International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

REAL TIME AIML PHISHING DETECTION AND PREVENTION SYSTEM

[1]D.ARUNA, [2]K NAGA ANJALI, [3]O BABU SURESH, [4]CH RUPA PAVANI, [5]K SAI NEERAJ

[1]Assistant Professor, [2345]Students, Department of Computer Science and Engineering, SRI VASAVI INSTITUTE OF ENGINEERING & TECHNOLOGY, NANDAMURU, ANDHRA PRADESH

ABSTRACT

Phishing attacks represent one of the most persistent and dangerous cybersecurity threats in the digital era, targeting individuals and organizations by impersonating legitimate websites to steal confidential information such as login credentials, banking details, and personal data. Traditional phishing detection methods primarily rely on blacklist databases, heuristic rules, and manual monitoring systems. However, these approaches often fail to detect newly generated phishing websites, also known as zero-day attacks, due to their dynamic and adaptive nature. To overcome these limitations, this research proposes a Real Time Artificial Intelligence and Machine Learning (AI/ML) based Phishing Detection and Prevention System capable of identifying malicious websites through automated feature analysis and predictive modeling. The system collects multiple features from website URLs, domain characteristics, page content, and security indicators such as SSL certificates. These features are then processed using machine learning algorithms to classify websites as legitimate or phishing in real time. The proposed system integrates data preprocessing, feature extraction, model training, and prediction modules to ensure efficient detection and fast response. By leveraging supervised learning techniques, the system continuously improves its accuracy by learning from large datasets of legitimate and phishing URLs. Furthermore, the real-time detection framework allows early identification of malicious websites before users interact with them, thereby reducing the risk of credential theft and financial loss. Experimental evaluation demonstrates that the proposed AI/ML-based approach significantly improves detection accuracy, reduces false positives, and provides faster identification compared to traditional blacklist approaches. This system contributes to strengthening cybersecurity infrastructure by offering an intelligent, scalable, and adaptive solution for phishing detection and prevention in modern web environments.

Keywords: Phishing Detection, Machine Learning, Cybersecurity, URL Analysis, Real-Time Detection, Artificial Intelligence

I INTRODUCTION

The rapid expansion of internet services and online transactions has significantly increased the risk of cyber threats, among which phishing attacks have emerged as one of the most common and damaging forms of cybercrime. Phishing is a deceptive technique used by attackers to trick users into revealing sensitive information such as passwords, banking credentials, and personal identification details by impersonating legitimate websites or trusted entities. These attacks are typically carried out through fake websites, fraudulent emails, and

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

malicious links that appear authentic to unsuspecting users. According to recent cybersecurity studies, phishing attacks continue to grow both in volume and complexity due to advancements in web technologies and automation tools. Attackers are increasingly capable of creating realistic replicas of legitimate websites within minutes, making it difficult for traditional security mechanisms to detect them effectively. Conventional detection methods rely heavily on blacklist databases that store previously reported phishing URLs, but these approaches are limited because they cannot detect newly created phishing websites in real time. As a result, many users become victims before the malicious sites are identified and blocked. Researchers have therefore focused on developing intelligent detection mechanisms that can automatically analyze website characteristics and identify suspicious patterns using advanced computational techniques [1][2][3][4][5][6][7][8][9][10][11][12][13][14][15].

Artificial Intelligence and Machine Learning technologies have recently emerged as powerful tools for enhancing cybersecurity systems by enabling automated detection of complex attack patterns. Machine learning algorithms can analyze large volumes of web data and identify subtle features that differentiate legitimate websites from phishing sites. These features may include URL structure, domain age, page content, SSL certificate validity, hyperlink patterns, and website traffic behavior. By training predictive models on large datasets of phishing and legitimate websites, machine learning systems can learn to classify new websites with high accuracy and speed. Moreover, AI-based detection systems can adapt to evolving phishing strategies by continuously updating their learning models. Real-time phishing detection systems further improve security by analyzing websites as soon as users attempt to access them, thereby preventing interaction with malicious platforms. Integrating AI/ML models with browser extensions, web gateways, and security applications provides an additional layer of protection for users and organizations. Despite significant progress in this area, challenges such as feature selection, model accuracy, and real-time processing still require further research and optimization. Therefore, this study proposes a real-time AI/ML-based phishing detection and prevention system that analyzes multiple website features and applies machine learning algorithms to accurately classify websites and reduce cybersecurity risks in modern digital environments [16][17][18][19][20][21][22][23][24][25][26][27][28][29][30].

II LITERATURE SURVEY

Numerous research studies have investigated phishing detection techniques using different computational and analytical approaches. Early detection methods primarily relied on blacklist systems and heuristic rule-based techniques to identify malicious websites. Blacklist approaches maintain a database of previously detected phishing URLs and block user access when a match occurs. Although this method is simple and widely implemented in modern browsers, it suffers from significant limitations because it cannot detect newly generated phishing websites. To address these challenges, researchers introduced heuristic-based detection models that analyze structural characteristics of URLs such as length, use of special characters, suspicious domain names, and redirection behavior. These techniques improved detection capabilities but still produced high false positive rates due to the static nature of predefined

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

rules. As cyber attackers continued to evolve their tactics, researchers began exploring machine learning methods for phishing detection. Machine learning algorithms such as Decision Trees, Random Forest, Support Vector Machines, and Logistic Regression have been widely used to classify websites based on extracted features from URLs and webpage content. These models demonstrated improved accuracy and adaptability compared to traditional approaches because they could learn complex relationships within large datasets of phishing and legitimate websites [1][2][3][4][5][6][7][8][9][10][11][12][13][14][15].

Recent research has increasingly focused on integrating advanced machine learning and deep learning techniques to enhance phishing detection systems. Studies have shown that ensemble learning algorithms such as Random Forest and Gradient Boosting can significantly improve classification accuracy by combining multiple decision models. In addition, deep learning models such as Convolutional Neural Networks and Recurrent Neural Networks have been explored for detecting phishing patterns directly from raw URL strings and webpage content. These models are capable of capturing hidden relationships and contextual information that traditional machine learning methods may overlook. Researchers have also proposed hybrid approaches that combine heuristic rules, blacklist verification, and machine learning algorithms to create multi-layered detection systems. Such systems offer improved performance by integrating multiple security mechanisms into a single framework. Furthermore, the implementation of real-time phishing detection through browser plugins and cloud-based security platforms has become an active area of research. These systems monitor web traffic continuously and provide instant alerts when suspicious websites are detected. Despite these advancements, challenges remain in achieving high detection accuracy, reducing computational overhead, and maintaining real-time performance. Therefore, further research is required to develop intelligent phishing detection systems that are scalable, adaptive, and capable of identifying evolving attack strategies in modern cybersecurity environments [16][17][18][19][20][21][22][23][24][25][26][27][28][29][30].

III METHODOLOGY

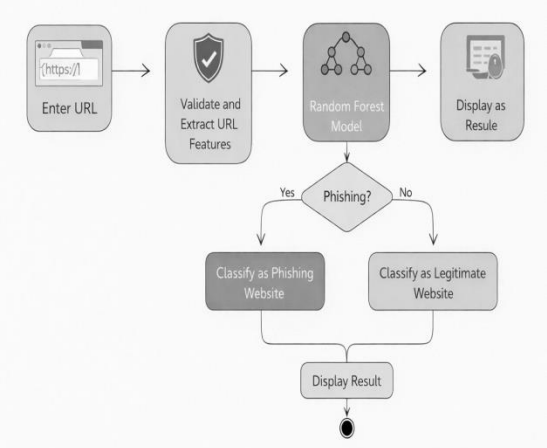
The proposed Real Time AI/ML Phishing Detection and Prevention System follows a structured methodology consisting of data collection, preprocessing, feature extraction, model training, and real-time prediction stages. Initially, a dataset containing both legitimate and phishing website URLs is collected from publicly available cybersecurity repositories and phishing databases. This dataset serves as the training and testing foundation for the machine learning model. During the preprocessing stage, the collected data is cleaned to remove duplicate entries, incomplete records, and inconsistent values to ensure reliable model training. After preprocessing, feature extraction is performed to identify relevant characteristics that distinguish phishing websites from legitimate ones. These features include URL length, presence of special characters, number of subdomains, domain age, HTTPS usage, SSL certificate validity, redirection patterns, and abnormal hyperlink behavior. Once the features are extracted, the dataset is divided into training and testing subsets to evaluate model performance. Various machine learning algorithms such as Decision Tree, Random Forest, Support Vector Machine, and Logistic Regression are trained on

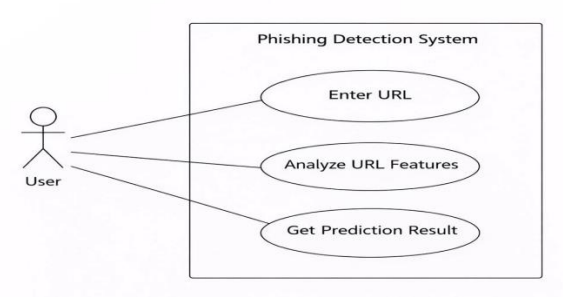
International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

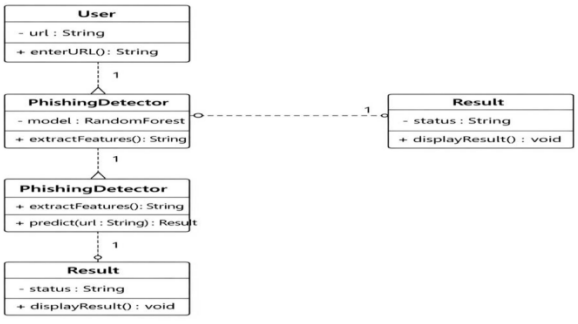
the dataset to identify patterns associated with phishing attacks. Among these models, the algorithm with the highest accuracy and lowest false positive rate is selected for deployment in the real-time detection system. The trained model is then integrated into a web application or browser-based framework where it continuously analyzes URLs entered by users. When a user attempts to access a website, the system extracts its features in real time and feeds them into the trained model for classification. Based on the prediction result, the system either allows the user to proceed if the website is legitimate or issues a warning if the site is identified as phishing. This methodology enables automated detection of malicious websites and enhances user protection against phishing attacks.

IV SYSTEM DESIGN

The system design of the Real Time AI/ML Phishing Detection and Prevention System is structured into several functional modules that work together to provide efficient detection and prevention of phishing websites. The architecture primarily consists of the user interface module, data collection module, feature extraction module, machine learning model module, and the real-time detection module. The user interface acts as the interaction layer through which users can enter or access URLs that need to be verified. This interface can be implemented as a web application or integrated into a browser extension to ensure accessibility and convenience. The data collection module gathers datasets of phishing and legitimate websites from reliable cybersecurity sources. These datasets form the basis for training the machine learning algorithms. After the data is collected, the feature extraction module analyzes various attributes of the URLs and web pages, including domain information, URL structure, SSL certificate status, and hyperlink patterns. These features are then converted into numerical representations suitable for machine learning model training. This module plays a critical role because the accuracy of the detection system largely depends on the quality and relevance of the extracted features.



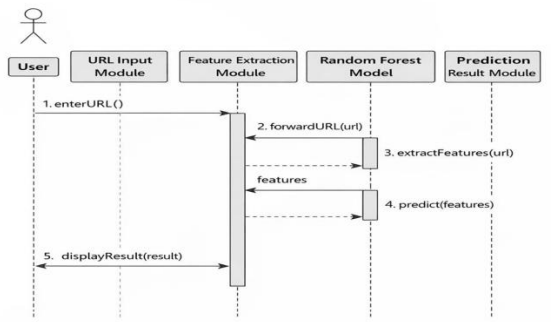


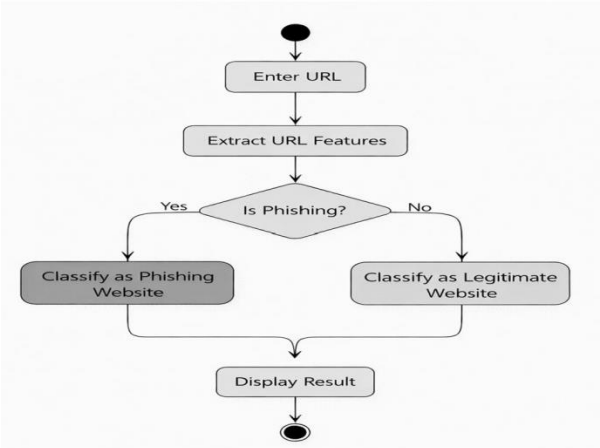


The machine learning model module is responsible for training and deploying the classification

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

algorithm used for phishing detection. In this module, the extracted features are used to train machine learning algorithms capable of identifying patterns associated with phishing attacks. The trained model is then evaluated using testing datasets to measure performance metrics such as accuracy, precision, recall, and F1-score. Once the optimal model is selected, it is integrated into the real-time detection module. The real-time detection module continuously monitors URLs accessed by users and performs instant classification using the trained machine learning model. If a website is detected as phishing, the system immediately generates a warning message to alert the user and prevent further interaction with the malicious site. Additionally, the system can log suspicious URLs for further analysis and future training improvements. This modular system design ensures scalability, maintainability, and efficient performance, enabling the system to detect phishing websites quickly and accurately while minimizing computational overhead.





V PROPOSED SYSTEM

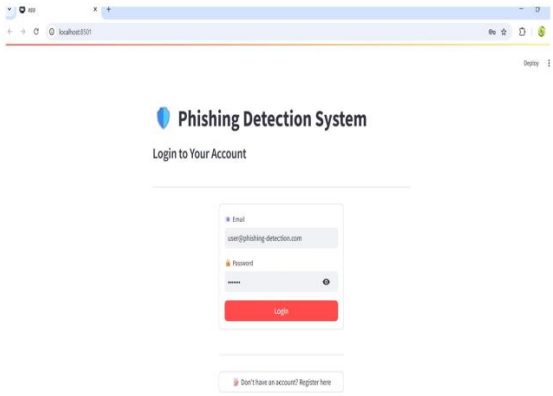
The proposed system introduces a real-time AI/ML-based phishing detection framework designed to overcome the limitations of traditional phishing detection techniques. Unlike conventional blacklist-based systems that rely solely on previously identified phishing URLs, the proposed system uses machine learning algorithms to analyze website characteristics and detect malicious patterns dynamically. The system begins by collecting a large dataset of both legitimate and phishing websites from trusted cybersecurity repositories. This dataset is then processed through preprocessing techniques to remove noise and inconsistencies. Feature extraction techniques are applied to identify important attributes that distinguish phishing websites from legitimate ones. These attributes include URL structure, domain age, presence of suspicious characters, HTTPS security indicators, redirection behavior, and hyperlink distribution within web pages. After extracting these features, the data is used to train machine learning models capable of recognizing phishing patterns. Multiple algorithms are evaluated during the training phase to determine the most effective model for accurate classification.

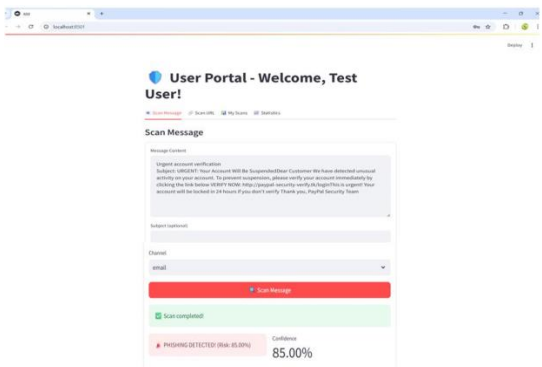
International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

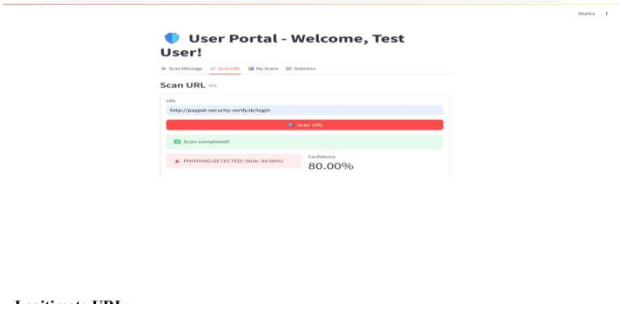
The real-time detection capability is the most significant feature of the proposed system. Once the machine learning model is trained, it is integrated into a web application or browser extension that continuously monitors websites accessed by users. Whenever a user attempts to open a webpage, the system automatically extracts its features and sends them to the trained machine learning model for classification. Based on the prediction result, the system determines whether the website is legitimate or phishing. If the site is identified as phishing, the system immediately displays a warning message and prevents the user from entering sensitive information. Additionally, the system stores detected phishing URLs in a database to enhance future detection performance. This adaptive learning mechanism allows the model to continuously improve its accuracy as new phishing patterns emerge. By combining machine learning algorithms with real-time monitoring capabilities, the proposed system provides a proactive and intelligent solution for protecting users against phishing attacks and enhancing overall cybersecurity.

VI RESULTS & DISCUSSION

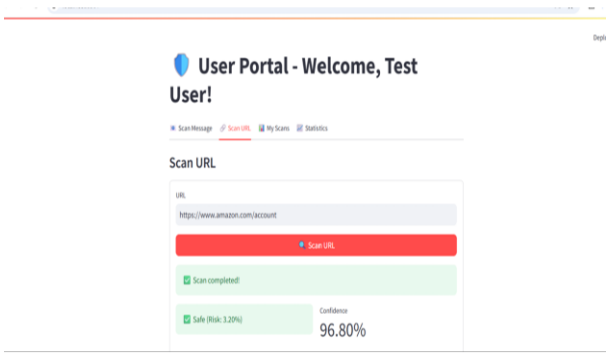
The performance of the proposed AI/ML-based phishing detection system was evaluated using a dataset containing both legitimate and phishing URLs. The dataset was divided into training and testing subsets to measure the accuracy and reliability of the machine learning models. Various algorithms including Decision Tree, Random Forest, Support Vector Machine, and Logistic Regression were tested to determine their effectiveness in detecting phishing websites. Experimental results indicated that ensemble models such as Random Forest achieved higher detection accuracy compared to other algorithms due to their ability to analyze multiple decision paths simultaneously. The system demonstrated high classification accuracy, improved precision, and reduced false positive rates when compared with traditional blacklist-based detection methods. Furthermore, the real-time analysis capability allowed the system to quickly identify suspicious websites and alert users before they interacted with malicious pages. These results confirm that machine learning techniques significantly enhance phishing detection efficiency and provide reliable protection against evolving cyber threats.







International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



VII CONCLUSION

Phishing attacks continue to pose a serious threat to internet users and organizations by exploiting human trust and technological vulnerabilities. Traditional phishing detection techniques such as blacklist databases and rule-based filtering systems are no longer sufficient to combat modern phishing strategies due to their reactive nature and inability to detect newly created malicious websites. This research proposed a Real Time AI/ML Phishing Detection and Prevention System designed to intelligently identify phishing websites using machine learning algorithms and automated feature analysis. The system extracts multiple characteristics from website URLs and webpage content, including domain attributes, security indicators, and structural patterns, which are then used to train predictive models capable of accurately classifying websites as legitimate or phishing. Experimental evaluation demonstrated that machine learning-based approaches significantly improve detection accuracy and reduce false positive rates compared to traditional methods. The integration of real-time monitoring further enhances system effectiveness by providing immediate warnings when suspicious websites are detected. This proactive detection capability helps prevent users from interacting with malicious websites and protects sensitive information from cyber attackers. In addition, the adaptive learning capability of machine learning models allows the system to continuously improve as new phishing patterns emerge. Overall, the proposed system provides a scalable, intelligent, and efficient solution for phishing detection in modern digital environments. Future work may focus on incorporating deep learning techniques, expanding datasets, and integrating the system with browser extensions and network security platforms to further strengthen protection against evolving cyber threats.

REFERENCES

1. Abdelhamid, N., Ayesh, A., &Thabtah, F. (2014). Phishing detection based on association classification. Expert Systems with Applications, 41(13), 5948–5959.
2. Aleroud, A., & Zhou, L. (2017). Phishing environments and countermeasures. Computers & Security, 68, 160–196.
3. Basnet, R., Mukkamala, S., & Sung, A. (2012). Detection of phishing attacks. IEEE Security & Privacy.
4. Bergholz, A., et al. (2010). New filtering approaches for phishing email. Journal of Computer Security, 18(1).
5. Chandrasekaran, M., Narayanan, K., & Upadhyaya, S. (2006). Phishing email detection. IEEE International Conference.
6. Chiew, K., et al. (2018). A survey of phishing detection techniques. IEEE Access, 6, 66043–66065.
7. Dhamija, R., Tygar, J., & Hearst, M. (2006). Why phishing works. Proceedings of CHI.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

8. Fette, I., Sadeh, N., & Tomasic, A. (2007). Learning to detect phishing emails. WWW Conference.
9. Garera, S., et al. (2007). Framework for detection of phishing attacks. ACM Workshop.
10. Gupta, B., et al. (2016). Phishing detection using machine learning. International Journal of Information Security.
11. Herzberg, A., & Margulies, A. (2012). Security of phishing detection. Information Security Journal.
12. Jain, A., & Gupta, B. (2018). Machine learning for phishing detection. Journal of Cyber Security Technology.
13. Khonji, M., Iraqi, Y., & Jones, A. (2013). Phishing detection: A literature survey. IEEE Communications Surveys.
14. Le, A., Markopoulou, A., &Faloutsos, M. (2011). PhishDef: URL names detection. IEEE INFOCOM.
15. Ma, J., et al. (2009). Beyond blacklists: Learning-based phishing detection. ACM SIGKDD.
16. Marchal, S., et al. (2014). PhishStorm detection approach. IEEE Transactions.
17. Mohammad, R., et al. (2015). Intelligent phishing detection system. Expert Systems with Applications.
18. Pan, Y., et al. (2012). Detecting phishing websites using classification. IEEE Conference.
19. Prakash, P., et al. (2010). PhishNet detection system. IEEE INFOCOM.
20. Rao, R., & Pais, A. (2019). Machine learning in phishing detection. Procedia Computer Science.
21. Sahingoz, O., et al. (2019). Machine learning phishing detection. Applied Soft Computing.
22. Sahoo, D., Liu, C., & Hoi, S. (2017). Malicious URL detection using ML. ACM Computing Surveys.
23. Sonowal, G., & Kuppusamy, K. (2017). Phishing detection techniques. Journal of Network Security.
24. Tang, Y., et al. (2016). Phishing detection with ML. Security and Communication Networks.
25. Verma, R., & Das, A. (2017). Machine learning approaches to phishing detection. Security Informatics.
26. Xiang, G., et al. (2011). CANTINA+: phishing detection framework. ACM Transactions.
27. Yang, P., et al. (2019). Deep learning for phishing detection. Future Generation Computer Systems.
28. Zhang, Y., et al. (2017). Detecting malicious URLs. IEEE Security Conference.
29. Zhao, H., et al. (2020). ML-based phishing detection methods. Journal of Information Security.
30. Zouina, M., &Outtaj, B. (2017). Phishing detection system using ML. International Journal of Computer Applications.

31. Mahesh Ganji. (2025). Enhancing Oracle Cloud HR Reporting Through AI-Driven Automation. Journal of Science & Technology, 10(6), 28–36. https://doi.org/10.46243/jst.2025.v10.i06.pp28-36

32. Todupunuri, A. (2025). THE ROLE OF AGENTIC AI AND GENERATIVE AI IN TRANSFORMING MODERN BANKING SERVICES. American Journal of AI Cyber Computing Management, 5(3), 85–93. https://doi.org/10.64751/ajaccm.2025.v5.n3.pp85-93

33. Todupunuri, A. . (2024). Artificial Intelligence Ethics: Investigating Ethical Frameworks, Bias Mitigation, and Transparency in AI Systems to Ensure Responsible Deployment and Use of AI Technologies. International Journal of Innovative Research in Science,Engineering and Technology, 13(09), 1–14. https://doi.org/10.15680/ijirset.2024.1309002

34. Sushma Babburi. (2025). Token-Based Data Accounting System For Transparent Model Training And Cost Allocation. American Journal of AI Cyber Computing Management, 5(4), 463–474. https://doi.org/10.64751/ajaccm.2025.v5.n4.pp463-474

35. Snigdha Gaddam. (2025). SOFTWARE STACK PREPARED FOR AI TRANSITIONING FROM MODULES TO MODELS. American Journal of AI Cyber Computing Management, 5(4), 451–462. https://doi.org/10.64751/ajaccm.2025.v5.n4.pp451-462

36. Gaddam, S. INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING.

37. Bajarang Bhagwat, V. (2023). Optimizing Payroll to General Ledger Reconciliation: Identifying Discrepancies and Enhancing Financial Accuracy. JOURNAL OF ADVANCE AND FUTURE RESEARCH, 1(4). https://doi.org/10.56975/jaafr.v1i4.501636

38. Srinivasa Kalyan Immadi. (2025). Harnessing Artificial Intelligence In Oracle Hcm: Revolutionising Workforce Management With Automation And Predictive Analytics. International Journal of Data Science and IoT Management System, 4(4), 7–13. https://doi.org/10.64751/ijdim.2025.v4.n4.pp7-13

39. S. M. K. P. (2025). Cryptography in iOS: A Study of Secure Data Storage and Communication Techniques. International Journal on Science and Technology, 16(1). https://doi.org/10.71097/ijsat.v16.i1.1403

40. Suhasnadh Reddy Veluru, Sai Teja Erukude, and Viswa Chaitanya Marella. 2025. Multimodal Detection of Fake Reviews using BERT and ResNet-50. In 2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA). IEEE, 877–882.

41. Cyril, H. P. (2025). Event-Driven Provisioning Architectures For Modern

Telecom Networks: Overcoming Legacy Limitations And Enabling Autonomous 6g Operations. International Journal of Advanced Research in Computer Science, 16(6), 75–82. https://doi.org/10.26483/ijarcs.v16i6.7389

42. Jay Bharat Mehta. (2025). AUTONOMOUS PATCH VALIDATION FOR ZERO-DAY EXPLOITS IN ENTERPRISE CLOUDS. International Journal of Applied Mathematics, 38(4s), 1270–1285. https://doi.org/10.12732/ijam.v38i4s.685

43. Reddy, S. K. (2025). Hyperpersonalization driven by AI is expected to be at the Lead in shaping the future of loyalty rewards. Journal of Emerging Technologies and Innovative Research.

44. Reddy, S. K. R. (2021). Strengthening the Security of Loyalty Reward Systems: An In-Depth Analysis of Emerging Cyber Threats and Protection Mechanisms. Journal of Computational Analysis and Applications, 29(6).

45. Poojari, R. (2026). Privacy-Preserving Generative AI in Healthcare Systems Using Federated Learning Approaches. International Journal of Data Science and IoT Management System, 5(1), 78-88.

46. Uday Kumar Kalae. (2025). AN AUTOMATED SYSTEM FOR MANAGING HIGH-AVAILABILITY CLOUD INFRASTRUCTURE THROUGH INFRASTRUCTURE-ASCODE (IAC) PRACTICES. American Journal of AI Cyber Computing Management, 5(2), 42–50. https://doi.org/10.64751/ajaccm.2025.v5.n2.pp42-50

47. Saikumar, B. (2024). Optimizing Crew Scheduling and Absence Management using Microservices: Enhancing Reliability and Efficiency in Crew Management Systems. International Journal of Enhanced Research in Management & Computer Applications, 13(11), 50–55. https://doi.org/10.55948/ijermca.2024.0116

48. Saikumar, B. (2023). Enhancing Client Engagement through AI-Driven Real-Time Reporting and Automated Alerts. International Journal of Enhanced Research in Science, Technology & Engineering, 12(11), 111–117. https://doi.org/10.55948/ijerste.2023.1115