CSE7101- Capstone Project
Review-1

Secure AI Chat System (End-to-End Encryption + AI Moderation)

Batch Number: COM_58

Roll Number :
20221COM0136

Name:
R. Charu Swathi Sree

Under the Supervision of,

Dr.Ruhin Kouser R
Assistant Professor – Senior Scale
School of Computer Science and Engineering
Presidency University

Name of the Program: B.tech. Computer Engineering (AI&ML)
Name of the HoD: Dr.Pallavi R
Name of the Program Project Coordinator: Dr.Debasmita Mishra
Name of the School Project Coordinators: Dr. Sampath A K , Dr. Geetha A

Problem Statement Number: PSCS_275

- Organization:

  Presidency University, School of Engineering, Department of Computer Science and Engineering.

- Category (Hardware / Software / Both) :

  Software

- Problem Description:

- With the rapid growth of digital communication platforms, issues such as **data privacy breaches, cyberbullying, hate speech, spam, and phishing attacks** have increased significantly.
  Although popular chat applications like WhatsApp and Instagram use **end-to-end encryption**, they lack effective **real-time content moderation**, allowing harmful messages to reach users.
- There is a need for a secure communication system that ensures:
- **Complete message privacy**
- **Proactive prevention of harmful content**
- **Real-time communication without server-side data exposure**
- This project addresses this gap by combining **end-to-end encryption with AI-based moderation before encryption**.

Content

- Problem Statement
- Objectives
- Background and Related work for title Selection
- Analysis of Problem Statement
- Innovation or Novel Contributions
- Git-hub Link
- Timeline of the Project
- References

Problem Statement:

- To design and develop a **secure real-time chat application** that provides **end-to-end encrypted communication** along with **pre-encryption AI-based content moderation**, ensuring both **user privacy** and **online safety**.

Objectives

- To develop a real-time chat application with secure communication.
- To implement **End-to-End Encryption (AES & RSA)** for message confidentiality.
- To ensure messages are accessible only to sender and receiver.
- To integrate **AI-based moderation before encryption**.
- To detect and prevent toxic, abusive, spam, and phishing messages.
- To provide instant warnings or message blocking.
- To maintain low latency and scalability.

Background and Related work for title Selection:

- Applications such as **WhatsApp and Signal** focus primarily on **privacy using E2EE**.
- Due to encryption, these platforms cannot analyze message content in real time.
- Harmful messages are moderated **only after user reporting**, not proactively.
- Research shows **Transformer-based models (BERT)** are highly effective in detecting hate speech and toxicity.
- However, existing systems do not combine **client-side AI moderation with encryption**, motivating the proposed solution.

Analysis of Problem Statement :

- **Frontend:** React.js / HTML / CSS
- **Backend:** Python (Flask) / Node.js
- **Real-Time Communication:** WebSockets / Socket.IO
- **Encryption:** AES-256, RSA-2048
- **AI Model:** BERT-based Toxicity Detection
- **Database:** MongoDB / Firebase
- **Authentication:** JWT

Innovation or Novel Contributions:

- **Pre-Encryption AI Moderation (Key Innovation)**
- Real-time blocking of harmful messages before sending
- Combination of **AI safety + cryptographic security**
- Zero-trust server architecture (server never sees plaintext)
- Proactive prevention of cyberbullying and scams

Github Link

The Github link provided should have public access permission.

Github Link

https://github.com/Charu/Secure AI Chat System (E2EE + AI Moderation)

Analysis of Problem Statement (contd...)

Software and Hardware Requirements:

- **Software Requirements:**
- Windows / Linux
- Python, JavaScript
- TensorFlow / PyTorch
- Crypto libraries
- **Hardware Requirements:**
- Intel i5 or higher
- Minimum 8 GB RAM
- Stable Internet connection

Analysis of Problem Statement (contd...)

Proposed System / Methodology:

- **User Roles**: Admin, Registered User, **AI Moderation Engine**

- **Modules:**
- **User Authentication & Authorization**
  - Secure login and user identity verification
- **Key Management & Encryption Module**
  - RSA-based key exchange
  - AES-based message encryption
- **Real-Time Chat Module**
  - One-to-one and group messaging using WebSockets
- **AI-Based Content Moderation Module**
  - Toxicity, hate speech, spam & phishing detection
- **Message Control & Warning Module**
  - Message blocking / warning before sending
- **Encrypted Data Storage Module**
  - Secure storage of encrypted messages
- **Analytics & Reporting Module**

- Moderation statistics (blocked messages, warnings)

Analysis of Problem Statement (contd...)

Workflow (High-Level):

- User registers and logs into the system securely.
- System generates and manages encryption keys for users.
- User composes a message in the chat interface.
- **AI moderation engine analyzes the message before encryption.**
- If the message is safe, it is encrypted and sent to the receiver.
- If harmful, the system blocks or warns the user.
- Receiver decrypts and reads the message securely.
- Admin monitors overall system activity and moderation reports.

Timeline of the Project (Gantt Chart)

- **Weeks 1–2:** Literature review & requirement analysis
- **Weeks 3–4:** System design & encryption module
- **Weeks 5–6:** AI moderation model implementation
- **Weeks 7–8:** Real-time chat integration
- **Week 9:** Testing & validation
- **Week 10:** Documentation & Review-2 preparation

References (IEEE Paper format)

- [1] A. Singh and R. Sharma, “Secure end-to-end encrypted messaging systems,”
- *IEEE Communications Surveys & Tutorials*, 2021.
- [2] J. Davidson et al., “Toxic content detection using deep learning,” *IEEE Access*, 2021.
- [3] T. Brown et al., “AI-based content moderation systems,” *IEEE Transactions on AI*, 2022.


Thank
You!