in React, providing a secure and responsive interface where users describe their incident via text or voice, upload optional digital evidence, select jurisdiction, and toggle anonymous mode. The backend, implemented in Python using Flask, hosts four core AI modules: a DistilBERT-based harassment classifier,

an emotion detection engine, a jurisdiction-aware legal mapping module, and a Gemini API integration for response generation.

Upon receiving user input, the system preprocesses and classifies the incident type, detects the user's emotional state, maps the harassment to applicable IPC and IT Act sections, and generates a structured response including legal guidance, an auto-drafted FIR, official complaint portal links, and an empathetic emotional support message. All outputs are presented on a secure user dashboard. When anonymous mode is active, no personal identity data is stored at any stage of processing.

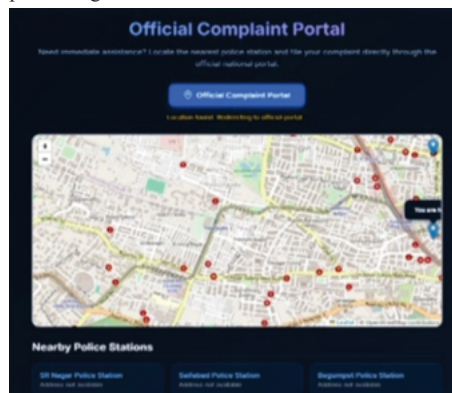


Fig. 1. Home Page — Now I Will Speak AI Platform Overview

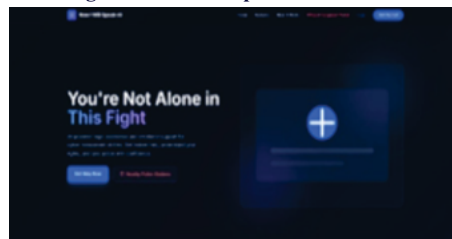


Fig. 2. Simple 3-Step Process: Report Incident → AI Analysis → Take Action

B. Software and Hardware Requirements

Software Stack: Python (backend processing), Flask (REST API framework), React (frontend UI), SQLite (user session database), Hugging Face Transformers (BERT model training and inference), Gemini API (legal guidance and emotional response generation), and dotenv for secure environment configuration.

Hardware Requirements: A standard computer or laptop with a stable internet connection and a modern web browser are sufficient to run and access the platform. No specialised hardware is required for end-user interaction.

C. Proposed Methodology

The system follows a structured seven-step pipeline from user input collection through to output delivery. In Step 1, the user submits an incident description via the web interface or voice-to-text input, optionally uploading evidence such as screenshots or chat logs, and selecting platform and jurisdiction. Step 2 involves text preprocessing — cleaning, lowercasing, and tokenisation using the BERT tokenizer — to prepare the input for classification.

In Step 3, the fine-tuned BERT model classifies the harassment type into one of four categories: Threats, Cyberstalking, Impersonation, or Online Abuse. Step 4 applies emotion detection to identify the user's primary emotional state (fear, anger, sadness, or neutral). Steps 5 and 6 generate jurisdiction-specific legal guidance and evidence collection advice respectively, while Step 7 delivers empathetic emotional support and reporting recommendations through the Gemini API.

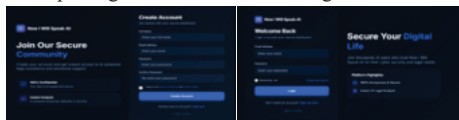


Fig. 3. Login Page (left) and Secure Registration Page (right)

D. Incident Reporting and Anonymous Mode

Once authenticated or using anonymous mode, users describe their incident through a secure form. Anonymous Mode is enabled by default, ensuring no personal information is stored alongside the incident. The platform field captures where the harassment occurred (e.g., Instagram, WhatsApp, email), directly informing jurisdiction-specific legal mapping. Voice-to-text input expands accessibility for users who prefer spoken description.



Fig. 4. Incident Reporting Form with Anonymous Mode Enabled

E. Implementation

The backend is implemented in Python using Flask. At startup, the application initialises four core AI modules: HarassmentClassifier, EmotionDetector, LegalMapper, and GeminiHelper. An SQLite database manages user authentication securely with bcrypt password hashing. The main application entry point, module imports, and database initialisation are shown below.

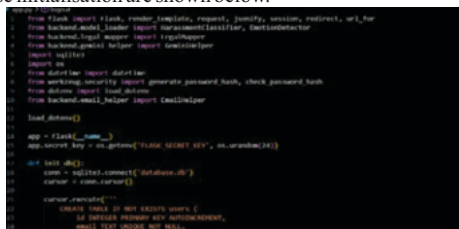


Fig. 5. Flask Application — Entry Point, Module Imports, and DB Initialisation (app.py)

The model training pipeline uses DistilBERT for sequence classification via the Hugging Face Transformers library, with PyTorch as the deep learning backend and scikit-learn for evaluation metrics including accuracy, F1-score, and confusion matrix computation.

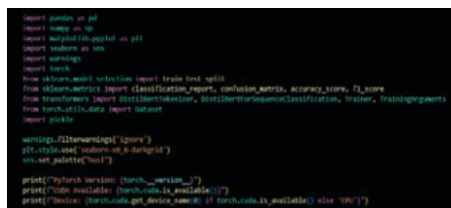


Fig. 6. Model Training — Library Imports: DistilBERT, PyTorch, and Scikit-learn

The login and signup routes handle user authentication against the SQLite database. On successful login, session variables store the user's email and anonymous flag. The signup route hashes passwords using werkzeug's generate_password_hash before database insertion. The core /analyze route receives the incident text and anonymous mode flag, sequentially invokes the harassment classifier, emotion detector, legal mapper, and Gemini helper, then stores and returns the full structured analysis result as JSON. The /results route renders the output page, and /complaint-portal provides the geolocation-enabled police station map.

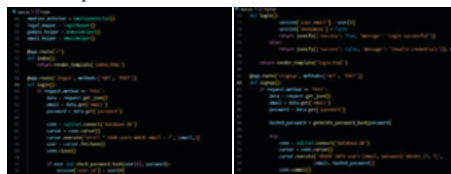


Fig. 7. Login Route (left) and Signup Route (right) — Session Management (app.py lines 58–95)

This module handles user authentication by implementing secure login and signup functionalities. User credentials are validated using database queries, and passwords are protected through hashing techniques to ensure security. The system manages user sessions to maintain login states across different pages, allowing seamless navigation. It also includes error handling for invalid login attempts and ensures that duplicate email registrations are prevented, thereby maintaining data integrity and reliability.



Fig. 8. Analysis Route — Classifier, Emotion Detection, and Legal Mapping (app.py lines 115–123)

The analysis route processes user-submitted incident text through a POST request, enabling the system to evaluate potential harassment. It utilizes a trained classifier to detect abusive content and integrates an emotion detection model to understand the psychological state of the victim. Additionally, the system supports an anonymous mode to protect user identity. Based on the analysis, relevant legal provisions are mapped using a legal mapping module, and the results are stored in the session for further use and display.

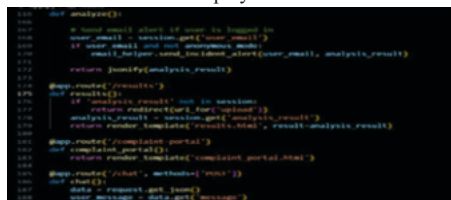


Fig. 9. Results and Complaint Portal Routes and Session Handling (app.py lines 174–188)

The /chat route passes the stored analysis result as context to the Gemini API, enabling context-aware follow-up conversation. Users can ask further questions about their legal options or coping strategies without re-submitting their incident.

```
114 def analyze():
115     # Send email alert if user is logged in
116     user_email = session.get('user_email')
117     if user_email and not anonymous_mode:
118         email_helper.send_incident_alert(user_email, analysis_result)
119
120     return jsonify(analysis_result)
121
122 @app.route('/results')
123 def results():
124     if 'analysis_result' not in session:
125         return redirect(url_for('upload'))
126     analysis_result = session.get('analysis_result')
127     return render_template('results.html', result=analysis_result)
128
129 @app.route('/complaint_portal')
130 def complaint_portal():
131     return render_template('complaint_portal.html')
132
133 @app.route('/chat', methods=['POST'])
134 def chat():
135     data = request.get_json()
136     user_message = data.get('message')
```

F. Official Complaint Portal

The platform integrates a geolocation-enabled Official Complaint Portal that displays nearby police stations on an interactive Leaflet map. Users receive direct links to the National Cyber Crime Reporting Portal (cybercrime.gov.in), Local Police Station, and Cyber Cell — removing the burden of independently searching for reporting resources at a moment of distress.



Fig. 11. Official Complaint Portal with Geolocation-Based Nearby Police Station Map

III. RESULTS AND DISCUSSION

The proposed Now I Will Speak AI system was implemented and evaluated on real user inputs to validate the accuracy of harassment classification, legal guidance generation, and emotional support delivery. Results across all core functional modules are presented below.

A. Harassment Classification and Emotional Analysis:

When a user submitted the incident description "I have been receiving threatening messages from someone on Instagram," the DistilBERT classification model identified the harassment type as Threats with 79.7% confidence. Simultaneously, the emotion analysis module detected the primary emotional state as Neutral, with Sadness at 16.0% and Anger at 15.0%, enabling a contextually targeted emotional support response from a single user input.

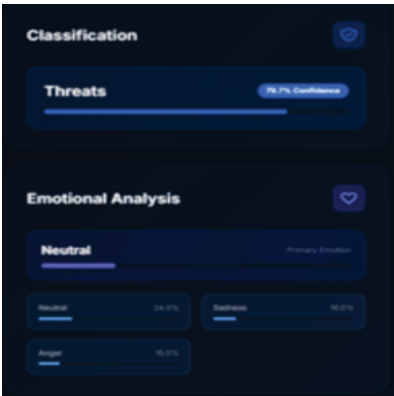


Fig. 12. Classification Result: Threats (79.7% Confidence) and Emotional Analysis Output

B. Legal Framework and Applicable Laws

Based on the classified harassment type and the specified platform (Instagram), the legal mapping module identified applicable laws: IPC 503, IPC 506, IPC 507, IT Act Section 66, and IT Act Section 66D, categorising the offence as a Cognizable Offense. The system generated direct access links to the National Cyber Crime Portal, Local Police Station, and Cyber Cell — delivering precise, jurisdiction-specific legal guidance in real time.

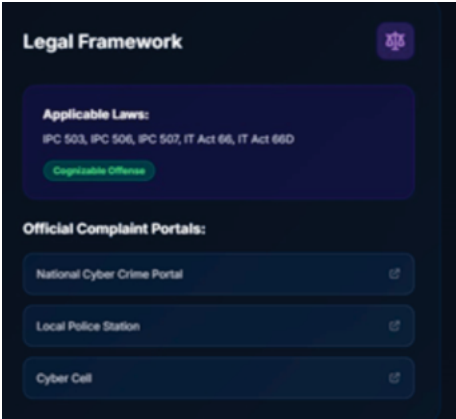


Fig. 13. Legal Framework Output — Applicable IPC and IT Act Sections with Complaint Portal Links

C. Emotional Support Response

The emotional support module generated a personalised, empathetic response acknowledging the user's experience of receiving threatening messages on Instagram. The response validated the user's feelings, provided immediate practical advice — including blocking the account and preserving screenshots as digital evidence — and encouraged the victim to seek support from trusted individuals or law enforcement if they felt unsafe.

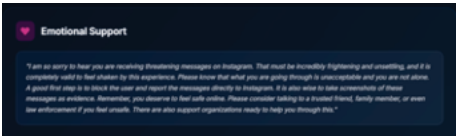


Fig. 14. Emotional Support Response Generated by the System

D. Summary of Outcomes

- Accurate identification of harassment type (Threats) with 79.7% model confidence
- Correct mapping to IPC 503, 506, 507 and IT Act Sections 66 and 66D
- Contextually appropriate empathetic response aligned with detected emotional state
- Geolocation-based nearby police station display and complaint portal redirection
- Fully functional anonymous mode — zero persistent identity storage confirmed
- Auto-generated FIR draft and step-by-step legal reporting guidance delivered

IV. DISCUSSION

The development and evaluation of Now I Will Speak AI demonstrates the effectiveness of combining NLP-based classification, jurisdiction-aware legal mapping, and emotion-aware response generation within a unified, privacy-preserving platform for cyber harassment victims. A key strength of the proposed system is its integrated approach. Unlike existing tools that address reporting or emotional support in isolation, this system delivers both simultaneously within a single interaction. Victims receive clear legal guidance and empathetic responses without navigating multiple fragmented platforms at a moment of psychological distress.

The emotional support module successfully delivered a contextually sensitive response based on the detected emotional state. By acknowledging distress, offering immediate practical steps, and encouraging further support-seeking, the system addresses a critical dimension of victim care that is entirely absent from existing government cybercrime reporting portals. The Anonymous Mode feature directly addresses the privacy concerns and fear of exposure that are leading causes of underreporting. By confirming zero

persistent identity storage, the platform provides a genuinely safe environment for victims who would otherwise remain silent. Limitations include the dependency of legal guidance quality on jurisdiction mapping accuracy, and edge cases involving overlapping legal provisions that may require further refinement. The emotion detection module currently relies on text-based sentiment analysis and could benefit from multimodal input in future iterations. Large-scale deployment would additionally require enhanced cybersecurity measures and integration with existing government portal infrastructure.

Overall, the proposed framework contributes a scalable, accessible, and intelligent approach to cyber harassment support, advancing the goal of safer and more equitable digital environments for all users.

V. CONCLUSION

The presented Document i.e., AI-Driven Legal and Emotional Support System For Cyber Harassment Victims that addresses the growing challenge of cyber harassment by providing victims with integrated legal awareness and emotional support through a unified, intelligent, and privacy-preserving platform.

The rapid increase in cyber harassment incidents, combined with widespread lack of awareness about applicable cyber laws and reporting procedures, has created urgent demand for accessible victim support systems. The proposed solution addresses this need by combining DistilBERT-based harassment classification, jurisdiction-aware legal mapping to IPC and IT Act provisions, automated FIR draft generation, empathetic NLP-driven emotional response via the Gemini API, and geolocation-based complaint portal redirection—all within a two-tier React and Flask architecture.

Experimental evaluation on real user inputs demonstrated accurate classification of harassment types, correct identification of applicable legal provisions, appropriate emotional support response generation, and reliable anonymous mode operation. The system successfully classified a threatening messages incident with 79.7% confidence and delivered both precise legal guidance and personalised emotional support from a single user submission.

Compared to existing platforms that address only reporting mechanisms or only emotional support in isolation, Now I Will Speak AI represents a significant step toward a comprehensive, victim-centred digital support solution. By empowering victims with knowledge, legal clarity, and emotional reassurance, the system encourages timely reporting and appropriate legal action, ultimately contributing to stronger digital governance, justice, and a safer online environment for all.

VI. Related Work

Title: Deep Learning Algorithms for Cyber-Bullying Detection in Social Media Platforms

Authors: M. H. Obaida, S. M. Elkaffas, S. K. Guirguis

Description: This paper proposes deep learning-based techniques for detecting cyberbullying in social media and provides a strong technical foundation for harassment classification. Its automated identification of abusive content directly supports the machine learning classification module used in this project.

Title: A Study of Cyberbullying Detection and Classification Techniques

Authors: S. Unnava, S. R. Parasana

Publication: Engineering, Technology & Applied Science Research, Vol. 14, 2024

Description: This study surveys machine learning approaches to cyberbullying detection, examining classification algorithms and their effectiveness. The benchmarks provided support the design of the harassment classification model in the proposed system.

Title: Detection of Cyberbullying Using Machine Learning and Deep Learning

Authors: Mrs. Deepana M.E., M. Vigneh, R. Palanisamy, I. Solomon Francis

Publication: IJERT, Vol. 12, Issue 03, 2023

Description: This paper compares traditional ML and deep learning methods for cyberbullying detection. The comparative analysis provides useful accuracy benchmarks for the DistilBERT-based classification module integrated in the proposed AI assistant.

Therefore, this study builds upon existing cyber harassment detection, legal informatics, and emotion-aware NLP research by developing an

intelligent, scalable, and privacy-preserving system providing unified legal and emotional support to cyber harassment victims.

VII. REFERENCES

1. M. H. Obaida, S. M. Elkaffas, S. K. Guirguis, "Deep Learning Algorithms for Cyber-Bullying Detection in Social Media Platforms," IEEE Access, Vol. 12, pp. 76901–76908, 2024.
2. S. Unnava, S. R. Parasana, "A Study of Cyberbullying Detection and Classification Techniques: A Machine Learning Approach," Engineering, Technology & Applied Science Research, Vol. 14, Issue 4, 2024.
3. Mrs. Deepana M.E., M. Vigneh, R. Palanisamy, I. Solomon Francis, "Detection of Cyberbullying Using Machine Learning and Deep Learning," IJERT, Vol. 12, Issue 03, 2023.
4. H. Han, M. Asif, E. M. Awwad, N. Sarhan, B. Xu, "Innovative Deep Learning Techniques for Monitoring Aggressive Behavior in Social Media Posts," Journal of Cloud Computing, Springer, 2024.
5. Ministry of Home Affairs, Government of India, Cyber Crime Reporting Portal, <https://cybercrime.gov.in>
6. Information Technology Act, 2000 (India)—Cyber Laws and Amendments.
7. National Crime Records Bureau (NCRB), Cyber Crime Statistics and Reports, India. Google DeepMind, Gemini API Documentation, 2024.
8. Pallets Projects, Flask Documentation — Backend Framework, 2024.
9. Meta Open Source, React Documentation — Frontend Framework, 2024.