MULTIDISCIPLINARY RESEARCH INSTITUTE INDIA

Multidisciplinary Journal of Research in Engineering and Technology

ISSN: 2348-6953

Volume 13 Issue 02 (2026) | Pages: 418-423

A Comprehensive Result Paper on UnveilThreatAI: See the Invisible Threat in What You Share

A. B. Gavali[1], Aniket Baral[2], Ketaki Pawar[3], Gitanjalee Rakshe[4]

[1,2,3,4]Department of Artificial Intelligence and Data Science, S. B. Patil College of Engineering, Indapur, Pune, India

[1]*dnyane.ash@gmail.com*, [2]*aniketbaral08966@gmail.com*, [3]*ketakipawar877@gmail.com*, [4]*rakshegitanjalee@gmail.com*

Peer Review Information

***Type**: Article*
***Received**: 22 March 2026*
***Revised**: 06 April 2026*
***Accepted**: 24 May 2026*
***Published**: 05 June 2026*

Abstract

The project, "UnveilThreatAI: See the Invisible Threats in What You Share," provides an AI-powered system to support users in identifying cybersecurity risks hidden in images, documents, and URLs. The platform combines image metadata inspection, document parsing, suspicious link analysis, explainable AI, and risk visualization to help users make safer sharing decisions. The system is implemented as a web-based application using React for the frontend, Django REST Framework for the backend, PostgreSQL for storage, and Python-based AI/ML utilities for analysis. It generates a consolidated report that includes a severity score, metadata exposure findings, suspicious URL indicators, and a consequence chain showing how detected issues may lead to phishing, privacy loss, identity theft, or financial fraud. Experimental observations show that the system processes moderate-size inputs reliably and achieves observed detection performance in the range of 80–90% on representative test samples. The platform improves user awareness by combining automated detection with simple, interpretable output.

Keywords: Cybersecurity Risk Analysis; Phishing Detection; URL Analysis; Document Inspection; Metadata Leakage; Explainable AI.

How to Cite This Article

Gavali, A. B., Baral, A., Pawar, K., & Rakshe, G. (2026). A comprehensive result paper on UnveilThreatAI: See the invisible threat in what you share. *Multidisciplinary Journal of Research in Engineering and Technology*, 13(2), 418–423.

Introduction

Cybersecurity threats have become increasingly common in day-to-day digital communication [5]. Users frequently share screenshots, scanned documents, PDF files, links, and images without realizing that these files may contain exposed metadata, suspicious URLs, hidden text, or privacy-sensitive information [10]. Traditional user behavior often depends on trust and convenience rather than security validation, which creates a need for systems that can automatically analyze shared content before it causes harm [5].

To address this issue, this project introduces *UnveilThreatAI*, an integrated web-based platform that uses artificial intelligence and content-aware analysis to detect hidden cyber risks [5]. The system is designed to analyze multiple content types within a single workflow and produce a unified risk report for the user [5]. Instead of providing only a safe-or-unsafe label, the platform explains why a threat is important and what consequences may follow if the content is shared without caution [4].

The proposed system improves cybersecurity awareness by combining content inspection, machine learning, explainable output, and visual analytics [9]. By leveraging metadata extraction, URL analy-sis, document inspection, and consequence visualization, the platform helps users make better security decisions and reduces the gap between technical detection systems and practical user understanding [3].

Literature Survey

This section summarizes the key research studies considered in this work and highlights how the present session focuses on malicious URL detection, document threat analysis, explainable AI, visualization, and hidden-information exposure in shared digital content [1]. These studies provide the academic foundation for the design and evaluation of *UnveilThreatAI* [5].

Y. Tian, Y. Yu, J. Sun, and Y. Wang [1] addressed the continuing challenge of malicious URL detection in the presence of phishing, malware, and data-theft campaigns. Their survey reviewed blacklist, heuristic, machine learning, deep learning, transformer, graph-based, and LLM-driven approaches. The study showed that modern URL detection requires stronger multimodal features and better benchmark standardization for future progress.

S. Kuikel, A. Piplai, and P. Aggarwal [2] examined large language models for phishing detection, with particular attention to self-consistency, faithfulness, and explainability. Their work highlighted that explainable decision support is important for security applications because high accuracy alone is not sufficient when users and analysts need understandable outputs.

S. Liu, J. Ming, G. Zhou, X. Liu, J. Fu, and G. Peng [3] focused on adversarially robust PDF malware analysis. They proposed a stronger representation of PDF nostructure and showed that document-based threats require analysis beyond superficial text extraction. Their findings support the importance of deep document inspection in cyber defense systems.

B. Lim, R. Huerta, A. Sotelo, A. Quintela, and P. Kumar [4] proposed an explainable phishing detection framework that combined machine learning with LIME, SHAP, and language-model-based interpretabil-ity. The study emphasized that trustworthy interfaces are essential when users need to understand risk predictions and act on them.

A. Al Siam, M. Alazab, A. Awajan, and N. Faruqui [5] reviewed the broader impact of AI in cyber-security, covering malware detection, phishing analysis, intrusion detection, and risk assessment. Their work showed that AI has significant promise in cybersecurity but also requires practical explainability, privacy awareness, and real-time usability.

W. Li, S. Manickam, Y. W. Chong, W. Leng, and P. Nanda [6] surveyed phishing website detection techniques and noted that blacklist-only methods are insufficient for fast-changing online attacks. Their review supported the need for intelligent analysis that considers lexical, structural, and behavioral signals in URLs.

U. Zara, K. Ayub, H. U. Khan, A. Daud, T. Alsahfi, and S. Gulzar [7] studied phishing website detection using deep learning models. Their results showed that evolving phishing patterns can be captured more effectively through deep architectures, though real-world deployment still requires robust interpretation and generalization.

A. Bensaouda, J. Kalita, and M. Bensaouda [8] surveyed deep learning methods for malware detec-tion across multiple environments. Their findings reinforced the value of intelligent classification but also pointed out continued challenges related to explainability, adversarial robustness, and operational deployment.

H. Ahmad, F. Ullah, and R. Jafri [9] analyzed immersive cyber situational awareness systems and showed that effective visualization reduces cognitive burden and supports faster understanding of cyber events. This work motivates the use of visual analytics in cybersecurity tools intended for end users.

O. Kuznetsov, E. Frontoni, K. Chernov, K. Kuznetsova, R. Shevchuk, and M. Karpinski [10] explored AI-based steganography detection in images. Their study demonstrated that image-based cyber risks can extend beyond visible content, supporting the need for hidden-pattern

and metadata-aware inspection in image analysis pipelines.

Limitations of Existing Work: Existing systems are often specialized for only one type of input, such as URLs, documents, or images, which makes them less useful for everyday sharing scenarios [6]. Many tools provide only binary classifications and do not explain the likely consequences of a detected threat [4]. Some approaches also ignore metadata leakage, hidden content, or the relationship between technical findings and real-world user harm [10]. In addition, fragmented workflows, limited user-friendly dashboards, and the absence of unified scoring reduce the practical usefulness of current tools for non-technical users [9].

Problem Statement

A web-based platform that scans user inputs, generates a detailed risk report, and visualizes the possible chain of consequences to improve user awareness and online safety [9].

Proposed System: *UnveilThreatAI* is designed as a unified cybersecurity analysis platform that inspects images, documents, and direct URLs through a single web interface [5]. The system extracts metadata, text, embedded links, lexical URL features, and other relevant indicators to evaluate the security posture of submitted content [3]. The image analysis component checks metadata exposure and text-based leakage, the document analysis component scans extracted content and embedded links, and the URL analysis component evaluates phishing and suspicious structure [1]. The backend aggregates these findings into a weighted risk score and classifies the overall severity level [4].

In addition to detection, the system provides explainable output and visual analytics [4]. The dashboard shows a risk meter, metadata exposure details, suspicious link findings, and a consequence chain that connects a detected issue to possible harms such as identity theft, privacy loss, account compromise, and financial fraud [9]. This design helps users move beyond simple detection results and understand why an alert matters [4]. The platform is implemented using React, Django REST Framework, PostgreSQL, and Python-based AI/ML utilities, making it suitable for modular deployment and future extension [5].

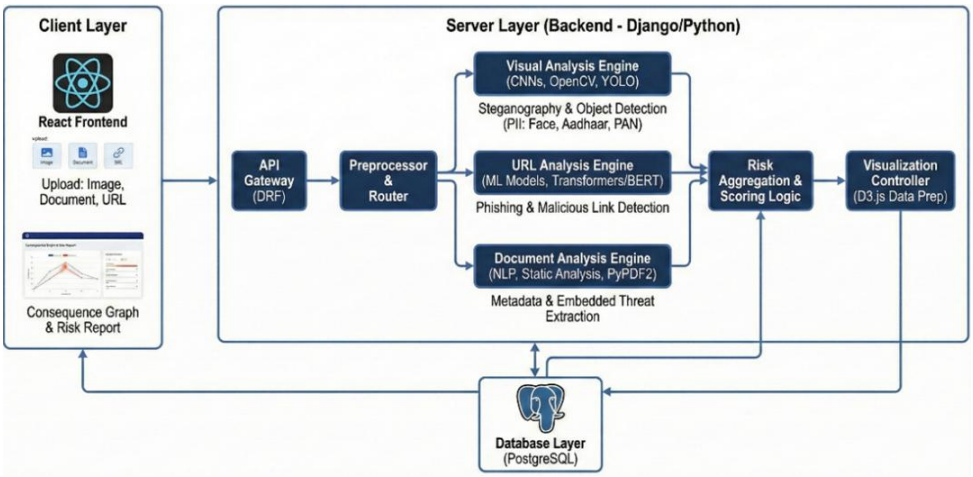


Fig. 1. *System architecture of UnveilThreatAI showing the interaction between input handling, analysis modules, and user-facing outputs.*

System Requirements

The efficient operation of *UnveilThreatAI* depends on a combination of suitable database support, compatible software tools, and adequate hardware resources [5]. The following requirements summarize the recommended environment for development, testing, and practical deployment [6].

Database Requirements:

- PostgreSQL for production data storage.
- SQLite may be used for lightweight local development and testing.

Software Requirements (Platform Choice):

- Operating System: Windows 10/11 or Linux.
- Coding Language: Python and JavaScript.
- Frontend: React. Backend: Django and Django REST Framework.
- IDE: VS Code or equivalent editor.
- Web Browser: Google Chrome or any modern browser.

Hardware Requirements:

- System: Intel i5 or equivalent processor.
- RAM: 8 GB minimum.
- Hard Disk: 256 GB minimum available storage.
- Network: Stable internet connection for web access and API-based updates.

Methodology

The development of the *UnveilThreatAI* system follows a structured methodology to ensure effective threat detection, interpretability, and reliable web-based deployment [4]. The methodology consists of multiple phases, including data preparation, preprocessing, model-based analysis, system integration, and result generation [3].

Data Collection and Preprocessing:

- Image inputs are processed to extract metadata, visible text, and sensitive visual indicators.
- Document inputs are parsed to extract text, embedded links, and metadata from uploaded files.
- URL inputs are analyzed for lexical patterns, structural anomalies, and suspicious characteristics.
- The extracted features are cleaned and structured into machine-readable representations for downstream analysis.

Threat Analysis Model Development:

- Rule-based and AI/ML-assisted modules are used to identify phishing, privacy leakage, suspicious links, and hidden risk indicators.
- Risk signals from image, document, and URL analysis are converted into comparable scores.
- Explainability methods are applied so that influential features can be presented in the final report.

System Development and Integration:

- A React-based user interface is used for content upload, report access, and visual dashboard rendering.
- Django REST Framework manages input validation, request handling, analysis orchestration, and report delivery.
- PostgreSQL stores reports, logs, and structured findings for later access and evaluation.

System Workflow:

- Users upload an image, document, or URL through the web interface.
- The backend validates the submitted input and extracts relevant threat-related features.
- Analysis modules inspect the content and produce threat indicators with severity values.
- The scoring engine combines the indicators into an overall risk level.
- The final dashboard displays the report, recommendations, and consequence visualization.

Results And Discussion

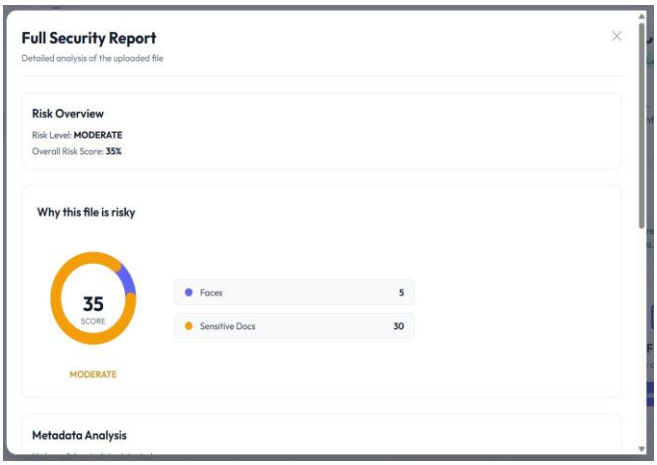
The *UnveilThreatAI* system was implemented and evaluated to assess its detection capability, processing stability, and user-facing interpretability. The platform successfully analyzed representative image, docu-ment, and URL inputs, generating consolidated reports with severity indicators and explanation-oriented summaries. The observed detection performance ranged between 80–90% on representative test samples, demonstrating that the system is effective for practical demonstration and awareness-oriented cybersecurity support.

The system produces multiple outputs that enhance both analytical depth and user understanding [5]. These include an overall risk score categorized as low, medium, or high severity, along with a metadata exposure report for image and document inputs [10]. It also identifies suspicious links and phishing indicators from URLs and embedded document content [6]. An explainable report highlights the key factors influencing the detection outcome [4]. The dashboard visualization presents insights through a risk meter and a consequence chain [9]. Additionally, the system provides actionable safety recommendations to promote safer user behavior [5].

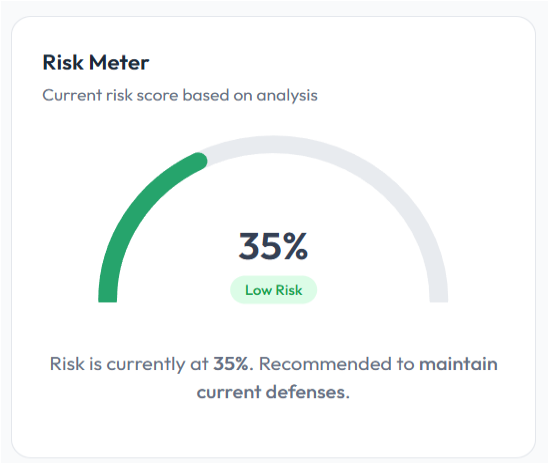
From a functional perspective, the image analysis pipeline effectively detects metadata exposure and visible text leakage [10]. The document inspection module identifies embedded threats and extracts risk-relevant content [3]. The URL analysis component performs phishing detection using lexical and structural indicators [1]. These outputs are integrated through a weighted scoring engine, which produces a unified severity assessment that users can interpret quickly and intuitively [4].

The user-facing dashboard significantly enhances usability and interpretability [9]. The risk meter simplifies threat evaluation, the metadata exposure panel increases awareness of privacy risks, and the consequence chain translates technical findings into potential real-world impacts [4].

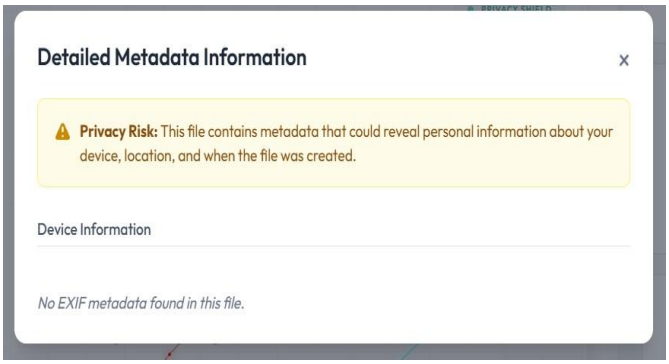
Despite these strengths, some limitations remain. High-resolution files require longer processing time, and the current evaluation is limited in scale. Expanding the dataset and conducting broader benchmarking would further strengthen the reliability and generalizability of the results.



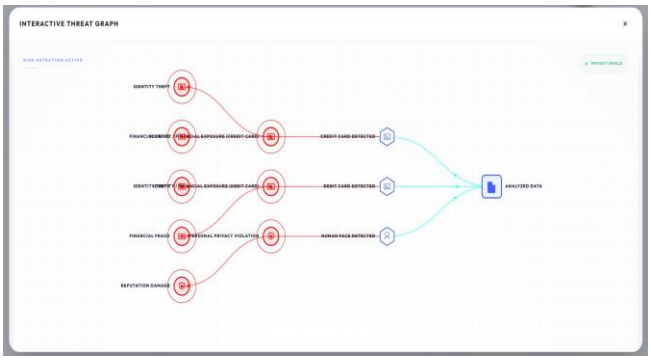
(a) Full report screen.



(b) Risk meter visualization.



(c) Metadata exposure result.



(d) Consequence-chain visualization.

Fig. 2. *Representative result outputs of UnveilThreatAI shown in a 2 × 2 layout.*

Conclusion

The *UnveilThreatAI* system successfully integrates content inspection, AI-assisted threat detection, explainable reporting, and web-based visualization to provide users with practical cybersecurity support [5]. By analyzing images, documents, and URLs within a single platform, the system helps users identify hidden threats that are often missed during routine digital sharing [3]. The observed results indicate that the platform can reliably process representative inputs and present understandable reports that improve cybersecurity awareness [9]. The project demonstrates that combining automated detection with consequence-oriented visualization can make cybersecurity tools more useful for non-technical users [4]. Future improvements may focus on expanding datasets, improving large-file processing efficiency, strengthening adversarial robustness, and extending the platform with multilingual support and broader real-time deployment capabilities [8].

References

1. Y. Tian, Y. Yu, J. Sun, and Y. Wang, “From Past to Present: A Survey of Malicious URL Detection Techniques, Datasets and Code Repositories,” 2025.
2. S. Kuikel, A. Piplai, and P. Aggarwal, “Evaluating Large Language Models for Phishing Detection, Self-Consistency, Faithfulness, and Explainability,” 2025.
3. S. Liu, J. Ming, G. Zhou, X. Liu, J. Fu, and G. Peng, “Analyzing PDFs like Binaries: Adversarially Robust PDF Malware Analysis via Intermediate Representation and Language Model,” 2025.
4. B. Lim, R. Huerta, A. Sotelo, A. Quintela, and P. Kumar, “EXPLICATE: Enhancing Phishing Detection through Explainable AI and LLM-Powered Interpretability,” 2025.
5. A. Al Siam, M. Alazab, A. Awajan, and N. Faruqui, “A Comprehensive Review of AI’s Current Impact and Future Prospects in

Cybersecurity," 2024.
6. W. Li, S. Manickam, Y. W. Chong, W. Leng, and P. Nanda, "A State-of-the-Art Review on Phishing Website Detection Techniques," 2024.
7. U. Zara, K. Ayub, H. U. Khan, A. Daud, T. Alsahfi, and S. Gulzar, "Phishing Website Detection Using Deep Learning Models," 2024.
8. A. Bensaouda, J. Kalita, and M. Bensaouda, "A Survey of Malware Detection Using Deep Learning," 2024.
9. H. Ahmad, F. Ullah, and R. Jafri, "A Survey on Immersive Cyber Situational Awareness Systems," 2024.
10. O. Kuznetsov, E. Frontoni, K. Chernov, K. Kuznetsova, R. Shevchuk, and M. Karpinski, "Enhancing Steganography Detection with AI: Fine-Tuning a Deep Residual Network for Spread Spectrum Image Steganography," 2024.