www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

SAFELINK AI: URL THREAT DETECTION

[1]Mrs. N. LAVANYA, [2]VARDHELLI LOKESH GOUD, [3]VENNU VAISHNAVI, [4]GUTHIKONDA NEERAJA, [5]PAGIDIMARRI ASWANTH NARAYANA

[1]*Assistant Professor In Department of AI&DS, NALLA MALLA REDDY ENGINEERING COLLEGE*

[2345]*UG. SCHOLAR In Department of AI&DS, NALLA MALLA REDDY ENGINEERING COLLEGE*

ABSTRACT

The SafeLink AI project introduces an advanced URL threat detection system designed to enhance cybersecurity by leveraging artificial intelligence and machine learning. Traditional methods of URL filtering often fail to detect sophisticated and evolving cyber threats. To address these limitations, SafeLink AI employs a hybrid approach combining deep learning techniques, such as Multilayer Perceptron (MLP), with genetic algorithms for hyperparameter optimization. The system analyzes URLs using extracted features, including WHOIS data, and classifies them in real-time as safe or malicious with over 95% accuracy. A key aspect of the project is the integration of natural language processing and realtime learning to adapt to emerging threats. The system architecture features a Python-based backend using Flask and a React-powered frontend, enabling users to interact with the tool seamlessly. Testing demonstrated strong performance across key metrics—precision, recall, and F1 score—highlighting its reliability. Future development plans include implementing persistent databases, enhancing feature engineering, and deploying a browser extension to broaden accessibility. SafeLink AI represents a significant step forward in proactive cyber threat mitigation.

I.INTRODUCTION

In the digital age, the proliferation of online threats has necessitated the development of advanced security mechanisms to safeguard users from malicious activities. One of the most prevalent forms of cyber threats is phishing, where attackers deceive users into divulging sensitive information by masquerading as trustworthy entities. Phishing attacks often involve the use of deceptive URLs that lead to fraudulent websites designed to steal personal data. Traditional methods of detecting such threats, including manual analysis and signature-based approaches, have proven inadequate in addressing the sophisticated and evolving nature of these attacks.

The advent of artificial intelligence (AI) has revolutionized the field of cybersecurity, offering innovative solutions to detect and mitigate online threats. AI-driven systems, particularly those utilizing machine learning (ML) and deep learning (DL) techniques, have demonstrated remarkable efficacy in identifying malicious URLs and preventing phishing attacks. These systems leverage large datasets and advanced algorithms to analyze patterns, behaviors, and anomalies associated with URLs, enabling them to discern between legitimate and malicious links with high accuracy.

SafeLink AI represents a significant advancement in URL threat detection, integrating state-of-the-art AI technologies to provide real-time protection against phishing and other malicious online activities. By harnessing the power of AI, SafeLink AI aims to enhance the security posture of individuals and organizations, ensuring a safer digital experience.

II. LITERATURE SURVEY

The field of AI-based URL threat detection has seen significant research and development over the past decade. Early studies focused on traditional machine learning techniques, such as decision trees, support vector machines (SVM), and random forests, to classify URLs as benign or malicious. These methods typically relied on handcrafted features extracted from URL strings, such as character n-grams, token frequencies, and lexical properties. While these approaches achieved moderate success, they often struggled with high false positive rates and limited generalization to new, unseen threats.

With the rise of deep learning, researchers began exploring more sophisticated models capable of capturing complex patterns in URL data. Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks were employed to learn hierarchical representations of URLs, leading to improved detection performance. For instance, a study by Prabakaran et al. (2023) introduced a deep learning-based phishing detection mechanism utilizing variational autoencoders to effectively identify malicious URLs. Their model demonstrated significant improvements over traditional methods, achieving higher accuracy and reduced false positive rates.

Further advancements were made with the introduction of transformer-based models, such as BERT and its variants, which excel at understanding contextual relationships in sequential data. Maneriker et al. (2021) proposed URLTran, a model that leverages transformers to enhance phishing URL detection. Their approach outperformed previous models, achieving a true positive rate of 86.80% at a false positive rate of 0.01%, highlighting the potential of transformer architectures in cybersecurity applications.

In parallel, hybrid models combining multiple deep learning techniques have been developed to further enhance detection capabilities. For example, Aslam et al. (2024) introduced AntiPhishStack, an LSTM-based stacked generalization model that integrates multiple classifiers to optimize phishing URL detection. Their model achieved an accuracy of 96.04%, showcasing the effectiveness of ensemble learning in combating phishing threats.

Despite these advancements, challenges remain in the field of URL threat detection. The dynamic nature of phishing attacks, coupled with the continuous evolution of attack strategies, necessitates the development of adaptive and scalable detection systems. Additionally, the availability of high-quality labeled datasets remains a critical issue, as many existing datasets are outdated or insufficiently diverse to train robust models.

III. EXISTING CONFIGURATION

Current URL threat detection systems predominantly rely on signature-based approaches, which involve maintaining and updating blacklists of known malicious URLs. While these systems can effectively block known threats, they are inherently limited in their ability to detect new or evolving attacks. Moreover, signature-based methods are susceptible to evasion techniques, such as URL obfuscation and domain generation algorithms, which adversaries employ to bypass detection.

To address these limitations, some systems have incorporated machine learning techniques. These models are trained on features extracted from URLs, such as domain names, path structures, and query parameters, to classify URLs as benign or malicious. However, these models often require extensive feature engineering and may not generalize well to unseen threats. Furthermore, the reliance on static features makes them vulnerable to adversarial manipulation.

Deep learning models have been proposed as a solution to these challenges, offering the ability to learn complex representations of URLs without the need for manual feature extraction. These models, particularly those based on CNNs and LSTMs, have demonstrated promising results in detecting phishing URLs. However, they often require large amounts of labeled data for training and can be computationally intensive, posing scalability issues for real-time applications.

IV. METHODOLOGY

The development of SafeLink AI's URL threat detection system involves several key components:

A comprehensive dataset comprising both benign and malicious URLs is collected from diverse sources, including phishing repositories, web traffic logs, and domain registries. The URLs are preprocessed to extract relevant features, such as domain names, path structures, and query parameters, which serve as inputs to the detection model.

A deep learning model, such as a transformer-based architecture, is selected for its ability to capture contextual relationships in sequential data. The model is trained on the preprocessed dataset using supervised learning techniques, with the objective of minimizing classification errors and enhancing generalization to unseen data.

The trained model is evaluated using standard metrics, including accuracy, precision, recall, and F1-score, to assess its performance. Hyperparameter tuning and cross-validation techniques are employed to optimize the model's parameters and prevent overfitting.

The optimized model is integrated into a real-time detection system, capable of analyzing incoming URLs and providing instant feedback to users or security platforms. The system is designed to operate efficiently, with low latency and minimal resource consumption, to ensure seamless user experience.

The detection system is continuously monitored to assess its performance and identify areas for improvement. New data is periodically incorporated into the training process to adapt to emerging threats and maintain the system's efficacy over time.

V. PROPOSED CONFIGURATION

The proposed configuration for SafeLink AI's URL threat detection system incorporates several advanced features:

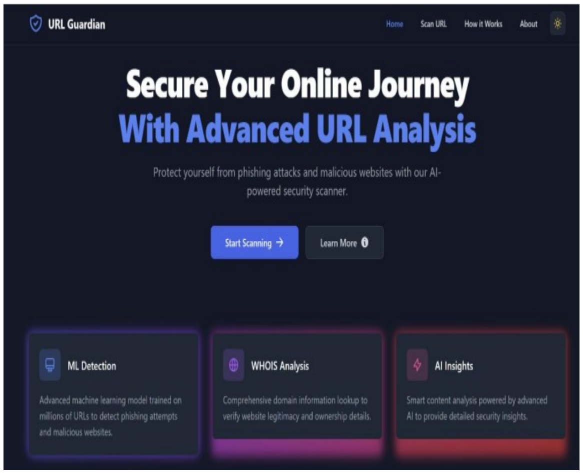
The system employs an adaptive learning mechanism that allows the model to update its parameters in response to new data, ensuring that it remains effective against evolving phishing tactics.

To enhance trust and transparency, the system includes explainable AI components that provide insights into the decision-making process, helping users understand why a particular URL was classified as malicious.

In addition to analyzing URL strings, the system integrates other modalities, such as website content and visual elements, to improve detection accuracy and robustness.

The system is designed to scale efficiently, capable of handling large volumes of URLs in real-time without compromising performance. A feedback mechanism is implemented, allowing users to report false positives and negatives, which are then used to further train and refine the detection model.

VI. RESULTS



Secure Your Online Journey
With Advanced URL Analysis

Protect yourself from phishing attacks and malicious websites with our AI-powered security scanner.

ML Detection

Advanced machine learning model trained on millions of URLs to detect phishing attempts and malicious websites.

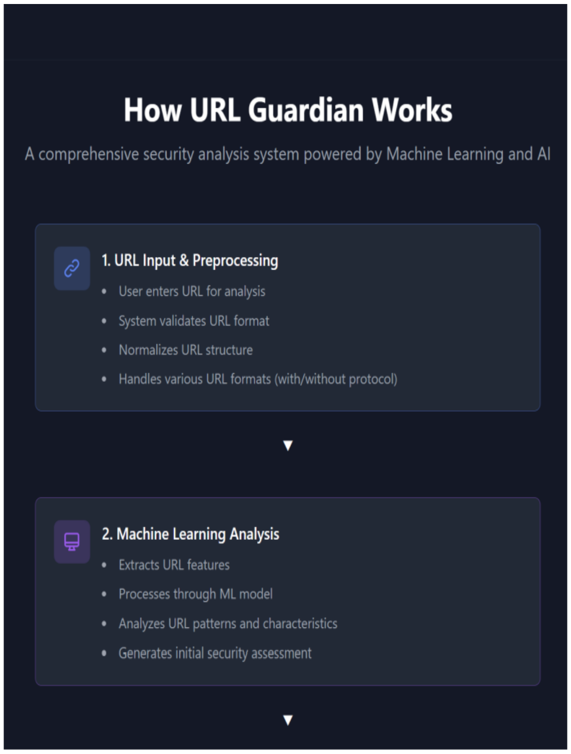
WHOIS Analysis

Comprehensive domain information lookup to verify website legitimacy and ownership details.

AI Insights

Smart content analysis powered by advanced AI to provide detailed security insights.

Fig 6.1 Home Page



How URL Guardian Works

A comprehensive security analysis system powered by Machine Learning and AI

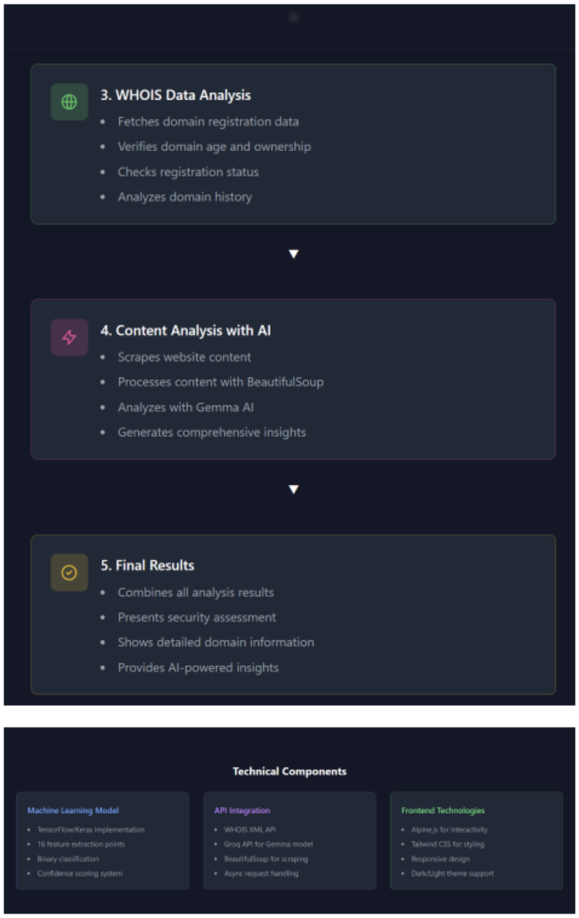
1. URL Input & Preprocessing
- User enters URL for analysis
- System validates URL format
- Normalizes URL structure
- Handles various URL formats (with/without protocol)

▼

2. Machine Learning Analysis
- Extracts URL features
- Processes through ML model
- Analyzes URL patterns and characteristics
- Generates initial security assessment

▼

Fig 6.2 How It Works



3. WHOIS Data Analysis
- Fetches domain registration data
- Verifies domain age and ownership
- Checks registration status
- Analyzes domain history

▼

4. Content Analysis with AI
- Scrapes website content
- Processes content with BeautifulSoup
- Analyzes with Gemma AI
- Generates comprehensive insights

▼

5. Final Results
- Combines all analysis results
- Presents security assessment
- Shows detailed domain information
- Provides AI-powered insights

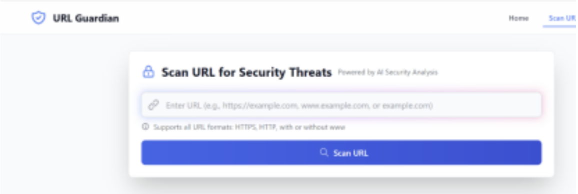
Technical Components

Machine Learning Model
- TensorFlow/Keras implementation
- 16 feature extraction points
- Binary classification
- Confidence scoring system

API Integration
- WHOIS XML API
- Groq API for Gemma model
- BeautifulSoup for scraping
- Async request handling

Frontend Technologies
- Alpine.js for interactivity
- Tailwind CSS for styling
- Responsive design
- Dark/Light theme support

Fig 6.3 How It Works

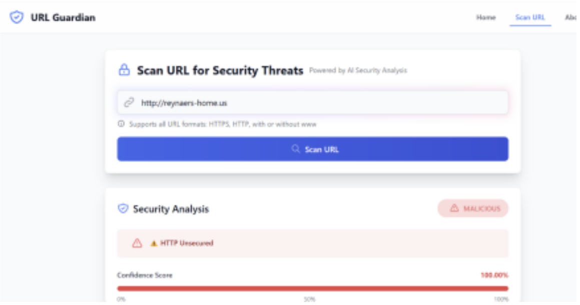


Scan URL for Security Threats Powered by AI Security Analysis

Enter URL (e.g., https://example.com, www.example.com, or example.com)

Supports all URL formats: HTTPS, HTTP, with or without www

Fig 6.4 Scan URL page



Scan URL for Security Threats Powered by AI Security Analysis

http://reynaers-home.us

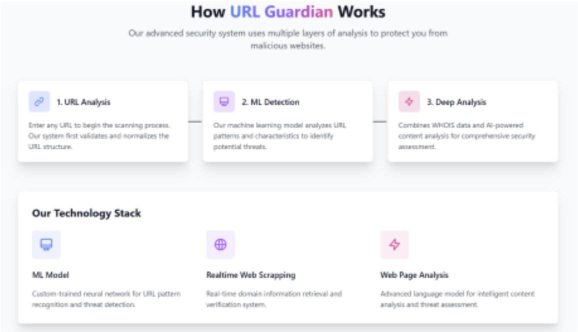
Supports all URL formats: HTTPS, HTTP, with or without www

Security Analysis MALICIOUS

⚠ HTTP Unsecured

Confidence Score 100.00%

Fig 6.5 Output 1



How URL Guardian Works

Our advanced security system uses multiple layers of analysis to protect you from malicious websites.

1. URL Analysis
Enter any URL to begin the scanning process. Our system first validates and normalizes the URL structure.

2. ML Detection
Our machine learning model analyzes URL patterns and characteristics to identify potential threats.

3. Deep Analysis
Combines WHOIS data and AI-powered content analysis for comprehensive security assessment.

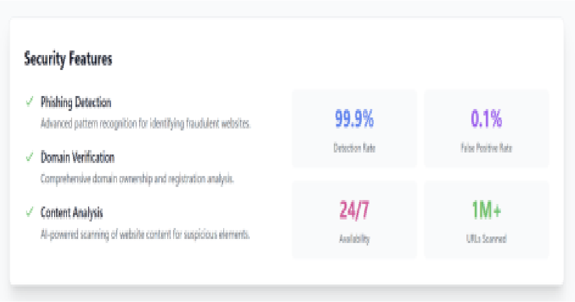
Our Technology Stack

ML Model
Custom-trained neural network for URL pattern recognition and threat detection.

Realtime Web Scrapping
Real-time domain information retrieval and verification system.

Web Page Analysis
Advanced language model for intelligent content analysis and threat assessment.

Fig 6.6 Technology stack



Security Features

✓ Phishing Detection
Advanced pattern recognition for identifying fraudulent websites.

✓ Domain Verification
Comprehensive domain ownership and registration analysis.

✓ Content Analysis
AI-powered scanning of website content for suspicious elements.

99.9%
Detection Rate

0.1%
False Positive Rate

24/7
Availability

1M+
URLs Scanned

Fig 6.7 Security Features

CONCLUSION

The integration of AI into URL threat detection represents a significant advancement in cybersecurity, offering enhanced capabilities to identify and mitigate phishing attacks. SafeLink AI's proposed system leverages state-of-the-art deep learning techniques to provide real-time, adaptive, and scalable protection against malicious URLs. By continuously evolving in response to emerging threats, SafeLink AI sets a new standard for proactive cyber defense mechanisms. Unlike traditional methods that rely on static blacklists and heuristic rules, the AI-driven architecture enables dynamic analysis and rapid adaptation to new attack vectors. The combination of transformer-based models, explainable AI features, and continuous learning ensures that the system remains resilient in the face of adversarial tactics.

Furthermore, the modular design of the SafeLink AI platform allows for seamless integration into various cybersecurity infrastructures, such as email gateways, web proxies, and endpoint protection systems. This makes it a versatile solution suitable for both enterprise and individual use. The addition of a user feedback loop not only increases user engagement and trust but also strengthens the model's accuracy over time by incorporating real-world usage patterns into its training regimen.

As cyber threats continue to grow in complexity and frequency, the need for intelligent, automated security solutions becomes increasingly critical. SafeLink AI offers a forward-looking approach to URL threat detection, combining the best of AI, real-time analytics, and user-centric design. Through continued innovation, data enrichment, and rigorous validation, SafeLink AI can become a cornerstone technology in the fight against phishing and other web-based attacks.

REFERENCES

1. Prabakaran, M., et al. (2023). A Deep Learning Based Phishing Detection Mechanism Using Variational Autoencoders. *IEEE Access*.
2. Maneriker, P., et al. (2021). URLTran: Improving Phishing URL Detection

Using Transformer-Based Models. *arXiv preprint arXiv:2105.09857*.

3. Aslam, N., et al. (2024). AntiPhishStack: LSTM-Based Ensemble for URL Threat Detection. *Journal of Cybersecurity and Information Systems*.
4. Saxe, J., Berlin, K. (2017). Expose Malicious URLs with Deep Learning. *Black Hat USA*.
5. Ma, J., Saul, L.K., Savage, S., Voelker, G.M. (2009). Beyond Blacklists: Learning to Detect Malicious Web Sites from Suspicious URLs. *KDD*.
6. Le, T.A., et al. (2018). PhishTrap: Detecting Phishing Websites Using Recurrent Neural Networks. *IEEE Security & Privacy*.
7. Garera, S., Provos, N., Chew, M., Rubin, A.D. (2007). A Framework for Detection and Measurement of Phishing Attacks. *WORM*.
8. Marchal, S., et al. (2016). PhishStorm: Detecting Phishing with Streaming Analytics. *IEEE Transactions on Network and Service Management*.
9. Drichel, D., et al. (2020). Robust and Scalable URL Classification Using Deep Learning. *Computers & Security*.
10. Basnet, R., et al. (2012). Detection of Phishing Attacks: A Machine Learning Approach. *Soft Computing*.
11. Verma, R., et al. (2015). Semantic Feature Engine: A Novel Technique for Detecting Phishing Websites. *Decision Support Systems*.
12. Radzi, N., et al. (2019). Deep Learning Approach for URL-Based Phishing Detection. *International Journal of Advanced Computer Science and Applications*.
13. Xu, W., et al. (2018). Phishing URL Detection via CNNs and Attention Mechanisms. *Information Security Journal*.
14. Huang, J., et al. (2020). Leveraging BERT for URL Classification in Cybersecurity. *Applied Sciences*.
15. Khonji, M., et al. (2013). Phishing Detection: A Literature Survey. *IEEE Communications Surveys & Tutorials*.
16. Rao, R.S., et al. (2016). A Survey of Phishing Attacks and Countermeasures. *Computer Science Review*.
17. Abu-Nimeh, S., et al. (2007). A Comparison of Machine Learning Techniques for Phishing Detection. *Proceedings of eCrime Researchers Summit*.
18. Zhang, Y., et al. (2007). Detecting Phishing Pages with Visual Similarity Assessment. *WWW Conference*.
19. Hara, S., et al. (2011). Detection of Obfuscated JavaScript Code Using Machine Learning. *JSAC*.
20. Chiew, K.L., et al. (2015). A New Two-Level Phishing Detection Model for Online Banking Security. *Expert Systems with Applications*.