GuardLY: A Sophisticated AI-Based Framework for Real-Time Identification of Personal Information Leakage

Prof. Anjali Nair, Prof. Amit Patil, Tanashvi Pujari, Vivek Maske, Rasika Shinde, and Dinesh Dorkar
Department of Computer Science Business Systems
Bharati Vidyapeeth Deemed University
Navi Mumbai, India

***Abstract*—The proliferation of digital ecosystems has led to an unprecedented volume of unintentional personal data exposure. Sensitive identifiers, including contact details and residential information, are frequently broadcast across various web platforms, creating significant privacy risks such as identity theft and targeted phishing. While conventional security measures effectively address network-level threats, they often overlook content-level privacy leaks. This research introduces GuardLY, an innovative AI-driven framework designed to monitor and detect Personally Identifiable Information (PII) in real-time. By integrating advanced Natural Language Processing (NLP) with optimized rule-based engines, GuardLY provides users with immediate visibility into their digital footprint. Our findings demonstrate that a hybrid detection approach significantly enhances accuracy and reduces response latency, empowering individuals to reclaim control over their digital identities in an increasingly transparent online world.**

***Index Terms*—Data Privacy, PII Detection, Natural Language Processing, Cyber Security, Real-Time Monitoring.**

I. Introduction

In the current era of hyper-connectivity, personal data has evolved into a critical digital asset, often shared inadvertently through social media and online service portals. Users frequently populate digital forms or maintain public profiles without fully recognizing the accessibility of their sensitive information to malicious actors. This systematic exposure forms the bedrock for sophisticated social engineering attacks and unauthorized profiling.

Most contemporary security solutions prioritize infrastructure integrity, focusing on encryption and access control while neglecting the visibility of data content. Consequently, there is a distinct lack of tools that provide users with instantaneous feedback regarding their level of personal data exposure. GuardLY addresses this critical gap by offering a proactive, content-aware monitoring system. This paper details the architectural design and performance metrics of GuardLY, highlighting its role in transforming passive browsing into a secure, privacy-conscious experience.

II. Literature Review

Contemporary academic discourse has shifted towards utilizing artificial intelligence to bolster digital privacy. Researchers have explored various methodologies for identifying PII leaks, particularly within the context of unstructured web data. Hybrid architectures that combine the efficiency of regular expressions with the contextual depth of Named Entity Recognition (NER) have shown promise in improving detection precision. However, many existing frameworks are designed for large-scale enterprise audits and often lack the low-latency performance required for individual real-time applications.

Furthermore, studies on automated redaction have highlighted the potential for masking sensitive entities to comply with global privacy regulations. While these tools are effective for document processing, they rarely offer the interactive, live feedback necessary for active web navigation. Lightweight browser-integrated scanners have emerged as an alternative, yet they frequently struggle with contextual ambiguity. GuardLY builds upon these foundations by implementing a streamlined, hybrid engine that balances high-speed pattern matching with deep semantic analysis, providing a scalable solution for end-user privacy protection.

III. Problem Statement

Internet users often disclose sensitive identifiers—such as home addresses and phone numbers—across public platforms without an adequate understanding of the associated risks. Current security architectures are primarily defensive against malware and external intrusions, failing to notify users when their own data is explicitly visible to the public. The absence of a transparent, real-time monitoring mechanism prevents timely corrective actions, leaving individuals vulnerable to identity fraud. There is an urgent need for an automated system that can continuously evaluate and report on data exposure to fortify personal security.

IV. Existing System

Traditional privacy tools predominantly focus on infrastructure-level protection, such as blocking tracking cookies or monitoring network traffic. Although these defenses are essential, they do not scrutinize the actual content displayed on a webpage for visible personal data. Consequently, a user might be protected from technical trackers while their sensitive information remains exposed on a public forum. Rule-based systems utilizing regex are

efficient but often produce high false-positive rates due to a lack of contextual understanding.

Advanced NER-based systems offer superior context awareness but are typically resource-intensive and relegated to backend environments. Data Loss Prevention (DLP) suites, while robust, are generally targeted at corporate entities and are often too complex for individual use. This leaves a significant void for a lightweight, user-centric tool that provides immediate, content-level insights to help individuals manage their digital footprints effectively.

TABLE I
COMPARISON OF GUARDLY WITH EXISTING SYSTEMS

Feature	**Regex**	**NER Systems**	**DLP Suites**	**GuardLY**
Method	Pattern Match	Contextual NER	Policy Based	Hybrid (Regex+NLP)
Real-time	Yes	No (Backend)	Yes (Enterprise)	**Yes**
Target	Simple Patterns	Entities	Network Data	**Contextual PII**
Interface	Basic	Backend	Console	**Dashboard**
Efficiency	High	Low	Very High	**High**

V. PROPOSED SYSTEM

GuardLY is engineered as a comprehensive monitoring solution featuring a browser extension and an interactive web dashboard. The architecture is designed to provide seamless PII detection while maintaining the highest standards of user privacy and data ethics.

A. System Architecture and Technical Flow

As depicted in Figure 1, the GuardLY platform employs a layered processing strategy. The Input Acquisition Layer retrieves content from the browser's Document Object Model (DOM) or user-uploaded documents, which is then processed by the Central Processing Engine.

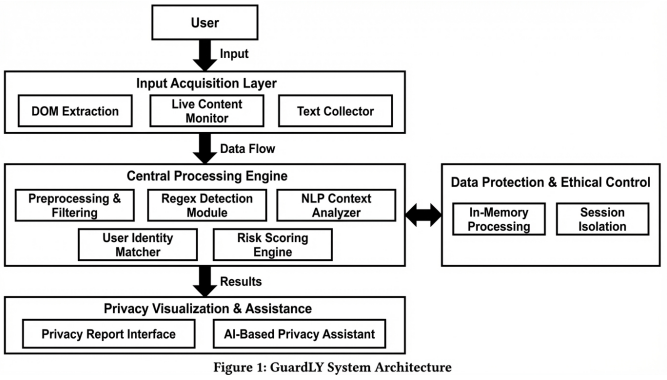

Figure 1: GuardLY System Architecture

Fig. 1. System architecture of the GuardLY platform, detailing the modular pipeline for real-time PII detection and risk assessment.

The processing pipeline consists of several specialized modules: 1) *Normalization*: Cleaning and formatting the input text. 2) *Pattern Identification*: Rapid detection of structured PII using optimized regular expressions. 3) *Semantic Analysis*: Utilizing NER to identify context-sensitive entities such as names and locations. 4) *Validation*: Cross-referencing findings with user profiles to minimize inaccuracies. 5) *Risk Evaluation*: Determining an exposure index based on the sensitivity of the detected data.

The Privacy Visualization layer then delivers real-time alerts and remediation guidance. All data processing is performed in-memory to ensure zero persistence of sensitive information.

VI. IMPLEMENTATION RESULTS

GuardLY was evaluated across diverse web environments and document types, demonstrating exceptional performance in identifying sensitive data with negligible latency.

TABLE II
PERFORMANCE EVALUATION OF GUARDLY DETECTION SYSTEM

Detection Type	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Email Detection	94%	92%	91%	91.5%
Phone Number Detection	86%	84%	82%	83%
Overall System Performance	90%	88%	86.5%	87.2%

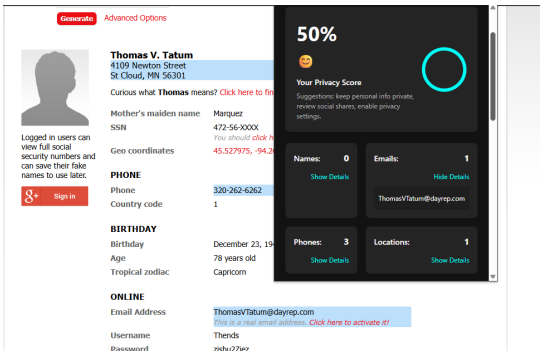

Fig. 2. GuardLY Web Extension interface: Real-time monitoring and detection of PII while browsing.

A. Risk Assessment and Reporting

The system generates a Privacy Risk Report (Figure 3), providing a comprehensive overview of exposure levels. The report identifies specific PII categories and calculates a risk score based on the sensitivity of the leaked data. The interface also provides automated remediation recommendations, enabling users to take immediate action.

Fig. 3. GuardLY Privacy Risk Report interface: Visualization of detected PII, risk metrics, and remediation strategies.

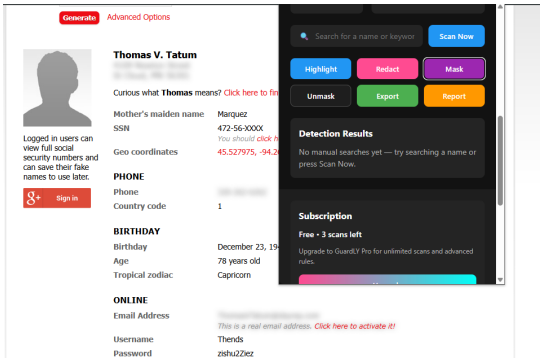

Fig. 4. Data masking features: Demonstrating the automated redaction capabilities of the GuardLY system to protect sensitive information.

VII. Conclusion

The GuardLY framework addresses a fundamental vulnerability in modern digital life by providing content-level visibility into personal data exposure. By synthesizing real-time analysis with sophisticated AI, it empowers users to manage their digital presence actively. Our implementation confirms that proactive PII detection is a viable and necessary addition to the contemporary security stack, offering a significant reduction in the potential for identity theft.

Unlike reactive security tools, GuardLY prioritizes user awareness and control. The integration of advanced semantic analysis ensures that even subtle data leaks are identified, while the interactive dashboard translates complex risk metrics into actionable intelligence. As digital footprints continue to expand, tools like GuardLY will become indispensable in safeguarding personal privacy.

Future development will focus on expanding multilingual support and integrating localized privacy regulations to ensure global applicability. We also intend to explore federated learning to enhance detection accuracy without compromising user data. GuardLY represents a pivotal step towards a more secure digital future, where privacy is an actively managed right rather than an overlooked consequence of connectivity.

Acknowledgment

The authors express their sincere gratitude to Prof. Anjali Nair and Prof. Amit Patil for their expert guidance and mentorship throughout this research. Their insights were vital to the successful development of the GuardLY framework.

References

[1] P. Acharya and L. Shrestha, "Comparative Study of Lightweight Language Models for PII Masking and their Deployment for Real Conversational Texts," Preprint, 2025.
[2] Z. Zhou, K. Vanacore, B. Ahtisham, et al., "Utility-Preserving De-Identification for Math Tutoring: Investigating Numeric Ambiguity in the MathEd-PII Benchmark Dataset," Preprint, 2026.
[3] A. S. Almasah and S. Alyaseri, "Kubernetes-Based Framework for Scalable PII Detection and Redaction in Application Logs," 2017 IEEE Region 10 Symposium (TENSYMP), 2025.
[4] A. S. Almasah and S. Alyaseri, "Kubernetes-Based Framework for Scalable PII Detection and Redaction in Application Logs," Preprint, 2025.
[5] S. Zhang, Y. Ma, Y. O. Hu, et al., "From Patient Burdens to User Agency: Designing for Real-Time Protection Support in Online Health Consultations," Preprint, 2025.
[6] H. Rajgarhia, S. Gupta, A. Shaik, et al., "An Evaluation Study of Hybrid Methods for Multilingual PII Detection," Preprint, 2025.
[7] S. Patel and P. Verma, "Detecting Sensitive Information from Documents," Preprint, 2025.
[8] L. Garza, A. Kotal, A. Piplai, et al., "PRvL: Quantifying the Capabilities and Risks of Large Language Models for PII Redaction," arXiv.org, 2025.
[9] S. Asthana, R. Mahindru, B. Zhang, et al., "Adaptive PII Mitigation Framework for Large Language Models," Preprint, 2025.
[10] Y. Zhou, X. Li, T. Lu, et al., "Intelligent Recognition of Sensitive Information in Mobile Office and Leakage Detection System," Advances in transdisciplinary engineering, 2025.
[11] E. Gouvea, A. Dadgar, S. Jalalvand, et al., "Trustera: A Live Conversation Redaction system," Preprint, 2023.
[12] Author, "Hybrid PII detection and anonymization in financial documents," 2025.
[13] T. W. Chiang and S. H. Lin, "Smartphone-based attendance tracking via GPS and NFC," IEEE Access, 2022.
[14] Shaswat, "Location tracking via Android GPS APIs," 2022.
[15] Sharma and V. Gupta, "Geo-Fencing for smart reminder systems," 2021.
[16] Nallusamy, "GPS-integrated attendance systems," 2020.
[17] Enikuomehin and Dosumu, "Geofencing for monitoring applications," 2021.
[18] Kumawat, "Biometric and Geofence based management systems," 2022.
[19] P. Iksan et al., "Design of administrative presence systems," JNKTL, vol. 3, 2020.
[20] Irwanda, "Waterfall methodology for UMKM information systems," 2022.