ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26mar1626

AI-Driven Phishing Detection & Honeypot-Based Analysis System

S. Ezhil Savier[1]; T. Balaguru[2]; E. Dhanushri[3]; A. Soumya[4]

1Department of AI & DS, AVS Egineering College, Salem, Tamil Nadu, India
2Department of AI & DS, AVS Egineering College, Salem, Tamil Nadu, India
3Department of AI & DS, AVS Egineering College, Salem, Tamil Nadu, India
4Department of AI & DS, AVS Egineering College, Salem, Tamil Nadu, India

Publication Date: 2026/04/06

Abstract: This paper presents the design and development of a consent-based AI-driven system for detecting phishing and social engineering attacks from spam emails. The proposed system integrates official email APIs to access only spam or phishing-flagged emails with explicit user consent, ensuring privacy protection and ethical compliance. A hybrid detection model combining Natural Language Processing, metadata analysis, and rule-based techniques is employed to classify emails and generate a risk score. The system introduces a Phishing Intent Timeline Reconstruction module to identify attack stages such as lure creation, delivery, exploitation, and credential harvesting, and explains potential consequences in clear and user-friendly language. Additionally, a Phishing DNA engine extracts structural and behavioral features including HTML patterns, redirection chains, and hosting attributes to cluster related phishing campaigns and detect phishing kit reuse. A secure backend honeypot environment safely interacts with suspicious links in an isolated environment to observe attacker behavior and infrastructure patterns without collecting real credentials. The system also incorporates an Explain-Before-Click interface and a local- language awareness module to enhance user understanding and prevention. The proposed solution improves phishing detection accuracy while maintaining ethical standards and practical feasibility for academic implementation.

Keywords: *Phishing Detection, Artificial Intelligence, Honeypot, Social Engineering, Cybersecurity.*

How to Cite: S. Ezhil Savier; T. Balaguru; E. Dhanushri; A. Soumya (2026) AI-Driven Phishing Detection & Honeypot-Based Analysis System. *International Journal of Innovative Science and Research Technology,* 11(3), 3286-3288. https://doi.org/10.38124/ijisrt/26mar1626

I. INTRODUCTION

Phishing and social engineering attacks exploit human psychology to steal sensitive information through deceptive emails and malicious links. Traditional spam filters based on keyword matching and blacklists are insufficient against evolving threats. This paper proposes a consent-based AI-driven phishing detection framework integrating machine learning, metadata analysis, and honeypot-based monitoring to provide detection, behavioral analysis, and user awareness through an Explain-Before-Click mechanism.

II. PROPOSED SYSTEM ARCHITECTURE

The system includes an Email Analysis module, an AI Detection module, and a Behavioral Monitoring module. Email metadata is accessed via official APIs with user consent. NLP and structural feature extraction generate a risk score, while the Phishing Intent Timeline and Phishing DNA components analyze attack stages and campaign similarities. A sandboxed honeypot safely interacts with suspicious links to capture infrastructure patterns.

III. RELATED WORK

Existing phishing detection systems using ML, NLP, & URL-based analysis to classify malicious emails. Honeypots are employed to observe attacker infrasctructure & behaviour. However, most solutions focus primarily on detection accuracy & lack integrated behavioral analysis & user-centric explanation features, which this work addresses.

IV. METHODOLOGY

A hybrid detection approach combines rule-based filtering and machine learning classification. Textual and metadata features are extracted to compute a weighted risk score. Suspicious URLs are executed in an isolated honeypot environment, and similarity-based clustering groups related phishing campaigns.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26mar1626

➢ *Architecture*

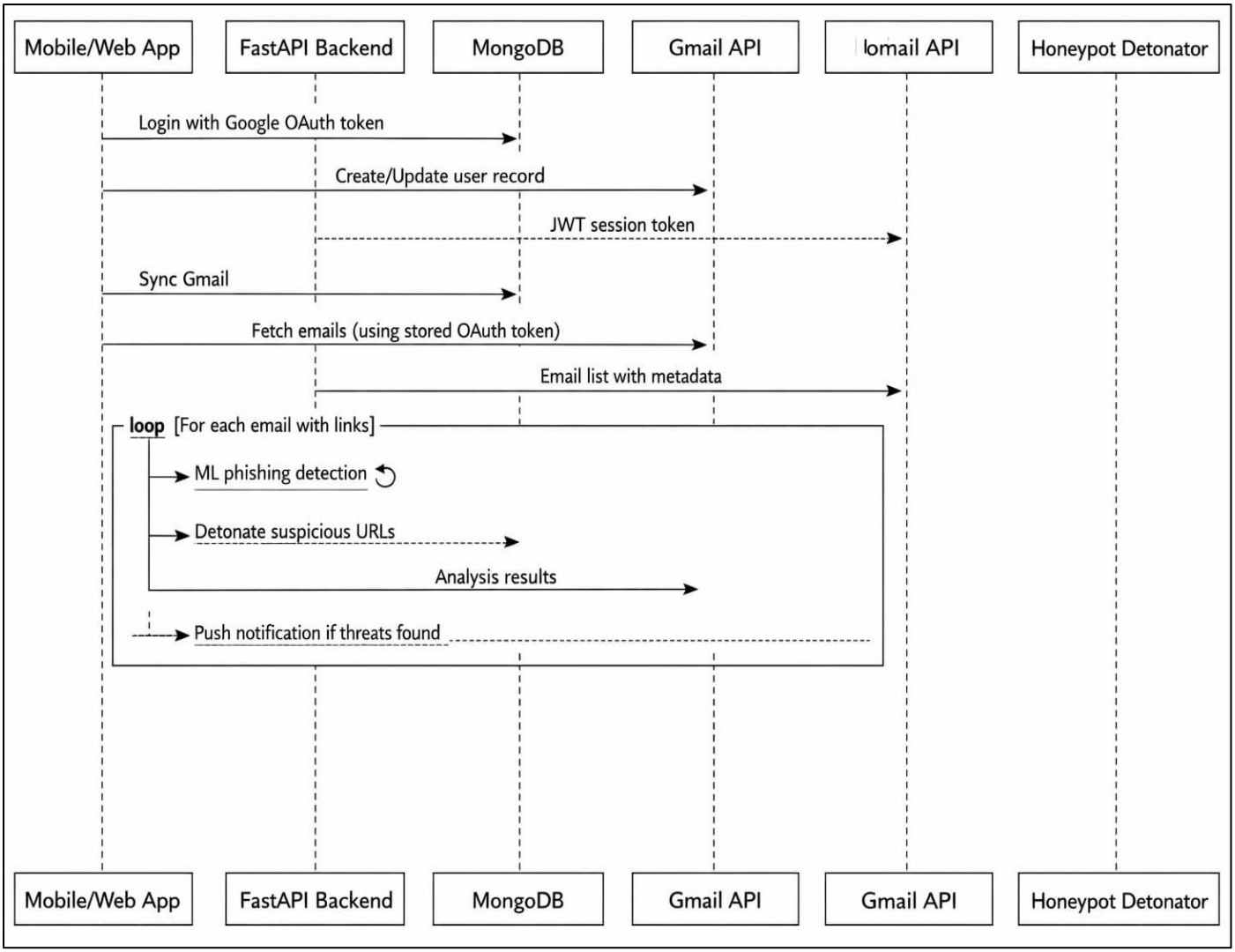

Fig 1 Architecture

V. IMPLEMENTATION & RESULTS

The framework is implemented using Python with read- only API access to ensure privacy compliance. Experimental evaluation indicates improved detection performance through combined linguistic and structural features. The Phishing DNA and Explain-Before-Click modules enhance campaign analysis and user awareness.

VI. ETHICAL CONSIDERATIONS

The system operates with explicit user consent and processes only necessary indicators. No credentials are stored, and the honeypot performs passive observation without intrusive actions.

VII. CONCLUSION

The proposed AI-driven phishing detection and behavioral analysis system improves detection accuracy while maintaining ethical standards and promoting user awareness, with future enhancements.

ACKNOWLEDGEMENT

The authors sincerely thank Mrs.N.SANKARI, Assistant Professor, Department of Artificial Intelligence and Data Science, AVS Engineering College, Salem, Tamil Nadu, India, for valuable guidance, continuous support, and technical insights provided throughout the development of this work.

REFERENCES

[1]. Y. Zhang, J. Hong and L. Cranor, "Cantina: A Content- Based Approach to Detecting Phishing Web Sites," *Proc. 16th Int. Conf. World Wide Web (WWW)*, pp. 639–648, 2007. Available: https://doi.org/10.1145/1242572.1242653

[2]. D. R. Thomas, C. Grier and V. Paxson, "Adapting Honeypots for Web Security Education," *Proc. 17th ACM Conf. Computer and Communications Security (CCS)*, pp. 50–61, 2010. Available: https://doi.org/10.1145/1866307.1866315

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26mar1626

[3]. S. Abu-Nimeh, D. Nappa, X. Wang and S. Nair, "A Comparison of Machine Learning Techniques for Phishing Detection," *Proc. IEEE 10th Int. Conf. Machine Learning Applications (ICMLA)*, pp. 254–260, 2011. PDF: https://ieeexplore.ieee.org/document/6147468

[4]. A. Khonji, Y. Iraqi and A. Jones, "Phishing Detection: A Literature Survey," *IEEE Commun. Surveys & Tutorials*, vol. 15, no. 4, pp. 2091–2121, 2013. Available: https://ieeexplore.ieee.org/document/6472480

[5]. R. Chandrasekaran, G. Narayanan and S. Upadhyaya, "Phishing Email Detection Based on Structural Properties," *Proc. 5th Int. Conf. Email and Anti-Spam (CEAS)*, 2008. Link: https://www.ceas.cc/papers-2008/78.pdf