IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, November 2025
Impact Factor: 7.67

SafeURL_AI_Extension: An Intelligent Browser Extension for Real-Time Detection of Malicious and Phishing Content

Mr. Omkar Santosh Kale , Mrs. Pratiksha Shivnath Dhumal , Mrs. Sakshi Sachin Thorat, Prof. K. S. Ambatkar, Dr. A. A. Khatri
Student, Computer Department
Jaihind College of Engineering Kuran, Pune, India
kaleomkar992@gmail.com , dhumalpratiksha57@gmail.com, thoratsakshi988@gmail.com

Abstract: *Modern cyberattacks, including phishing, fake popups, malicious scripts, and harmful downloads, have become more sophisticated and cannot easily be detected. Most of the traditional browser protections are based on static blacklists that cannot recognize newly emerging threats. This paper introduces SafeURL_AI_Extension, an intelligent browser extension based on machine learning, threat intelligence, and behavior analysis to identify and block malicious or phishing content hosted on legitimate websites in real time. The proposed system examines URLs, Web forms, pop-ups, embedded scripts, and download behaviors through an AI classifier continuously improved through user feedback and global threat data. SafeURL_AI_Extension combines real-time scanning, adaptive learning, and privacy-preserving analysis for online safety without disrupting user experience.*

Keywords: Browser Extension, Cybersecurity, Phishing Detection, Machine Learning, Threat Intelligence, Real-Time Protection

I. INTRODUCTION

With the rapid expansion of online services, users are often exposed to phishing pages and malicious scripts that masquerade as legitimate websites. The existing solutions like built-in browser protections and antivirus plugins protect based on static blocklists or signature-based detection, which are not able to handle zero-day threats.

SafeURL_AI_Extension addresses this limitation by integrating AI-driven URL and content analysis within a lightweight browser extension. It detects malicious behavior such as fake login forms, popups, injected scripts, and unsafe file downloads with the help of both local and cloud-based intelligence. The system, coded in JavaScript and Python APIs, along with a lightweight ML model, monitors the activities on a webpage, extracts security features, classifies potential threats, and automatically blocks unsafe interactions.

This solution offers continuous protection with respect to user privacy, doing most of the analysis locally..

II. PROBLEM STATEMENT

Cyberattacks, such as phishing, fake pop-ups, malicious scripts, and hazardous downloads, are on the rise in today's world with rapid growth in the use of the internet and web services. These kinds of attacks have now started bypassing the usual ways that browsers defend against them.Most of the current protection systems depend on static blacklists or signature-based detection, failing to identify new and evolving threats in real time.As a result, users remain vulnerable to data theft, financial loss, and privacy breaches even when browsing through what seems to be valid websites.

There is a demand for an intelligent, adaptive, real-time protection system that could analyze and detect malicious behaviors dynamically.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, November 2025
Impact Factor: 7.67

III. OBJECTIVES

• Develop an intelligent browser extension that can detect and block phishing, malicious scripts, and harmful downloads in real time.
• Using machine learning algorithms to classify safe and unsafe URLs, pop-ups, and embedded scripts with high accuracy.
• To incorporate global threat intelligence that provides updates on newly emerging cyber threats continuously.
• To carry out behavioral analysis of websites, forms, and downloads to identify suspicious or abnormal activities.
• To implement adaptive learning so that the system can improve its detection accuracy based on user feedback and from threat data.
• Ensuring privacy preservation by analyzing web content and user interactions without the collection of personal data.
• To offer real-time protection without compromising a smooth, non-intrusive user browsing experience.
• The goal is to reduce dependence on static blacklists and develop a dynamic, AI-driven defense mechanism against emerging cyber threats.

IV. LITERATURE REVIEW

As such, according to [1], traditional anti-phishing mechanisms highly rely on static URL blacklists and predefined heuristics. While these methods are fast, they fail to identify zero-day attacks and obfuscated phishing links. The study accentuates the call for AI-driven adaptive systems that can analyze URL structures, domain characteristics, and content semantics to identify previously unseen malicious sites. This insight forms a backbone for our project by integrating AI classifiers to dynamically identify emerging threats.

The work in [2] proposed a machine learning-based phishing URL detection system that uses both lexical and host-based features like URL length, presence of IP addresses, subdomain depth, and HTTPS certificate validity. It demonstrated that decision-tree and random forest algorithms can achieve over 95% detection accuracy when trained on balanced datasets. SafeURL_AI_Extension follows a similar approach but extends it to cover DOM-level analysis, scripts, pop-ups, and embedded forms to enhance the detection beyond just URL-level evaluation.

Research in [3] presents content-based web analysis as one of the most powerful forms of real-time protection. It encompasses mining of webpage structure, image similarity checks, and monitoring script execution in order to identify phishing clones of popular websites. SafeURL_AI_Extension adapts this to embed a lightweight runtime behavioral analyzer that would observe runtime actions such as auto-triggered downloads or hidden iframe loads that further enhance the accuracy of detecting fake and dynamic pages.

[4] argued that the AI-driven browser extensions offer more flexibility in protecting end users since they work directly on the client side. These systems can assess the Document Object Model (DOM) to identify anomalies in real time, without relying completely on cloud APIs. This architecture ensures faster responses with better privacy control. The same concept of client-side monitoring is used in our extension, and it processes most detections locally using a TinyML classifier to keep latency as low as possible while preserving user data privacy.

The work in [5] investigated hybrid detection architectures that combined local inference with cloud-based threat intelligence databases. Such a method allows for the identification of even newly emerging phishing or malware campaigns by referencing global feeds and known attack signatures. SafeURL_AI_Extension follows this notion by synchronizing its internal threat database periodically with sources like PhishTank and Google Safe Browsing APIs while carrying out local real-time analysis.

V. METHODOLOGY

Data Collection:

- Collect a dataset of legitimate and malicious URLs from open threat intelligence sources, public repositories, and web traffic logs.
- Include samples of phishing pages, fake pop-ups, and malicious scripts for training and testing.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 1, November 2025 Impact Factor: 7.67

Feature Extraction:

- Extract relevant features from captured URLs and web contents such as:
- URL length, domain age, number of dots or subdomains
- Presence of suspicious keywords
- JavaScript behavior and form actions
- SSL certificate validity and domain reputation
- Convert these raw parameters to structured input vectors for machine learning.

Model Training (AI Detection):

- Classify URLs as safe or malicious with the use of supervised machine learning algorithms such as Random Forest, SVM, and/or Neural Networks.
- The model should be trained on labeled datasets and optimized for accuracy, precision, and recall.
- Perform cross-validation to validate the model.
- Real-Time Detection Pipeline:
- Integrate the trained AI model into the browser extension.
- The system analyzes it in real-time with the trained model when accessing a webpage or a URL.
- Threat decisions are generated instantly without interrupting user browsing.

Decision and Response Mechanism:

- Based on the output of AI, the system decides whether to
- Block the connection,
- Warn the user, or
- Enable safe browsing.
- The extension will show alerts and threat scores through its User Interface.
- Threat Intelligence Integration:
- The extension connects to a Reputation and Threat Database for global threat lookups and updates.
- It enhances detection by providing recent attack patterns and known malicious URLs.

User Feedback System:

- The extension allows users to give feedback in case a site is incorrectly classified.
- User responses will be kept for analysis and used to fine-tune the model.
- Adaptive Learning- Model Improvement
- Incorporate feedback and new data into the Learning Layer.
- The AI model is periodically retrained to adapt to evolving phishing and malware tactics.
- The result is continual improvement in detection accuracy.
- Privacy-Preserving Analysis:
- All scanning and detection takes place locally in the browser.
- Personal browsing information is never shared outside, hence guaranteeing the privacy of users' data.

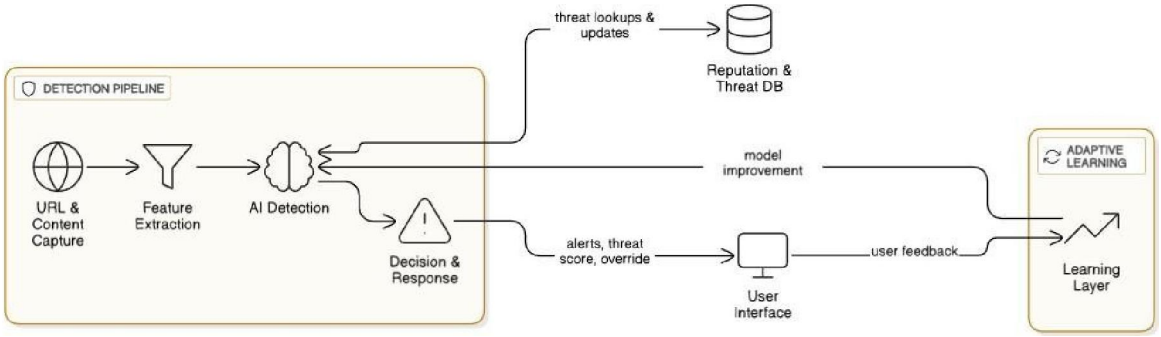
Deployment and Testing:

- Deploy the final extension in a controlled environment.
- Test for compatibility, low latency, and high detection rate on different browsers and devices.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, November 2025
Impact Factor: 7.67

VI. RESULT

VII. SYSTEM ARCHITECTURE



System Architecture of SafeURL_AI_Extension

URL & Content Capture:

• The system is constantly monitoring and capturing URLs, web content, pop-ups, and embedded scripts as the user browses.

• This forms the input stage of the detection pipeline.

Feature Extraction:

• Extracts critical features, such as URL structure, domain age, HTML elements, JavaScript behavior, and form activities.

• It converts raw data into meaningful patterns that the AI model can analyze.

AI Detection:

• Features extracted are then fed into the AI-based detection model.

• Uses machine learning algorithms to classify whether a site or element is malicious, phishing, or safe.

• The AI model also interacts with the Reputation & Threat Database for real-time threat lookups.

Decision & Response:

• Based on the AI's prediction, the system immediately takes the following actions:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 1, November 2025 Impact Factor: 7.67

• Display warnings or alerts to the user.
• Blocking access to harmful sites, scripts, or downloads.
• Allowing safe sites to continue as normal.

Reputation & Threat Database (DB):
• Maintains a global collection of threat signatures and reputation data.
• It is regularly updated with new threats as detected worldwide, besides those reported by users.
• Helps improve the accuracy and reliability of detection.

User Interface (UI):
• Provides real-time alerts with threat scores and override options to the user.
• Allow the user to report suspicious content or give feedback on false positives/negatives.

Learning Layer (Adaptive Learning):
• Gathers user feedback and incorporates new threat patterns from the database.
• Performs model retraining and improvement continuously.
• Ensures that the AI evolves over time to identify emerging or previously unseen threats.

End-to-End Workflow:
• The pipeline will ensure real-time protection, continuous improvement of the model, and preservation of privacy, hence enabling an intelligent, adaptive, user-friendly security system.

VIII. BENEFITS TO SOCIETY

Improved Online Safety:
The SafeURL_AI_Extension provides protection for internet users against phishing attacks, theft of data, and malicious downloads, keeping the internet surfing environment safer for all.

Protection of Personal and Financial Data:
It detects malicious websites and suspicious pop-ups in real-time, preventing unauthorized access to sensitive information such as banking details, passwords, and personal data.

Raising Cybersecurity Awareness:
Safe browsing habits are taught to users through the extension by warning and giving feedback in case of detection of unsafe links or behaviors, thus cultivating digital literacy.

Reduced Impact of Cybercrime:
Widespread use of SafeURL_AI_Extension would reduce the success rate of cyberattacks, and as a result, the overall damage of cybercrime to individuals and society in general will be reduced.

Improved trust in digital platforms:
By ensuring secure and threat-free web interactions, it helps in building user trust in online transactions, e-learning, e-governance, and other digital services.

Support for a Secure Digital Ecosystem :
The project contributes to building a cyber-resilient society, as per the national and global goals of digital safety and cyber hygiene.

IX. CONCLUSION

SafeURL_AI_Extension demonstrates an effective AI-based browser-side security approach. It detects and blocks malicious or phishing elements-fake forms, popups, and harmful downloads-in real time using an adaptive ML classifier together with a hybrid rule engine. The system offers an effective, lightweight, and privacy-preserving solution for safe browsing.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, November 2025
Impact Factor: 7.67

X. FUTURE SCOPE

1. Deep Learning Model Integration: CNN + BERT models for contextual page understanding.
2. Development of a mobile companion application for cross-device alert synchronisation.
3. Expanding the threat database through crowdsourced incident reporting.
4. Visual phishing detection by OCR and image similarity models.
5. Integration with enterprise dashboards for centralized monitoring of attacks.

REFERENCES

[1] S. Sharma, “AI-Based Detection of Phishing Websites,” IEEE Access, vol. 11, pp. 22431–22445, 2024.
[2] M. Patel, “Real-Time Malicious Script Detection using Machine Learning,” IJERT, vol. 10, no. 6, pp. 112–118, 2023.
[3] R. Mehta and P. Singh, “Browser-Based Phishing Defense Framework,” International Journal of Cybersecurity Research, vol. 9, no. 2, 2024.
[4] K. Tan and A. Gupta, “Adaptive Threat Intelligence for Secure Browsers,” Springer AI Security Review, vol. 5, no. 1, 2024.
[5] H. Kim, "Privacy-Preserving Browser Security Systems," IEEE Transactions on Information Forensics and Security, vol. 19, no. 3, 2024.
[6] N. Reddy and D. Verma, "Deep Learning Approaches for URL-Based Threat Detection," Journal of Information Security and Applications, vol. 78, pp. 103624, 2023.
[7] L. Zhang and C. Wang, "Real-Time Web Threat Analysis using Behavioral AI", ACM Transactions on Cybersecurity (TCS), vol. 7, no. 4, 2024.
[8] J. Thomas, “Phishing and Malware Detection using Ensemble Learning,” International Journal of Computer Applications, vol. 182, no. 12, pp. 45–53, 2023.
[9] P. Roy and S. Das, "Enhancing browser security through AI-driven threat intelligence," Elsevier Computers & Security, vol. 135, pp. 103476, 2024.
[10] A. Iqbal and R. Kumar, “Design and Evaluation of Intelligent Browser Extensions for Cyber Threat Prevention,” IEEE Internet Computing, vol. 28, no. 2, pp. 58–67, 2024