

A Hybrid AI Framework for Privacy Risk Analysis in Digital Footprints Using NLP, Machine Learning, and Large Language Models

Shaik Sumaiya Sultana¹, Palle Sahasra², Jade Vaishnavi³, Vyahruti Cheemalakonda⁴, Ms. K. Priyanka

¹²³⁴ Department of Computer Science and Engineering,
Keshav Memorial Institute of Technology, Hyderabad, India

Abstract - The rapid growth of online platforms has increased the risk of unintentional disclosure of sensitive personal information. This paper proposes an AI-powered Privacy Footprint Analyzer that detects and evaluates privacy risks in user-generated text. The system integrates Natural Language Processing (NLP) for identifying personally identifiable information, a Machine Learning model for risk classification, and a Large Language Model (LLM) for generating context-aware recommendations and privacy-safe text rewriting. A hybrid architecture combining detection, scoring, and recommendation modules is implemented within a unified pipeline. Experimental results demonstrate high accuracy and low response time, enabling proactive privacy protection. The proposed system enhances user awareness and supports safer digital content sharing.

Key Words: Digital Footprint, Privacy Risk Analysis, Natural Language Processing, Machine Learning, Large Language Models, Cybersecurity

1. INTRODUCTION

The rapid proliferation of digital platforms has significantly transformed communication and information sharing, resulting in the expansion of users' digital footprints. These footprints often contain sensitive personal information, making individuals vulnerable to privacy breaches and cyber threats [1].

Users frequently post content without recognizing implicit disclosures such as phone numbers, addresses, affiliations, or behavioral patterns. These exposures can be exploited for malicious purposes including social engineering, impersonation, and targeted attacks [3].

While traditional cybersecurity tools focus on securing systems and credentials, they fail to address pre-publication privacy risks in user-generated content. This creates a critical gap in proactive privacy protection.

To address this challenge, this paper presents an AI-driven Privacy Footprint Analyzer that evaluates textual data before it is shared. The system leverages a hybrid architecture combining NLP for detection, ML for classification, and Large Language Models (LLMs) [8], [9] for contextual reasoning and recommendation generation.

Key Contributions

- A hybrid AI pipeline integrating NLP, ML, and LLMs
- A multi-factor privacy risk scoring model
- Context-aware recommendation and text rewriting system
- End-to-end implementation with real-time user interaction

2. RELATED WORK

Early privacy detection systems relied on rule-based approaches, including regular expressions and keyword matching. While efficient for structured data, these methods lack contextual understanding and fail in dynamic real-world scenarios.

NLP-based systems improved detection through Named Entity Recognition (NER) [1], enabling identification of entities such as names, locations, and organizations. Tools like Microsoft Presidio [4] and spaCy [5] provide scalable solutions but primarily focus on extraction rather than interpretation.

Machine learning approaches introduced risk classification [6], leveraging features such as entity frequency and text complexity. However, these models depend heavily on feature engineering and lack semantic reasoning.

Recent advancements in Large Language Models (LLMs) [8], [9] enable contextual understanding and generation of human-like responses. LLMs enhance privacy systems by providing explanations and actionable recommendations.

Despite these advancements, existing systems suffer from:

- Lack of unified architecture
- Absence of explainable risk scoring
- Limited user-centric recommendations
- Fragmented workflows

The proposed system addresses these gaps through a fully integrated and intelligent pipeline.

3. PROBLEM STATEMENT

Despite increasing awareness of cybersecurity, users continue to expose sensitive information due to:

- Lack of real-time feedback before sharing content
- Absence of contextual understanding in existing tools
- No unified system for detection, scoring, and mitigation
- Limited accessibility for non-technical users

There is a need for a system that can proactively analyze textual data, quantify privacy risks, and provide actionable insights in a single workflow.

4. PROPOSED SYSTEM

The proposed system follows a layered hybrid architecture designed for accuracy, scalability, and interpretability.

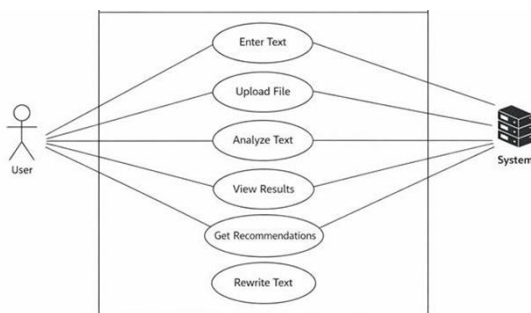


Fig 1: Sequence Diagram of Privacy Footprint Analyzer

The functional interactions of the system are illustrated in Fig 1.

1) 4.1 NLP Layer (PII Detection)

- Utilizes spaCy [5] and Microsoft Presidio [4] Detects:
 - Names
 - Emails
 - Phone numbers
 - Locations
 - Organizations

4.2 Feature Engineering Layer

Extracted features include:

- Entity count
- Entity density
- Sensitive keyword frequency
- Text length
- Contextual indicators

4.3 Machine Learning Layer

- Model: Random Forest Classifier [6]
- Output:
 - Risk Score (0–100)
 - Risk Category (Low / Medium / High)

4.4 LLM Layer (Intelligent Assistance)

- Integrated using LangChain agents [10]
- Generates:
 - Context-aware recommendations
 - Privacy-safe rewritten text
 - Explanations of detected risks

4.5 System Interface

- Frontend: React-based dashboard
- Backend: FastAPI
- Database: MongoDB

5. METHODOLOGY

The system follows a structured pipeline. The system workflow is illustrated in Fig 2.

1. Step 1: Input Acquisition

User provides text via UI or file upload.

2. Step 2: Entity Detection

NLP models extract PII entities with contextual tagging.

3. Step 3: Feature Extraction

Quantitative features are computed for ML processing.

4. Step 4: Risk Scoring Model

The privacy risk score is calculated as:

$$R = (S \times D \times F) / C$$

Where:

- S = Sensitivity of detected entities
- D = Entity density
- F = Frequency of occurrence
- C = Contextual protection factor

The approach is inspired by standard risk modeling techniques [3].

5. Step 5: Risk Classification

Random Forest classifies risk into discrete categories.

6. Step 6: LLM-Based Recommendation

LLM generates:

- Risk explanations
- Mitigation suggestions
- Safe rewritten content

Step 7: Visualization

Results displayed with:

- Risk score indicators
- Highlighted entities
- Recommendation panels

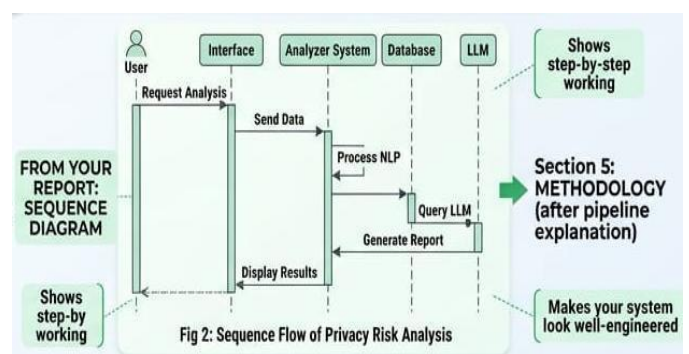


Fig 2: Sequence Diagram of Privacy Footprint Analyzer

Fig 2 illustrates the interaction flow of the system. The process begins with user input, followed by preprocessing and PII detection using NLP techniques. The extracted features are passed to the machine learning model for risk scoring and classification. The LLM module then generates recommendations and rewritten text, which are displayed to the user through the interface.

6. SYSTEM ARCHITECTURE

The system follows modular client-server architecture:

- User Interface (React)
- API Layer (FastAPI)
- AI Processing Pipeline
- Database Layer (MongoDB)

The overall system design is shown in Fig 3.

Table 1: Performance Evaluation Metrics of the Proposed System

Metric	Value
Accuracy	92%
Precision	0.91
Recall	0.90
F1-Score	0.905
Response Time	< 3 seconds

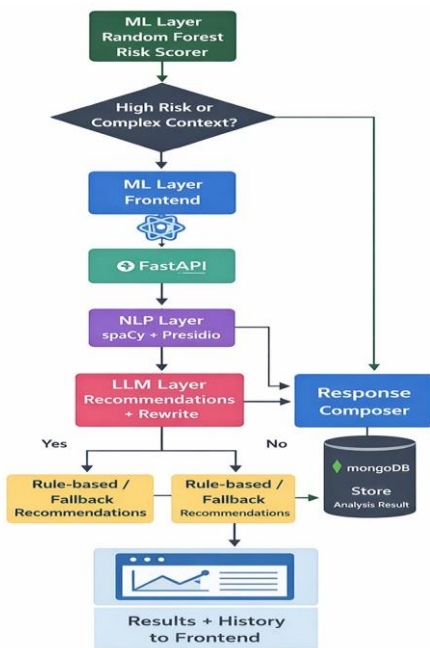


Fig 3: System Architecture of AI Privacy Footprint Analyzer

Fig 3 shows the overall architecture of the system. It consists of a user interface, API layer, AI processing pipeline, and database layer. The modular design ensures efficient data flow, scalability, and integration of NLP, ML, and LLM components.

This design ensures scalability, modularity, and efficient data flow.

7. RESULTS AND DISCUSSION

The proposed system was evaluated using standard performance metrics [3] to assess its effectiveness in

detecting and classifying privacy risks. The system outputs are shown in Fig 4, and performance evaluation is presented in Fig 5.

Performance Metrics

Table 1: Performance Evaluation Metrics of the Proposed System

8. Key Observations

- High accuracy in detecting PII across varied inputs
- Effective classification of privacy risk levels
- LLM-generated recommendations significantly improved usability
- Reduced false positives compared to rule-based systems

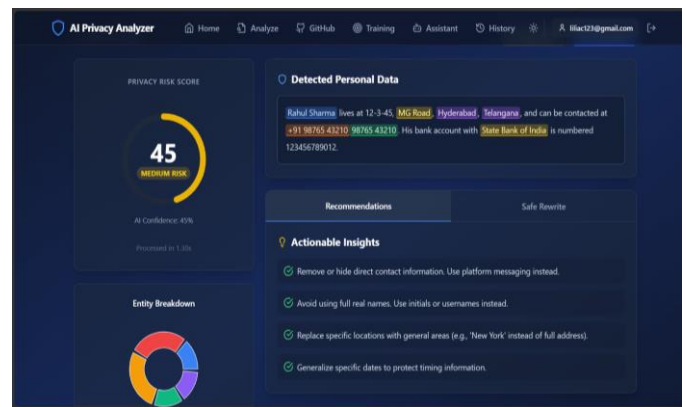


Fig 4: User Interface Showing Privacy Risk Detection and Recommendations

Fig 4 presents the system output displayed to the user. It highlights detected sensitive entities, calculated risk score, and AI-generated recommendations along with a privacy-safe rewritten version of the input text.

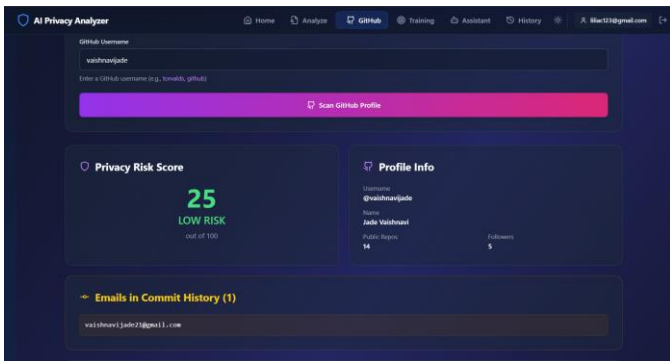


Fig 5: GitHub Profile Privacy Risk Analysis Output Interface

Fig 5 shows the system output for a GitHub profile analysis. It displays the calculated privacy risk score, detected sensitive information such as email exposure, and user profile details. The interface provides a clear visualization of risk levels along with categorized insights for improved user awareness.



Fig 6: Privacy Risk Analysis Performance Metrics

Fig 6 illustrates the performance evaluation of the proposed system using metrics such as accuracy, precision, recall, and F1-score. The results indicate high effectiveness in detecting and classifying privacy risks. **Comparative Insight**

Compared to traditional tools:

Table 2 presents the performance evaluation of the proposed system.

Feature	Traditional Tools	Proposed System
Context Awareness	Low	High
Risk Scoring	Limited	Advanced

Recommendations	None	AI-generated
Usability	Complex	User-friendly

8. CONCLUSIONS

This paper presents a comprehensive AI-driven solution for analyzing and mitigating privacy risks in textual data. By integrating NLP, ML, and LLM technologies, the system provides accurate detection, meaningful risk assessment, and actionable recommendations.

The proposed system shifts privacy protection from a reactive approach to a proactive, user-centric model [3], thereby enabling safer digital interactions and improved privacy awareness among users.

The experimental results validate the effectiveness and practicality of the proposed approach in real-world scenarios.

REFERENCES

- [1] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd ed., Pearson, 2023.
- [2] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv preprint arXiv: 1810.04805*, 2019.
- [3] T. Fawcett, "An Introduction to ROC Analysis," *Pattern Recognition Letters*, 2006.
- [4] Microsoft, "Presidio: Data Protection and PII Detection," 2023.
- [5] Explosion AI, "spaCy: Industrial-Strength NLP Library," 2023.
- [6] Scikit-learn, "Machine Learning in Python," 2023.
- [7] L. Tunstall et al., *Natural Language Processing with Transformers*, O'Reilly, 2022.
- [8] OpenAI, "GPT-4 Technical Report," 2023.
- [9] Meta AI, "LLaMA 3: Open Foundation Models," 2024.
- [10] LangChain, "LangChain Documentation," 2023.
- [11] MongoDB Inc., "MongoDB Documentation," 2023.
- [12] S. Ramírez, "FastAPI Documentation," 2023.