International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

CyberShield EDU: An AI-Powered Platform for Detection and Prevention of Student-Targeted Online Frauds

Mr. Akash Chavan[1], Mr. Om Deshmane[2], Ms. Diksha Deshmane[3], Ms. Jyoti Jadhav[4], Ms. Abha Pathak[5], Ms. Alpana Adsul[6]

[1,2,3,4]B E Computer Engineer(Final Year)
[5]Assistant Professor
[6]Professor

Abstract
The growing digitalization of education has made students heavily reliant on online platforms for various academic processes, including admissions, assessments, and recruitment activities. This increased connectivity has also exposed themto multiple forms of cyber threats, including phishing, impersonation, fake job postings, and deceptive admission portals. This study presents CyberShield EDU, an AI-powered web platform designed to detect, prevent, and raise awareness about cyber fraud targeting students. The system integrates real-time phishing analysis, educational awareness modules, and an efficient reporting mechanism that links students with institutional and government authorities.The proposed system strengthens the ability to detect online deception, promotes cybersecurity awareness, and encourages proactive digital safety habits among learners.

Keywords: Cyber fraud, phishing detection, artificial intelligence, cybersecurity awareness, educational institutions, online fraud prevention, student protection, digital safety.

I. INTRODUCTION

In recent years, advancements in digital infrastructure have significantly influenced the education system, encouraging students to perform most of their academic activities through online platforms. Processes such as application submission, fee payment, and participation in virtual examinations have become an integral part of academic routines. However, this growing digital reliance has also introduced new opportunities for cybercriminals to manipulate unsuspecting users through tactics such as phishing, impersonation, and deceptive scholarship or recruitment schemes. These malicious practices frequently target students, who often have limited awareness of cybersecurity risks.

According to reports from the Indian Cybercrime Coordination Centre (I4C), India experienced financial losses exceeding 22,845.73 crore in 2024 due to cyber-related incidents, emphasizing the increasing scale of digital threats. Furthermore, the absence of a unified reporting mechanism and delayed institutional responses

contribute to the persistence of online fraud, leaving students more exposed to financial and informational exploitation.

To address these issues, this paper introduces CyberShield EDU, an intelligent, AI-driven web platform designed to protect students from online deception. The proposed framework utilizes artificial intelligence to evaluate suspicious links and text content in real time, provides interactive awareness modules, and establishes direct communication channels with institutional and governmental authorities for rapid intervention. The overall objective is to foster a secure, well-informed, and resilient digital learning environment by combining AI-based detection, awareness training, and collaborative reporting.

Prior studies [4]–[6] indicate that Artificial Intelligence (AI) and Natural Language Processing (NLP) techniques are highly effective in analyzing online communication behavior and detecting anomalies linked to fraudulent or deceptive activities. Building on these findings, CyberShield EDU adopts a student-centered design that emphasizes proactive threat detection, continuous cyber awareness, and coordinated reporting to enhance security within academic ecosystems.

II. PROBLEM STATEMENT

India's rapidly evolving digital education ecosystem has led to a significant increase in the use of online platforms for academic activities such as registration, examinations, and recruitment. This expanded dependence on digital systems has simultaneously increased students' exposure to cyber risks, including phishing, impersonation, and fraudulent academic websites [3], [4].

Commonly observed scams targeting learners include:

- Phishing and social engineering-based deception
- Impersonation of faculty members, institutions, or officials
- Fake portals for admissions and examinations
- Misleading job or scholarship advertisements

Research indicates that, although AI and NLP-driven models have improved the detection of fraudulent patterns [5], [6], most existing tools are designed for corporate or recruitment environments rather than educational contexts. In addition, limited digital security awareness among students and the absence of an integrated, user-friendly reporting mechanism make it difficult to respond promptly to emerging threats [1].

To address these limitations, an AI-enabled, student-oriented framework is required—one that can identify potential scams in real time, issue timely alerts, and cultivate a culture of cybersecurity awareness within academic institutions. The proposed CyberShield EDU platform aims to fulfill these objectives by merging intelligent detection, preventive education, and streamlined reporting in a single accessible system.

III. RELATED WORK

Previous research in the domain of cyber fraud Earlier studies on cyber fraud mitigation have mainly focused on enhancing institutional and organizational defense mechanisms. While these techniques effectively protect enterprise systems, they do not adequately account for the personal-level risks and digital behavior patterns that make students particularly vulnerable to online scams. securing enterprise systems, they often fail to address the specific vulnerabilities and behavioral risks associated with individual users, particularly students. Tan et al. [1] presented a generative AI-based framework, ScamGPT-J, that models scammer behavior to

enhance phishing and messaging scam detection. Their work demonstrated the potential of large language models in understanding fraudulent intent but primarily focused on general online communication environments rather than academic settings.

Al-Yozbaky and Alanezi [4] utilized Natural Language Processing (NLP) techniques for phishing email classification, achieving high detection accuracy through linguistic pattern recognition. Similarly, Vidros et al. [3] introduced the EMSCAD dataset to analyze the characteristics of online recruitment frauds and employed Random Forest algorithms for identifying deceptive job postings. While effective, both studies primarily addressed recruitment-related fraud and not student-specific cyber threats.

Further studies, such as those by Taneja et al. [6] and Papasavva et al. [5], applied advanced transformer and AIbased models for online fraud detection and content classification. These works highlight the evolving use of AI in mitigating fraud, but they emphasize generalized frameworks rather than domain-specific security applications. Akram et al. [7] also demonstrated the role of deep learning approaches in improving detection precision for online recruitment fraud, though scalability and accessibility for educational users remain limited.

Mahbub et al. [9] explored the impact of contextual features, showing that integrating regional business and institutional data can significantly enhance fraud detection accuracy. Kumar and B. B. M. [8] implemented a Random Forest model for identifying fake job advertisements, underscoring the practicality of lightweight machine learning solutions.

In contrast to these prior studies, government-led initiatives and awareness programs have focused mainly on public education or institutional-level security but often lack AI integration or real-time digital accessibility. The cybercrime.gov.in portal, although valuable for reporting incidents, remains underutilized among students due to its complex interface and multistep reporting structure.

The proposed CyberShield EDU system builds upon these foundational works and bridges the existing research gap by providing an AI-driven, student-centric web platform that integrates fraud detection, awareness generation, and simplified reporting workflows. This unified approach ensures accessibility, real-time analysis, and active collaboration between students, institutions, and cyber authorities, addressing the unique challenges of cybersecurity in academic environments.

IV. LITERATURE REVIEW

In Paper [1], Tan, See, and Kok proposed ScamGPT-J, a generative-AI framework developed to mitigate instantmessaging scams through large language models (LLMs) [1]. The research highlights that conventional binary classifiers often overlook the conversational flow and psychological manipulation strategies typical of social-engineering scams. To overcome this limitation, the authors fine-tuned the opensource GPT-J model using a curated dataset of 902 simulated scam conversations representing themes such as authority, employment, romantic, and investment frauds. The fine-tuned model generates potential scam-like replies in real time, enabling users to recognize deceptive message patterns by comparing system responses with authentic conversations.

Rather than automatically filtering or blocking suspicious messages, ScamGPT-J emphasizes user awareness and self-assessment. The framework leverages principles from heuristic-systematic information processing theory, encouraging users to shift from intuitive to analytical reasoning when confronted with questionable

messages. Experimental evaluation showed that the tuned model improved cosinesimilarity alignment with real scam texts by approximately 25% over baseline GPT-J performance. In a small-scale user study involving 20 participants, the system achieved a usefulness rating of 4.4 out of 5, reflecting its educational and preventive potential. This approach aligns closely with the goals of CyberShield EDU, demonstrating how generative AI can reinforce cyber-awareness and active scam-recognition capabilities among students.

In Paper [2], Al-Yozbaky and Alanezi proposed an intelligent framework utilizing Support Vector Machines (SVM) and Random Forest (RF) classifiers for detecting online recruitment fraud using the EMSCAD dataset [2]. Their method focused on extracting semantic and structural features such as company profile completeness, job title authenticity, and contact consistency. The ensemble model achieved a remarkable accuracy of 97.41%, proving that combining linear and tree-based classifiers yields robust discrimination between legitimate and fraudulent postings. The study also highlighted the importance of feature selection, as irrelevant or redundant inputs negatively affect fraud detection performance.

In Paper [3], Vidros et al. defined and analyzed the characteristics of Online Recruitment Fraud (ORF) and introduced the first publicly available EMSCAD dataset comprising 17,880 labeled job advertisements [3]. The authors experimented with bag-of-words models and Random Forest classifiers, achieving 91.22% accuracy. Their work laid the groundwork for subsequent fraud detection studies by emphasizing the need for real-world, labeled datasets to train and benchmark AI models. However, they also noted that lexical features alone are insufficient for detecting complex frauds, encouraging further research into contextual and behavioral cues.

In Paper [4], Al-Yozbaky and Alanezi developed a Natural Language Processing (NLP)-based phishing detection approach for Arabic emails [4]. The system employed a hybrid mechanism combining blacklisted phishing terms, rootbased tokenization, and sentence-level similarity analysis. The authors reported 99% accuracy in detecting legitimate emails and 96% in identifying phishing content. The study's main contribution lies in addressing language-specific phishing challenges, demonstrating that linguistic diversity demands localized solutions for effective detection.

In Paper [5], Papasavva et al. performed a comprehensive Systematic Literature Review (SLR) of 223 research papers published between 2019 and 2024, exploring AI and NLP methods used for online fraud detection [5]. The study revealed that fraud research remains fragmented across 16 domains, including phishing, recruitment scams, and fake reviews. They concluded that most models lack explainability and long-term adaptability, with many studies omitting model bias and generalization analysis. Their review underscores the necessity of transparent AI frameworks for sustainable cybersecurity solutions.

In Paper [6], Taneja, Vashishtha, and Ratnoo introduced Fraud-BERT, a transformer-based, context-aware model for online recruitment fraud detection [6]. The framework addressed limitations of earlier context-free methods such as TF-IDF and word2vec. Fine-tuned on the EMSCAD dataset, Fraud-BERT achieved 99% accuracy and a 0.93 F1-score without oversampling. The authors concluded that contextual embeddings derived from BERT substantially improve detection precision and recall by capturing semantic nuances ignored by traditional vectorization techniques.

In Paper [7], Akram et al. investigated the impact of class imbalance in online recruitment fraud detection using transformer architectures such as BERT and RoBERTa [7]. The authors experimented with multiple oversampling techniques, including ten SMOTE variants, to improve model performance on skewed datasets.

Their findings revealed that the SMOBD-SMOTE method yielded balanced accuracy and recall of approximately 90%. The study demonstrated that data balancing significantly enhances transformer model robustness when applied to fraud detection tasks with limited positive samples.

In Paper [8], Kumar and B. B. M. applied the Random Forest algorithm to detect fake job postings [8]. Their framework extracted lexical and metadata features such as job title repetition, email domain anomalies, and contact inconsistency. The model achieved high predictive accuracy and demonstrated potential for deployment in lightweight web applications. The researchers recommended future integration with deep learning to improve adaptability to evolving scam techniques.

In Paper [9], Mahbub, Pardede, and Kayes emphasized the significance of contextual and regional attributes in improving fraud detection outcomes [9]. They collected Australian job listings and derived features from business registry validations (SEEK, ABR, and Indeed) to assess advertiser legitimacy. Machine-learning classifiers, including J48, Na¨ıve Bayes, and Random Forest, all showed improved accuracy and recall when contextual features were incorporated. The results confirmed that localized and multi-feature approaches yield better generalization in fraud detection systems.

Collectively, these eight studies demonstrate a clear progression from rule-based detection to data-driven, transformerbased learning approaches. While earlier works focused on structural or textual attributes, recent research highlights the need for hybrid frameworks that integrate contextual, behavioral, and semantic cues. Despite high reported accuracy levels, existing models often lack adaptability for real-time or domain-specific deployment, particularly in the education sector. Hence, the development of CyberShield EDU aligns with this research direction—leveraging AI for detection, while incorporating user awareness and reporting mechanisms tailored for students.

V. SYSTEM DESIGN AND METHODOLOGY

A. Objectives

1. To detect student-targeted phishing and fraud patterns using AI.
2. To provide an interactive awareness and prevention interface.
3. To establish an efficient fraud reporting workflow linking students with authorities.

B. Architecture Overview

The CyberShield EDU platform follows a modular, layered design that integrates user interaction, AI-driven detection, and secure data management. The system architecture is composed of the following components:

- Frontend: An online user interface designed for reporting incidents, accessing awareness content, and viewing security analytics. It allows users to report suspicious activities, access awareness resources, and view visualized insights
- Backend: A server-side engine responsible for AI-driven fraud detection, data storage, and report management.
- AI Components: Includes URL and text classification models trained on phishing datasets and a natural language model for identifying impersonation or scamrelated content.
- Database: Maintains anonymized records of incidents, alerts, and detection outcomes to ensure privacy and facilitate continuous learning.

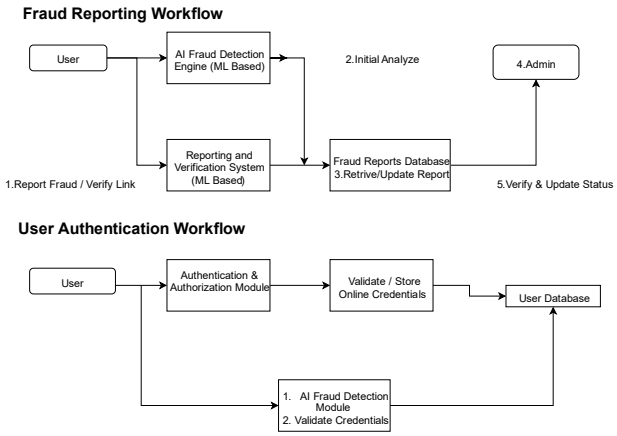

Fig. 1: System Architecture of CyberShield EDU

C. Workflow

The operational workflow of CyberShield EDU ensures seamless detection and awareness generation for users through the following process:

1. A user visits the CyberShield EDU web portal and submits a suspicious message, email, or URL.
2. The AI engine analyzes the content using phishing detection and text classification models.
3. If malicious patterns are detected, the system issues an alert and recommends actions such as reporting, payment suspension, or notifying authorities.
4. Educational content and preventive awareness tips are presented to the user to reinforce safe online behavior.

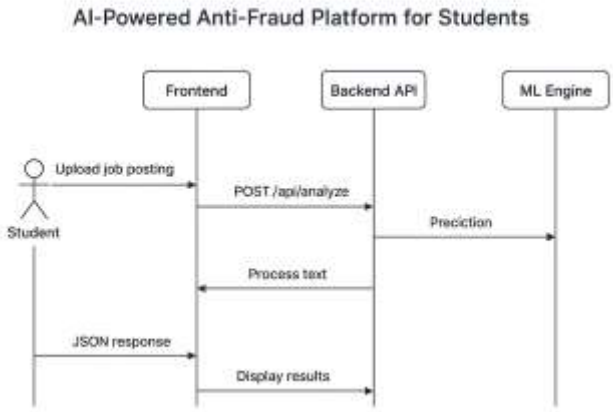

Fig. 2: Workflow of CyberShield EDU Platform

Empirical studies indicate that contextual alerting can significantly reduce response times and improve remediation outcomes [5], [6]. For instance, dashboards powered by contextual AI not only list detected threats but also visualize their relationships with network topology, dependencies, and historical attack data [1]. Such visualization supports continuous risk assessment and assists organizations in building resilient defensive mechanisms, which are essential for next-generation networked systems.

VI. CONCLUSION

Cyber threats are increasingly directed toward students in digital learning environments, taking advantage of their growing dependence on online systems for academic purposes such as online assessments, institutional communication, and placement activities. Common attack types, including phishing, impersonation, and fake educational websites, have led to significant disruption and challenges within the academic sector.

To address these challenges, the CyberShield EDU framework introduces an integrated, AI-assisted approach that combines intelligent detection of suspicious links and messages with interactive awareness modules developed specifically for students. The system emphasizes proactive prevention through awareness, real-time risk identification, and a simplified reporting process that connects users with institutional and governmental authorities.

By merging technological detection with continuous cyber awareness, CyberShield EDU encourages students to follow safer online practices and improves their capability to identify and handle cyber threats more effectively. Deploying such a system across educational institutions can substantially reduce student-targeted online fraud and strengthen a culture of secure and responsible digital engagement within academia.

Future enhancements may include integrating multilingual awareness content and automated reporting to national cybercrime databases for broader scalability.

REFERENCES

1. X. W. Tan, K. See, and S. Kok, “ScamGPT-J: Inside the Scammer’s Mind—A Generative AI-Based Approach Toward Combating Messaging Scams,” in Proc. 45th Int. Conf. on Information Systems (ICIS), Bangkok, Thailand, 2024.
2. R. Sh. Al-Yozbaky and M. Alanezi, “Detection and Analyzing Phishing Emails Using NLP Techniques,” in Proc. 5th Int. Congress on HumanComputer Interaction, Optimization and Robotic Applications (HORA), 2023.
3. S. Vidros, C. Kolias, G. Kambourakis, and L. Akoglu, “Automatic Detection of Online Recruitment Frauds: Characteristics, Methods, and a Public Dataset,” Future Internet, vol. 9, no. 1, p. 6, 2017.
4. R. Sh. Al-Yozbaky and M. Alanezi, “Detection and Analyzing Phishing Emails Using NLP Techniques,” in Proc. 5th Int. Congress on HumanComputer Interaction, Optimization and Robotic Applications (HORA), 2023.
5. Papasavva et al., “Applications of AI-Based Models for Online Fraud Detection and Analysis,” Crime Science, vol. 14, no. 7, 2025.
6. K. Taneja, J. Vashishtha, and S. Ratnoo, “Fraud-BERT: TransformerBased Context-Aware Online Recruitment Fraud Detection,” Discover Computing, vol. 28, no. 9, 2025.
7. N. Akram et al., “Online Recruitment Fraud (ORF) Detection Using Deep Learning Approaches,” IEEE Access, 2024.
8. V. Kumar H. L. and B. B. M., “Machine Learning for Fake Job Detection,” Int. J. of Advanced Research in Computer and Communication Engineering, vol. 13, no. 8, pp. 170–174, 2024.
9. S. Mahbub, E. Pardede, and A. S. M. Kayes, “Online Recruitment Fraud Detection: A Study on Contextual Features in Australian Job Industries,” IEEE Access, vol. 10, pp. 82776–82787, 2022.