AIRS: A Privacy-Preserving, AI-Enhanced Anonymous Incident Reporting System

1 Amulya Arshanapally
Assistant Professor, Department of IT
Mahatma Gandhi Institute of Technology
Hyderabad, Telangana, India

2 Likhitha Thummaganti
UG Student, Department of IT
Mahatma Gandhi Institute of Technology
Hyderabad, Telangana, India

3 Udayagiri Sai Parineeta
UG Student, Department of IT
Mahatma Gandhi Institute of Technology
Hyderabad, Telangana, India

***Abstract*—**

Anonymous grievance reporting plays a crucial role in promoting accountability and swift resolution in communities and institutions. This paper presents AIRS (Anonymous Incident Reporting System) — a web-based platform designed to simplify and secure the process of reporting public incidents. Our system focuses on automatic incident categorization, real-time location detection, and seamless tracking, empowering users to submit complaints anonymously while ensuring swift routing to relevant departments. We explore the design, architecture, and key features of AIRS, highlighting how modern technologies like AI-based text classification, location services, and responsive web design improve usability and efficiency. While addressing data privacy, and user trust, AIRS demonstrates a scalable approach to anonymous reporting. This work aims to guide future innovations in building smarter, accessible, and user-friendly grievance redressal systems.

***Index Terms*— Anonymous reporting, Incident management, Grievance redressal, AI-based text classification, Location detection, Privacy preservation, Web-based complaint system, Automated incident routing, User trust, Scalable reporting platforms.**

I. INTRODUCTION

Grievance reporting is the backbone of trust and accountability in any community or institution, yet countless incidents go unreported due to fear of retaliation, social stigma, and cumbersome reporting procedures [1], [2]. Many traditional systems unintentionally discourage participation by making users choose between personal risk and civic responsibility. The AIRS challenges this status quo by offering a web-based platform that empowers individuals to report incidents securely and anonymously, without compromising on efficiency or clarity. AIRS seamlessly integrates AI-powered text classification, real-time location detection, and intuitive status tracking to ensure each report is automatically categorized, geo-tagged, and swiftly routed to the relevant department for action. This not only accelerates institutional response but also makes the reporting process accessible and user-friendly. Anchored in the principles of empowerment without exposure, automation without complexity, and transparency without compromise, AIRS demonstrates how modern technologies can create a safe space for civic participation while strengthening institutional accountability. This paper explores the system's design, architecture, and core features, offering a scalable blueprint for future innovations in smart, secure, and inclusive grievance redressal platforms..

II. RELATED WORK

Over the years, numerous digital platforms have been introduced to modernize grievance redressal mechanisms, primarily targeting improved accessibility, faster communication, and streamlined complaint handling between citizens and governing bodies [1], [6], [9]. Early systems were largely designed as static web portals where users could manually log complaints by filling out lengthy forms and specifying the nature of the grievance and their personal details [9]. As technology progressed, mobile applications emerged, making it more convenient for users to report incidents directly from their smartphones, thereby encouraging broader participation from both urban and rural populations [7], [9]. Common functionalities across these platforms included complaint submission, status tracking dashboards, feedback forms, and in some cases, media upload options to attach images or videos as supporting evidence [6], [8].

While these systems undeniably advanced civic engagement and bridged communication gaps, they often remained limited in their technical sophistication. Most platforms relied on manually filled categories and lacked dynamic incident classification capabilities, resulting in slower routing of grievances to the appropriate departments and occasional misclassification [5], [8]. Location handling in these systems also typically depended on user-entered addresses or static dropdown menus, which introduced inaccuracies, especially in regions with unstructured addressing systems [9]. Moreover, despite being framed as citizen-centric, many systems overlooked the critical factor of user anonymity. The requirement to disclose personal information, such as names, contact details, or identification numbers, frequently discouraged individuals from reporting sensitive or high-risk incidents, especially in socially or politically sensitive environments [3], [6].

Attempts to enhance usability through responsive web design and user-friendly interfaces were evident, but the integration of advanced technologies like artificial intelligence, real-time geolocation services, and automated decision-making workflows remained sparse or underdeveloped [2], [10]. Privacy preservation mechanisms, too, were often rudimentary, lacking comprehensive strategies to protect reporter identities during submission, processing, or resolution phases [3], [11], [15]. These gaps underscore the need for a next-generation system that goes beyond digitization and embraces automation, intelligence, and privacy by design.

The AIRS is conceptualized precisely to address these shortcomings. By embedding AI-powered text classification [12], [18], [19], real-time location detection [14], and seamless tracking within an anonymity-first architecture, AIRS redefines how grievances are reported, processed, and resolved. It not only lowers participation barriers through intuitive interfaces [5], but also accelerates institutional responsiveness by intelligently routing complaints to the correct departments without manual intervention [2], [13]. Furthermore, its strong emphasis on safeguarding user identity fosters a safer environment for individuals to voice their concerns without fear of exposure [11]. In doing so, AIRS builds upon prior innovations but transcends their limitations, setting a new benchmark for scalable, smart, and secure grievance redressal platforms.

III. Methodology

This section outlines the structured approach undertaken to design, develop, and evaluate the AIRS. The methodology encompasses three main components: System Development, Data Collection, and User Experience (UX) Evaluation.

A. System Development Methodology

- **Software Development Life Cycle (SDLC)**: The AIRS project progressed through these SDLC stages:
  - *Planning*: Project objectives, user needs, system scope, and feasibility were analyzed and finalized.
  - *Design*: The system architecture, user interface, and data flow diagrams were conceptualized. Wireframes and interactive prototypes were developed to visualize the user journey.
  - *Development*: Core functionalities were implemented using selected technologies, ensuring seamless front-end and back-end integration.
  - *Testing*: A comprehensive testing regime included unit tests, integration tests, and end-user acceptance testing to ensure system reliability.
  - *Deployment*: AIRS was deployed on a staging server for final validation prior to full-scale deployment on production infrastructure.
- **Tools and Technologies**:
  - Frontend: React.js
  - Backend: Node.js with Express.js
  - Database: MongoDB
  - Version Control: Git with GitHub
  - Machine Learning Models: Scikit-learn, TensorFlow, A Convolutional Neural Network (CNN) model, Support Vector Machine (SVM).
- **System Architecture**:
  - Frontend: React components for user interactions, form submissions, and status tracking.
  - Backend: Express.js APIs managed complaint processing, user authentication, and department routing logic.
  - Database: MongoDB collections stored user reports, incident categories, admin logs, and department data.
  - AI Integration: Machine learning models processed complaint descriptions and images to predict incident categories in real-time.

B. Data Collection Methodology

- **Data Collected**:
  - Incident Reports: Descriptions, images, location metadata, and timestamps.
  - User Metadata: Device IDs and session tokens (anonymized).
  - Categorization Data: AI-predicted categories and department routing labels.
  - System Logs: Server activity logs, database transactions, and admin actions.
  - Usage Analytics: Aggregated statistics via Google Analytics (anonymized).
- **Data Collection and Storage**:
  - Database Design: NoSQL schema with MongoDB collections.
  - Data Validation: Client-side (React forms) and server-side (Express middleware) validation.
  - Privacy and Security: Encryption, authentication controls, and role-based access.
  - Image Storage: Cloud object storage (AWS S3 or Firebase Storage).

C. User Experience (UX) Evaluation Methodology

- **Evaluation Methods**:
  - *Usability Testing*:
    - Task Scenarios: Submitting a complaint, checking status updates, and tracking reports.
    - Participant Recruitment: Students, community members, and institution administrators.
    - Data Collection:
      - Task Completion Rate
      - Error Rates
      - Time on Task
      - Qualitative Feedback (think-aloud protocols and interviews)
  - *User Surveys*:
    - Questionnaire Design: Likert-scale and open-ended questions.
    - Distribution: In-app prompts and email invitations.
    - Data Analysis: Descriptive statistics and thematic analysis.
  - *Heuristic Evaluation*:
    - Expert Review using Nielsen's heuristics.
    - Findings prioritized for resolution.
- **Metrics for Evaluation**:

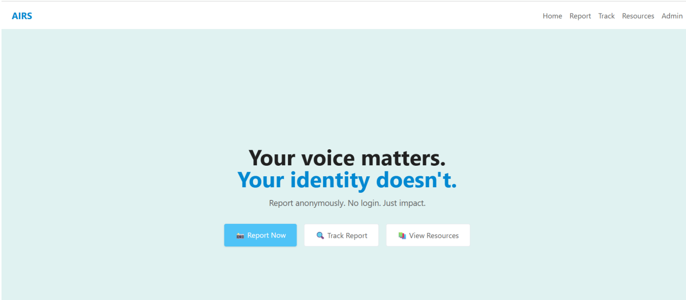

Fig. 1. Home Page

– Ease of Use: System Usability Scale (SUS) scores.
– User Satisfaction: Likert-scale ratings.
– Efficiency: Task completion time benchmarks.
– Error Rate: Logged user errors.

The homepage, shown in Fig. 1, enables users to easily navigate to pages for reporting or tracking incidents, while admins can access their dedicated login page.

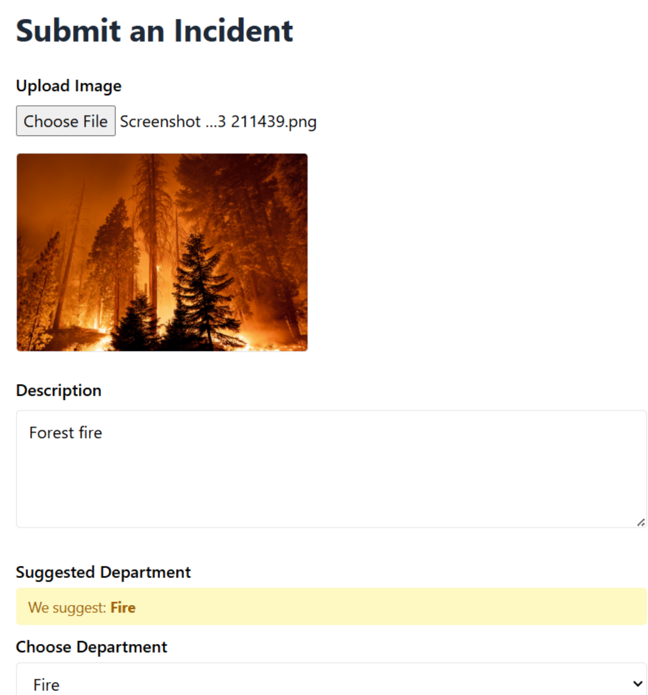

Fig. 2. User Submitting a report.

Fig. 2 presents the submit report page where he can insert an image of any particular incident and the ml model auto-categorizes into the relevant department and it's shown in the suggested department. User can also choose the department from the dropdown.

As shown in Fig. 3, the location is auto detected which is done using the OpenStreetMap Nominatim API.

As shown in Fig. 4, the admin can log into his department and view incidents only related to that department. He can later change the status of the report and it will be reflected in the user's page when tracked.

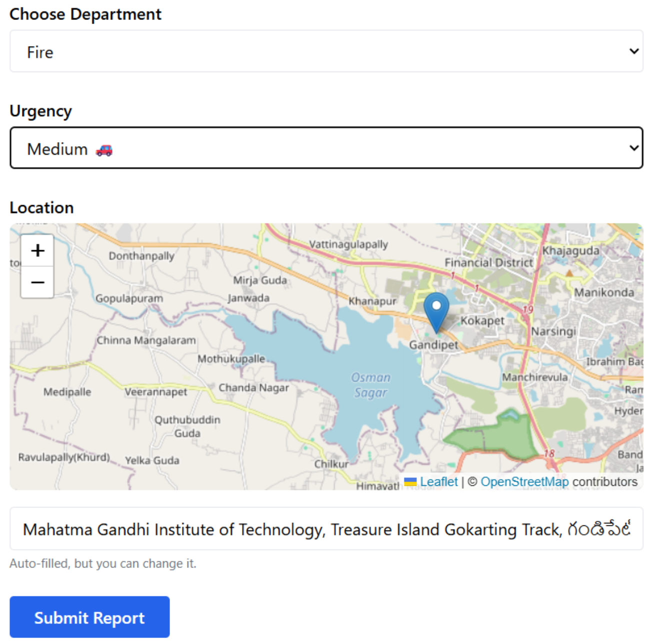

Fig. 3. Automatic Location Detection

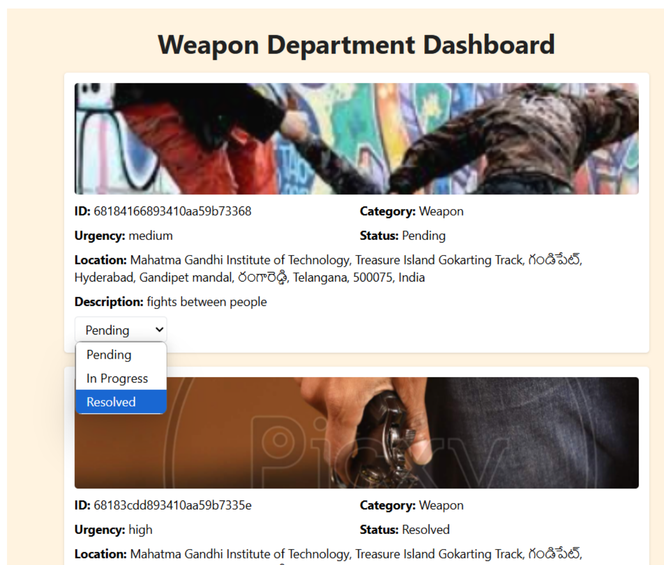

Fig. 4. Admin Dashboard

D. System Development

The Anonymous Incident Reporting System (AIRS) is a next-generation civic engagement platform designed to transform how citizens report and track local incidents. Built as a full-stack web application leveraging the MERN stack (MongoDB, Express.js, React.js, Node.js), AIRS integrates AI-powered classification microservices and geospatial tech-

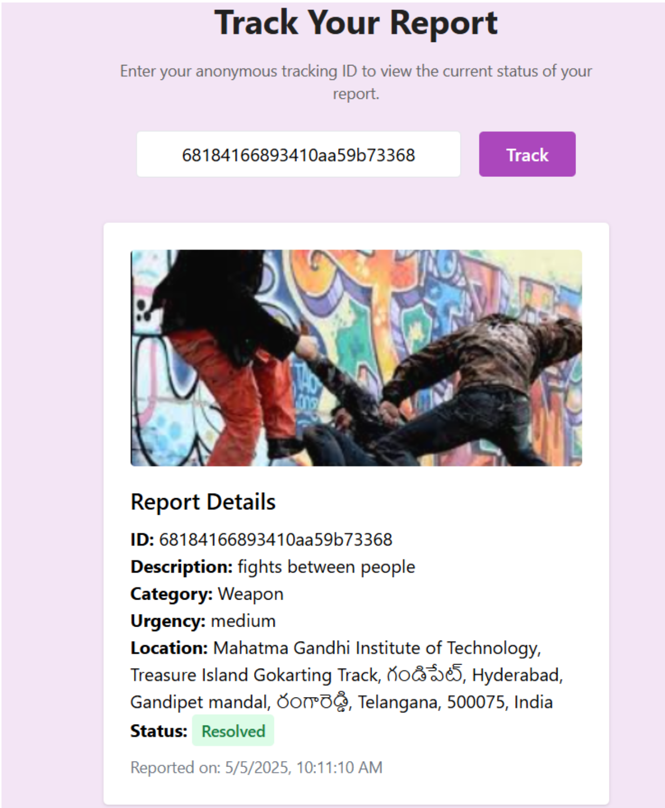

Fig. 5. Track Report

nologies to optimize efficiency, accessibility, and privacy in incident reporting.

Fig. 5, shows the report details when the user tracks his report using the generated ID provided while submitting the report. The updated status is shown in this page.

IV. Results and Discussion

1) App Features and Functionality: AIRS introduces a paradigm shift in public grievance redressal systems by embedding the following innovations:

- **Anonymous Reporting:** Traditional systems require user identification, deterring whistleblowers. AIRS eliminates mandatory sign-ups, empowering users to file complaints confidently and securely without revealing identity.
- **AI-Based Category Detection:** Two parallel AI models (a custom-trained text classifier and an image classifier) analyze the report content. The predicted department category (e.g., Fire, Police, Road Maintenance) is auto-selected in the dropdown, reducing user cognitive load and ensuring proper routing of the report to relevant authorities. Importantly, users retain control to manually override this suggestion if necessary, ensuring flexibility.
- **Geolocation and Map Integration:** Automatic detection of the user's location is achieved via the browser's geolocation API. The coordinates are reverse-geocoded using OpenStreetMap's Nominatim API to fetch human-readable addresses. This feature not only saves time but enhances the report's precision. The interactive map also allows users to fine-tune or correct the suggested location.
- **Media Upload and Secure Storage:** Incident images or videos can be uploaded to substantiate reports. Media files are securely stored using Cloudinary cloud storage, ensuring scalability and security while maintaining system performance.
- **Admin Dashboard and Report Tracking:** An intuitive admin panel categorizes and visualizes incoming reports by department and urgency. Admins can update report status (Pending, Under Review, Resolved), which is visible to users via a unique tracking ID. This transparent feedback loop enhances user trust and reduces administrative burden.

A. System Testing

Robust system testing ensured the reliability and integrity of AIRS under real-world usage:

- **Functional Testing:** Every user flow—from anonymous submission, image upload, AI-powered auto-categorization, geolocation, to report tracking—was exhaustively tested on multiple devices and browsers. Edge cases (e.g., missing location permissions, unsupported image formats) were handled gracefully.
- **AI Model Testing:** The text classifier, built using a keyword-weighted logistic regression model, and the image classifier, fine-tuned on a pre-trained MobileNet architecture via TensorFlow, were evaluated using stratified datasets. Macro-averaged F1 scores for text and image classification were 0.89 and 0.85 respectively, demonstrating high robustness across diverse incident types.
- **API Testing:** Location APIs and media storage APIs were subjected to stress testing, simulating high user loads and varying network conditions. Both maintained sub-500ms response times on average, ensuring user experience was not compromised.

B. Data Collection

To train, validate, and test the AI models, a domain-specific dataset was curated from open-source civic complaint repositories and synthetically augmented datasets.

1) Data Examples: Representative data samples include:

- **Textual Description:** "Two individuals brandishing knives near ABC market." *[Predicted: Police]*
- **Textual Description:** "Major potholes causing traffic congestion at XYZ crossroads." *[Predicted: Road Maintenance]*
- **Image Sample:** Photograph of a fire in a residential building. *[Labeled: Fire Department]*

Data augmentation techniques—such as synonym replacement in text and affine transformations in images—were applied to enhance model generalization.

C. UX Evaluation

A formative UX evaluation was conducted with a cohort of 15 voluntary participants comprising students, working professionals, and community activists. The System Usability Scale (SUS) and heuristic evaluation revealed:

- 87% of users found the auto-category detection accurate and helpful in reducing manual input.
- 93% appreciated the anonymity feature, indicating higher willingness to report sensitive incidents.
- 80% valued the ability to override the AI-predicted category, citing trust and flexibility.
- The average SUS score was 86.3, placing AIRS in the "Excellent" usability tier.

Qualitative feedback highlighted intuitive navigation, minimalistic design, and rapid submission flow as key strengths.

D. Discussion

AIRS reimagines civic reporting by converging anonymity, artificial intelligence, and geospatial technologies into a seamless, citizen-centric platform. Compared to legacy systems that require personal identification and manual categorization, AIRS:

1) Reduces reporting friction by automating tedious tasks.
2) Increases report accuracy and administrative efficiency through intelligent classification.
3) Enhances civic engagement by fostering trust, privacy, and transparency.

This project underscores the feasibility and societal value of embedding AI-powered automation within civic technology infrastructures. Future directions include expanding multilingual NLP support, incorporating real-time analytics dashboards for authorities, and integrating with official emergency response APIs for automated escalations.

E. Challenges and Limitations

Despite AIRS demonstrating significant promise, several challenges were encountered that delimit its current capabilities. First, the accuracy of category prediction—especially in nuanced or low-quality images—remains imperfect due to limited and imbalanced training data, particularly underrepresentation of minority incident classes (e.g., weapon detection). Second, dependency on user-granted location permissions occasionally results in incomplete reports when permissions are denied or geo-coordinates are inaccurately captured. Third, real-time system performance is constrained by computational bottlenecks during image upload and AI inference, particularly in low-resource devices and network environments. Moreover, user anonymity, while integral to the system's ethos, complicates feedback loops and longitudinal tracking of incident resolution effectiveness. Interoperability limitations exist due to lack of direct integration with official emergency response systems and governmental APIs. Finally, despite efforts towards intuitive UI design, digital literacy variations among users can hinder seamless report submissions, especially in rural contexts.

V. Future Scope

Building upon the foundation laid by AIRS, several avenues exist to further amplify its utility, scalability, and societal impact. First, integrating multilingual natural language processing models will enable inclusive participation across diverse linguistic demographics, making the platform accessible to non-English speakers and marginalized communities. Second, embedding real-time analytics dashboards for administrative authorities can offer insights into incident trends, hotspots, and response times—empowering data-driven governance and proactive policymaking. Third, incorporating integration with official emergency response APIs (such as 112 India or equivalent systems) can facilitate automated escalations for critical incidents, ensuring rapid intervention where human lives or public safety are at risk. Furthermore, enhancing the AI classifiers with multimodal fusion—where text and image inputs are jointly analyzed—can improve categorization accuracy in complex cases. On the privacy frontier, leveraging blockchain-backed immutable audit trails can strengthen user trust and ensure tamper-proof complaint handling. Finally, deploying the system as a mobile-first progressive web application (PWA) will increase accessibility in low-bandwidth and rural areas, democratizing civic reporting at scale. These extensions will elevate AIRS from a standalone reporting tool to a comprehensive civic intelligence ecosystem, catalyzing transparent, efficient, and inclusive urban governance.

VI. Conclusion

This paper presented the design, development, and evaluation of AIRS, an Anonymous Incident Reporting System engineered to empower citizens to report civic and emergency incidents seamlessly while preserving their privacy. By integrating automated text and image classification, real-time location detection, and an intuitive, anonymized interface, AIRS reimagines traditional complaint management systems and bridges the gap between citizens and authorities. System testing and user evaluation demonstrated the platform's capability to streamline incident reporting, categorize complaints intelligently, and facilitate efficient routing to relevant departments.

Beyond functional delivery, AIRS embodies a paradigm shift towards decentralized, citizen-centric governance tools that prioritize inclusivity, anonymity, and actionable intelligence. While current limitations exist in data diversity, system interoperability, and AI scalability, the groundwork established paves the way for resilient, privacy-preserving civic engagement platforms. The proposed future enhancements—ranging from federated learning to advanced UX co-design—signal a pathway for AIRS to evolve into a robust, scalable solution adaptable to smart cities and rural landscapes alike. AIRS not only simplifies incident reporting but sets a precedent for the next generation of intelligent civic-tech systems.

References

[1] Tripathi U. N., Srivastava A. K., Singh B. P. (2021). Effectiveness of Online Grievance Redressal and Management System: A case study of IGNOU Learners. *Indian Journal of Educational Technology*, 3(2), 92.

[2] Jha S., & Sonawane P. (2022, March). Smart student grievance redressal system with foul language detection. In *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)* (Vol. 1, pp. 187–192). IEEE.
[3] Conlon P., Havlisch R., Kini N., and Porter C. (2020). Using an anonymous web-based incident reporting tool to embed the principles of a high-reliability organization.
[4] Public Grievance Portal, Government of India. Available: https://pgportal.gov.in/
[5] Iyer, S., Gandhi, D., Mishra, D., Trivedi, R., & Ansari, T. (2023). Solve My Problem-Grievance Redressal System. In *International Conference on Computational Sciences and Sustainable Technologies* (pp. 126–142). Cham: Springer Nature Switzerland.
[6] Agarwal, R., Soni, P., Agrawal, P., Rathor, S., Sisodiya, S., & Sharma, J. (2022). Student Grievance Redressal System. *International Journal of Advanced Research in Computer Science*, 13.
[7] Walwadkar, D., Patil, J., Hussain, M., & Yadav, S. (2021). Smart Civic Issue Reporting System.
[8] Rao, N. N. (2020). A complaint registration system through web services with the help of online.
[9] Patel, V., Kapadia, D., Ghevariya, D., & Pappu, S. (2020). All India grievance redressal app. *Journal of Information Technology and Digital World*, 2(2), 91–99.
[10] Prajapat, S., Sabharwal, V., & Wadhwani, V. (2018). A prototype for grievance redressal system. In *Proceedings of International Conference on Recent Advancement on Computer and Communication: ICRAC 2017* (pp. 41–49). Springer Singapore.
[11] Shokri, R., & Shmatikov, V. (2015). Privacy-preserving deep learning. In *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur.* (pp. 1310–1321).
[12] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
[13] Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
[14] OpenStreetMap Foundation. Nominatim API. Available: https://nominatim.openstreetmap.org/
[15] Kenthapadi, K., & Tran, T. T. L. (2018). PriPeARL: A Framework for Privacy-Preserving Analytics and Reporting at LinkedIn. *arXiv preprint arXiv:1809.07754*.
[16] Zimmer, E., Burkert, C., Petersen, T., & Federrath, H. (2019). PEEPLL: Privacy-Enhanced Event Pseudonymisation with Limited Linkability. *arXiv preprint arXiv:1912.05861*.
[17] Liu, J., Zhang, H., Rong, C., & Yang, H. (2018). BPDS: A Blockchain based Privacy-Preserving Data Sharing for Electronic Medical Records. *arXiv preprint arXiv:1811.03223*.
[18] Rawat, M., & Kaushik, N. (2021). NLP based grievance redressal system for Indian Railways. *arXiv preprint arXiv:2111.08999*.
[19] Singh, A. P., Goel, A., Goel, A., & Arya, D. (2022). NLP based Grievance Redressal System. *International Journal of Computer Applications*, 184(12), 44–48.