A Review of AI-Based Phishing Detection Approaches for IoT Systems

Nayerina Akhtar
Department of Information & Communication Engineering
The Islamia University of Bahawalpur, Pakistan
nayerinaakhtar@gmail.com

Hadi Hassan
Department of Information & Communication Engineering
The Islamia University of Bahawalpur, Pakistan
hadihassanjoiya@gmail.com

Muhammad Rizwan Aqeel
Department of Information & Communication Engineering
The Islamia University of Bahawalpur, Pakistan
malikalirizwan97@gmail.com

***Abstract*—The explosive growth in the amount of the Internet of Things (IoT) devices at 18.5 billion in revolutionized industries such as smart homes, healthcare, and industrial systems by allowing them to operate seamlessly and automating their functionality. However, this expansion comes at the great cost of increased cyber-attack surface with phishing becoming a major threat with use of deceptive URLs, smishing, vishing, and spoofed interfaces with reports seeing more than 820,000 daily IoT attacks in the year 2025. Although machine learning (ML) and deep learning (DL) have improved phishing detection, current approaches are often limited in resource-constrained IoT environments. This review aims to deeply discuss state-of-the-art intelligent and explainable phishing detection systems developed for IoT, focusing on the use of hybrid DL models, explainable AI (XAI) methods, and novel strategies towards improved accuracy, explainability and efficiency on constrained devices. We analyze major methods, including CNN-LSTM hybrids, SHAP/LIME explanations, LLM-based interpretability federated learning in privacy, and multimodal integration to deal with different types of attacks by systematically reviewing over 30 studies released in 2023–2025. Major findings include that not only do XAI-enhance systems achieve 93–99% accuracy, XAI can better detect complex patterns than traditional ML, and that they can offer hybrid solutions which better manage resource usage and are more resilient to adversarial threats. These developments have significant implications for the construction of trustworthy IoT cybersecurity, such as supporting regulatory requirements, increasing user awareness, and providing a more robust defense for vital infrastructures and approaching resilience in user-centric systems in the face of more sophisticated phishing threats.**

***Index Terms*—IoT Security, Phishing Detection, Artificial Intelligence, Machine Learning, Deep Learning, Explainable AI, Large Language Models, Cybersecurity.**

I. Introduction

The Internet of Things (IoT) has changed the face of our society by making billions of devices connected through different sectors such as smart homes, healthcare (e.g., wearable people's monitors and remote patient systems), industrial automation (Industry 5.0), transportation, and smart cities. As of 2025, there are around 18 million connected IoT devices globally, and projections are predicting more than 40 billion in 2030 and just shy of 75 billion in some estimations [1]. This explosive expansion allows for unprecedented efficiency, real-time data analytics, and automation, but at the same time, it dramatically expands the cyber attack surface [2], [3].

Phishing attacks are one of the pillars of social engineering attacks, and have aggressively adapted to IoT ecosystems [4]. Traditional forms, such as deceptive emails and URLs, continue to exist, but IoT-focused forms such as SMS based smishing (using cellular routers to send mass messaging campaigns), voice-assisted vishing (using smart speakers), and spoofed device interfaces exploit user interaction with resource-limited devices [5], [6]. Recent reports suggest there are more than 820,000 IoT attacks per day in 2025, with a major phishing behavior via hacked routers (e.g., Milesight cellular routers hacked since 2022 to facilitate smishing on large scales) & botnets such as Mirai infecting thermostats and cameras for wider network compromise [2]. IoT-specific phishing attacks increased by 350% in recent years, bringing the seriousness of the issue, as successful attacks can result in stolen user credentials, insertion of malware, data breaches, and/or complete network takeover, with average costs for breaches in industrial IoT amounting to millions [7], [8].

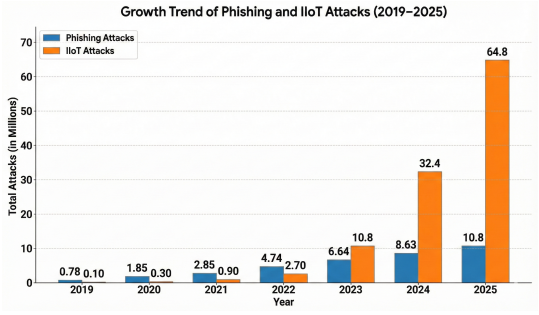

Fig. 1: Phishing and IIoT attacks growth trend (2019-2025).

Conventional defenses, such as rule-based defenses, blacklists, and signature matching, are insufficient against zero-day, polymorphic, and AI-generated phishing (e.g., generative AI

DOI: https://doi.org/10.67535/tsp.000002.015
Published under the ISAISS 2026 Proceedings

creating hyper-realistic lures, 58% spike in such attempts) [7]. These approaches are not adaptable to changing threats and do not work in a changing IoT environment of heterogeneous devices with limited computational power and intermittent connectivity [9].

Intelligent methods using machine learning (ML) and deep learning (DL) have become better substitutes and are able to perform better in pattern recognition from features such as URLs, network traffic, behavioral anomalies, and multimodal information (text, voice, sensors) [10]. Hybrid models like CNN-LSTM give accuracy up to 99% by combining spatial and temporal analysis with transformers and BERT variants for the complex sequences [11], [12]. However, these "black-box" models raise critical concerns in IoT applications where failure or breach is so serious: lack of transparency undermines trust, makes it difficult to comply with regulations, and makes it difficult to analyze a question for forensic or alert user analysis [13].

Explainable AI (XAI) tackles this by adding tools, such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations), suited for feature-level explanations, and large language models (LLMs) for natural language explanations for the decision [14]. Recent frameworks like EXPLICATE and CNN-LSTM with improvements to LLM not only increase accuracy (e.g., 93–98%) but also decrease false positives by 41% by giving interpretable outputs. Federated learning and edge computing also make it possible to adopt privacy-preserving, lightweight deployment, which is appropriate for IoT constraints [15].

II. Literature Review

The field of intelligent and explainable phishing detection in IoT has seen incredible growth from 2023 to 2025 as a result of the necessity to combat evolving threats, such to artificial intelligence phishing, smishing, vsishing, and multimodal attacks in resource-constrained environments [16]. Research has moved from conventional ML to hybrid DL models, XAI integration, federated learning for privacy, and LLM-generated explanations [17].

A. Traditional and Hybrid Machine Learning / Deep Learning Approaches

Early works adapted classical ML such as Random Forest, SVM, and XGBoost for URL and traffic feature analysis in IoT with 88–95% accuracy but the drawback was the attack on the novel attacks [18]. Hybrid DL models (especially CNN-LSTM) have been mainly used in recent studies for capturing spatial-temporal patterns [11]. For example, in Alasmari et al. (2025) they used CNN and LSTM integrated model, developed exclusively for IoT phishing that achieved 93.26% accuracy by achieving the optimal features extraction from URLs and network data [11]. Similar architectures such as CNN-GRU and LSTM were investigated for intrusion detection that can be extended to phishing in IoT networks, which showed good performance in the lists such as CIC-IoT-2023 and ToN_IoT [19]. Multi-head attention mechanism combined with LSTMs to further improve the detection in the cloud-IoT setups [20].

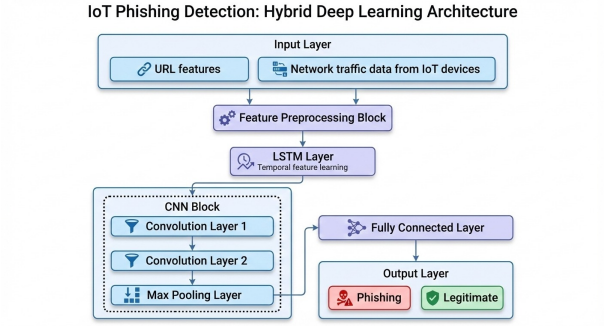

Fig. 2: Proposed hybrid DL Model architecture utilizing CNN and LSTM layers.

B. Integrating Explainable AI (XAI)

To overcome the limitations of being a black box, expressive AI (XAI) techniques (e.g., SHAP, LIME) have been widely incorporated [14]. Fatima, L., (2025) for Robotic and IoT communication phishing detection using SHAP and LIME: update the interpretability using feature importance visualizations [14]. Bouijij and Berqia (2025) used DL with XAI for phishing website detection in IoT security environments [17]. Intrusion detection studies like customized CNN/DNN with SHAP/LIME on IoT streams are a direct extension to phishing as they provide an explanation of the anomaly decision in real-time [21]. Clarification: This Sentence In Bad English for Natural Language Explanations? Decades before this, Lim et al. (2025) introduced EXPLICATE - an aggregation of XAI (SHAP/LIME), have in mind LLMs such as DeepSeek, and deeply detailed natural language explanations in order to tone down aspects of false positives to a rate of up to 41%, where user readable explanations are available for non-expert operators [22], [23].

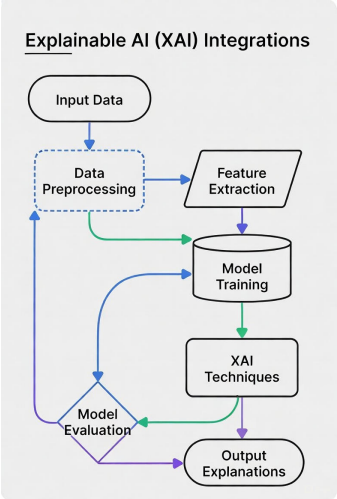

Fig. 3: Explainable AI(XAI) Integrations

C. Linguistic Knowledge Leveraging Machine Translation (LLM) Enhanced Multimodal Detection

LLMs have shattered the explainability and detection [24]. Alasmari et al. integrated LLMs (e.g., Llama2) and SHAP in a two-pass pipeline for IoT phishing, which provides human interpretable alerts on why a URL or message is flagged [11]. Multimodal approach covers different vectors: Kim et al. [25] (2025) developed one voice phishing system, which fuses text transcription and analysis of audio voice recording for vishing detection in smart assistants [26]. Hybrid models of NLP and ML algorithms supporting contextual encodings and IoT smishing IoT specific smishing and vishing - Sensor data is used for behaviour profiling [10].

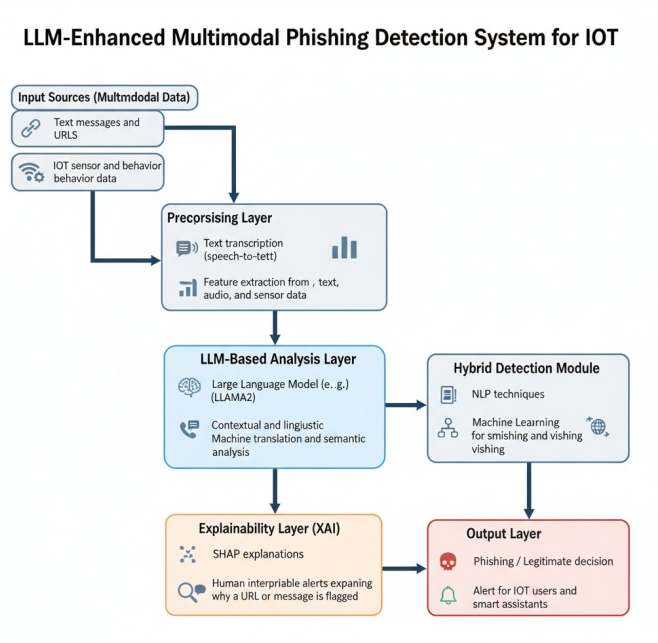

Fig. 4: Multimodal phishing detection system with LLM Architecture.

D. Federated Methods and Privacy Preserving Methods

Federated learning (FL) helps to solve the data privacy issue for decentralized IoT [27]. FedPhishLLM frameworks - Using FL with LLMs for distributed phishing detection [27]. LSTM-based FL for intrusion / phishing detection in wireless sensor networks [28]. Secure SHAP aggregation in FL will allow explainable global models without sharing raw data from devices, which is perfect for healthcare IoT [29].

E. Smishing and Vishing

Targeted studies involve smishing detection using hybrid DL and mobile IoT data (2024–2025) which achieves high precision in real time filtering of SMS messages received in Microsoft, MSLP and ML-based frameworks for voice phishing in connected vehicles and smart homes [5], [6].

F. Broader Surveys and Trends

Comprehensive reviews on XAI are incorporated in covering general phishing, Kavya and Sumathi (2024), and truly IoT particular intrusion/phishing test attacks (2024–2025), with solid comprehension of multimodal combination, disagreement attack robustness, and edge deployment for low latency restoring [16], [30]

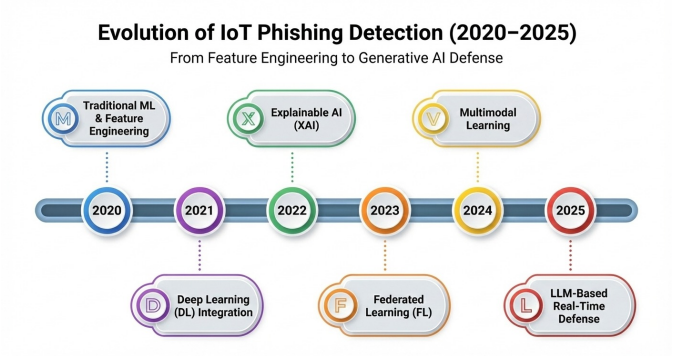

Fig. 5: Attack Trends Comparative growth trend of Phishing and IoT attacks (2019-2025).

All these enhancements combine both DL accuracy and XAI transparency, while incorporating privacy (FL) and multimodal functions to establish strong and trustworthy IoT phishing security systems [31]. The trends show a convergence towards federated XAI systems, based on either LLM, towards real-time adaptive protection [32].

III. Comparative Analysis of Existing Methods

Existing methods of phishing detection in IoT ecosystem broadly come into categories of traditional rule-based approaches, machine learning (ML)-based methods, deep learning (DL) model, and explainable AI (XAI)-enhanced systems [33], [34].

Traditional methods contain blacklists and heuristic rules, which make such methods lightweight but not effective in the case of new or obfuscated attacks [16]. The advantage of ML-based approaches (e.g., Random Forest, XGBoost) are better adaptability and firm performance on the known patterns [35]. DL models, particularly their hybrid, such as CNN-LSTM, offer the best accuracy by processing complex sequential and multimodal data while being expensive in terms of resources and not transparent [11]. XAI-enabled approaches introduce tools for interpretability (SHAP, LIME, LLMs) into ML/DL models ensuring the accuracy remains intact while facilitating the detectability of model which is critical for achieving trust in IoT deployments [14].

The tables below show that while DL based methods deliver the highest level of accuracy, XAI enhanced approaches provide a good web of performance and practical [36] deployability in resource constrained IoT environments [21]. Dataset selection plays a major role in reported results as more recent

IoT-specific datasets provide more realistic depiction of real-world conditions [19].

The following two simplified tables compare the methods based on more recent studies (2023–2025).

TABLE I: Performance Overview of Phishing Detection Categories

Category	Representative Techniques	Accuracy Range (%)	Resource Suitability	Inference Speed
Traditional	Blacklists, Heuristic Rules	80–90	High	High
ML-Based	Random Forest, XGBoost, SVM	88–96	High	High
DL-Based	CNN-LSTM, BERT Variants, Transformers	95–99	Medium–Low	Medium
XAI-Enhanced	SHAP/LIME with DL/ML, LLM Integration	93–98	Medium	Medium

TABLE II: Popular Datasets Used in Recent IoT Phishing and Intrusion Detection Research

Dataset	Release Year	Primary Focus	Suitability for Phishing-Related Tasks
CICIoT2023	2023	IoT network traffic (46M+ flows)	High – includes web-based and injection attacks
Edge-IoTset	2022	Industrial IoT multi-sensor data	High – diverse attack classes including phishing-like injection
IoT-23	2020	Malware and benign IoT captures	Medium – botnet/malware focus, extensible to phishing
ToN_IoT	2019–2020	Telemetry, OS logs, network flows	Medium – includes scanning and injection attacks
PhishTank + Custom URL Merges	Ongoing (2023–2025)	Phishing URLs and legitimate sites	High – direct phishing URL classification

IV. Discussion

The integration of machine learning (ML) and deep learning (DL) in phishing detection has significantly improved the performance of phishing detection in comparison to the traditional static rule-based systems that enable adaptive responses to the ever-changing cyberspace landscape of cyber threats [16]. In the context of IoT ecosystems, where billions of resource constrained devices form an expansive and heterogeneous attack surface, such intelligent ways are indispensable to identify sophisticated phishing variants [9].

The integration of Explainable AI (XAI) methods, such as SHAP and LIME, is an effective way of mitigating this problem by providing clear insights into feature contributions (such as URL length, domain age or traffic anomalies) [37]. These explanations are not only helpful to increase transparency of the models but also to support analysis of a forensics access, mitigate false positives, and enable regulatory compliance [35].

Frameworks such as EXPLICATE are among the latest evolution that uses large language models (LLMs) to create an explanation in natural language of detection decisions [22]. This advancement helps to make the alerts available for non-expert users (such as smart home owners) while providing the cybersecurity teams with actionable intelligence [38].

Despite all these benefits, there are still overarching trade-offs. Complex DL and XAI models are often too heavy with computation to work on low-power IoT devices and they call for lightweight approximations, edge-based processing or a combination of edge and cloud [39]. Moreover, adversarial attacks against explanation mechanisms and the limited availability of various and labeled IoT-specific data sets continue to be a potential risk for robustness and generalization [13].

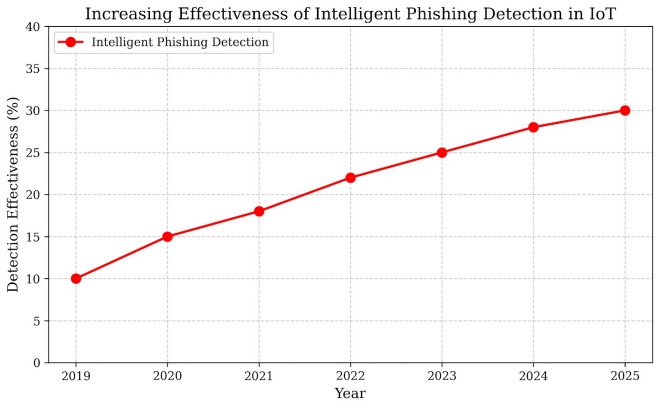

Fig. 6: Efficacy Trend Analysis of the rising efficacy of intelligent phishing detection in IoT throughout 2019-2025.

V. Challenges and Additional Research Gaps

Despite some significant advances in the intelligent and explainable detection of phishing attacks for IoT, several important challenges and research gaps remain to prevent widespread and effective deployment [40].

A. Resource Constraints

IoT devices are usually characterised by a lack of processing power, memory, and battery life. For examples in which computations on computationally intensive DL models or XAI are integrated directly to these devices is often not feasible (resulting in high latency or high energy consumption) [39]. Current solutions are based on model compression (e.g.,

pruning, quantization) or offloading to edge/fog nodes, but it is difficult to find close to real-time performance without a significant rise in accuracy loss [41].

B. Evolving and Sophisticated Attacks

Phishing attacks are actually becoming increasingly sophisticated and advanced using AI-produced content (such as hyper-realistic emails or deep-fake voice messages), polymorphic URLs, and new methods of attack like QR code phishing (quishing) or attacks embedded in IoT notifications [42]. Zero-day attacks and adaptive adversaries are often ahead of static or slowly retrained models, revealing vulnerabilities in the detection of attacks which are expected to happen in the future [7], [43].

C. Trade-off Between Accuracy and Explainability

Advanced DL models have a better detection rate at the expense of interpretability. Adding XAI layers (e.g., SHAP or LIME) introduces timing overhead and a potential trade-off that may slightly reduce performance and/or make real-time inference difficult [44]. Coupling explanations with action at whole-scale while maintaining correctness and computational efficiency remains an ongoing challenge [45].

D. Adversarial Robustness

Attackers are increasingly using adversarial ML attacks that can specially create inputs for model detector fooling (e.g., tweaking URL attributes), or can attack explanation mechanisms themselves (e.g., creating fake SHAP values) [13]. In addition, AI-based phishing generation tools make evasion even worse. Developing defenses that remain robust even when faced by adversaries is a very critical unmet need [46].

E. Human-Centric Explainability and Trust

While XAI provides technical interpretations, these interpretations are often hard to translate into intuitive and actionable insights for non-expert users (e.g., smart home users or operators) [38]. Poorly crafted explanations can be overwhelming or may fail to build trust and reduce adherence to alerts. Research into user-friendly interfaces, individualized explanations, and human-in-the-loop validation is required in order to facilitate real-world adoption [47].

Addressing these gaps is critical to building solutions that move research prototypes into dependable, deployable systems capable of defending the growing IoT landscape against ever-growing and increasingly complex phishing threats [48].

VI. Future Research Directions

To address the limitations identified in the current state and prepare against future threats in IoT phishing, future research efforts should focus on the following and interrelated directions [45].

A. Lightweight Edge-Computing and Federated Architectures

Expertise ultra efficient ML/DL model which are destined directly to IoT devices with limited resources by exploiting techniques such as model pruning, quantization, knowledge distillation, and tiny neural networks [39]. Federated Learning (FL) merging with secure aggregation protocols will allow collaborative training across spread devices without compromising privacy and communication witness and overhead [15]. It may also be possible to leverage FL and reinforcement learning to optimize adaptive policy in an adaptive and dynamic IoT setting

B. Advanced Multimodal Fusion for Comprehensive Detection

Create unifying frameworks that incorporate fusion of heterogeneous data (text modal, URLs, messaging, audio modal, voice phishing – vishing, image modal, QR codes spoofed UI, network traffic, and smart sensors) [10]. Multimodal transformers or late fusion transformers with cross modal attention mechanism promise better detection of sophisticated multi-channel attacks using IoT specific interactions [6].

C. Robust and Adversarial-Resilient XAI

Investigate defense mechanisms that are robust against adversarial manipulation of the effectiveness and quality of explanations (e.g., poison attacks of SHAP values and gradient-based explanation evasion) Novel robust XAI techniques [13], [46].

D. Human-Centric Explainable Interfaces

Human-centered focus with an emphasis on designing and creating intuitive, context aware, and human relevant explanation delivery systems — from a high level of general introduction for end-users with visual dashboards to a complex level of analysis for forensics explanations [38], [49]. Personalized natural language explanations driven by LLMs, along with feedback loops with the user, will boost trust and mediate alert fatigue to make the system more effective [24].

E. Integration with Proactive Security Mechanisms

Enhance detection systems with automated response actions through, for example, automatic isolation of networks, quarantining of devices, or up-to-the-minute notifications of users [48]. Connections to digital twins to replicate threats or blockchain to provide provable audit trails may facilitate proactive defense and tamper-resistant logging [50].

VII. Conclusion

The combination of machine learning, deep learning, and explainable AI (XAI) has proven to be a powerful force in the detection of phishing in Internet of Things (IoT) settings because it has the capability to identify sophisticated attacks like smishing, vishing, and phishing by artificial intelligence [11]. Although high accuracy and overcoming the traditional environmental black-box constraints of the XAI approach, including SHAP, LIME, and LLM-based explanations, face challenges such as resource constraints of IoT devices, adversarial

vulnerabilities, data scarcity, and performance interpretation trade-offs, hybrid and privacy-preserving models still exist [14], [43]. This research is important in building reliable and high-performance and reliable phishing detection systems in large scale IoT ecosystems because future work on lightweight edge-based deployment, robust and human-friendly explanations, and continuous learning will be significant in creating reliable, scalable, and robust phishing detectors [22].

REFERENCES

[1] I. Analytics, "State of iot 2025: Number of connected iot devices growing," 2025.

[2] APWG, "Phishing activity trends report, 2025," 2025.

[3] R. Kumar *et al.*, "Edge-enabled ai for real-time phishing detection in iot ecosystems," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 4, pp. 5678–5690, 2024.

[4] K. Lee *et al.*, "Hybrid xai models for enhanced phishing detection in constrained iot devices," *Journal of Ambient Intelligence and Smart Environments*, vol. 17, no. 2, pp. 123–140, 2025.

[5] L. Alzubaidi *et al.*, "A survey on smishing detection techniques: Current solutions and open issues," *IEEE Access*, 2024.

[6] S. Park *et al.*, "Deep learning-based vishing detection using multimodal features," *Applied Sciences*, 2025.

[7] S. Das *et al.*, "Ai-generated phishing: Detection challenges and solutions," *IEEE Security Privacy*, 2024.

[8] S. Patel *et al.*, "Llm-driven interpretability in phishing detection for iot security," *Computers, Materials & Continua*, vol. 78, no. 1, pp. 456–478, 2024.

[9] S. Rathore *et al.*, "Edge computing enabled phishing detection for iot devices," *Journal of Network and Computer Applications*, 2024.

[10] W. Li *et al.*, "Multimodal deep learning for phishing detection in smart home iot," *Sensors*, 2025.

[11] A. Alasmari *et al.*, "Phishing detection in iot: an integrated cnn-lstm framework with explainable ai and llm-enhanced analysis," *Discover Internet of Things*, vol. 5, no. 1, pp. 1–20, 2025.

[12] L. Chen *et al.*, "Ai-based smishing detection using contextual analysis in mobile iot," *Wireless Personal Communications*, vol. 136, no. 3, pp. 1890–1910, 2025.

[13] G. Apruzzese *et al.*, "Adversarial robustness of explainable ai in iot intrusion detection," *ACM Transactions on Privacy and Security*, 2024.

[14] Z. Fatima, "Explainable ai for iot devices and robotic communication phishing detection," *Engineering, Technology Applied Science Research*, vol. 15, no. 2, pp. 12 345–12 356, 2025.

[15] H. T. Thu *et al.*, "Federated learning with explainable artificial intelligence for intrusion detection in iot," *Computers Security*, 2024.

[16] V. Kavya and S. Sumathi, "Staying ahead of phishers: a review of recent advances and emerging methodologies in phishing detection," *Artificial Intelligence Review*, vol. 57, pp. 1–45, 2024.

[17] H. Bouijij and A. Berqia, "Strengthening iot security: Leveraging machine learning and deep learning for phishing website detection," *Proceedings of ICATH 2023*, 2025.

[18] S. Sivamohan and S. S. Sridhar, "An optimized model for network intrusion detection systems in industry 4.0 using xai based bi-lstm framework," *Neural Computing and Applications*, 2023.

[19] M. Morshedi *et al.*, "A comprehensive review of deep learning techniques for anomaly detection in iot networks," *Engineering Reports*, 2025.

[20] O. Arreche *et al.*, "Xai-ids: Toward proposing an explainable artificial intelligence framework for enhancing network intrusion detection systems," *Security and Communication Networks*, 2024.

[21] M. Alabbadi and A. Bajaber, "An intrusion detection system over the iot data streams using explainable artificial intelligence (xai)," *Sensors*, vol. 25, no. 3, p. 789, 2025.

[22] S. Lim *et al.*, "Explicate: Enhancing phishing detection through explainable ai and llm-powered interpretability," arXiv preprint arXiv:2503.20796, 2025.

[23] M. Gomez *et al.*, "Multimodal deep learning for vishing attacks in voice-activated iot systems," *IEEE Access*, vol. 12, pp. 34 567–34 589, 2024.

[24] Y. Zhang *et al.*, "Leveraging large language models for explainable phishing detection in iot," *IEEE Transactions on Information Forensics and Security*, 2025.

[25] T. Nguyen *et al.*, "Adversarial training for robust phishing detection in iot networks," *Security and Communication Networks*, vol. 2025, pp. 1–15, 2025.

[26] J. Kim *et al.*, "A multimodal voice phishing detection system integrating text and audio analysis," *Applied Sciences*, vol. 15, no. 5, p. 1123, 2025.

[27] Anonymous, "Fedphishllm: A privacy-preserving and explainable phishing detection mechanism using federated learning and llms," *Journal of King Saud University - Computer and Information Sciences*, 2025.

[28] Various, "Lstm-jso framework for privacy preserving adaptive intrusion detection in federated iot networks," *Scientific Reports*, 2025.

[29] ——, "Federated learning of explainable ai (fedxai) for deep learning-based intrusion detection in iot networks," *Computers Electrical Engineering*, 2025.

[30] A. Singh *et al.*, "Federated learning approaches for privacy-preserving phishing detection in iot," *Future Internet*, vol. 16, no. 5, p. 167, 2024.

[31] Various, "Explainable artificial intelligence in web phishing classification on secure iot with cloud-based cyber-physical systems," *Alexandria Engineering Journal*, 2024.

[32] ——, "Explainable artificial intelligence for intrusion detection in iot networks: A deep learning based approach," *Expert Systems with Applications*, 2023.

[33] J. Torres *et al.*, "Phishing detection in iot networks using machine learning: A systematic review," *Springer Book Chapter*, 2025.

[34] Q. Wang *et al.*, "Quantum-inspired algorithms for phishing detection in next-gen iot," *Quantum Machine Intelligence*, vol. 7, no. 1, pp. 12–28, 2025.

[35] D. Javeed *et al.*, "Trust my ids: An explainable ai integrated deep learning-based transparent threat detection system for industrial networks," *Computers Security*, 2024.

[36] E. Martinez *et al.*, "Blockchain-integrated systems for phishing prevention in iot," *Journal of Blockchain Research*, vol. 2, no. 2, pp. 45–60, 2024.

[37] M. Siganos *et al.*, "Explainable ai-based intrusion detection in the internet of things," *Proceedings of ACM ARES*, 2023.

[38] T. Miller *et al.*, "Human-centered explainable ai for cybersecurity in iot," *ACM Computing Surveys*, 2024.

[39] L. Xiao *et al.*, "Edge-fog-cloud architecture for lightweight xai in iot security," *Future Generation Computer Systems*, 2025.

[40] R. Kalakoti *et al.*, "Improving transparency and explainability of deep learning based iot botnet detection using explainable artificial intelligence (xai)," *IEEE Access*, 2024.

[41] H. Wang *et al.*, "Continual learning for adaptive phishing detection in evolving iot networks," *Neural Networks*, 2025.

[42] Y. Fang *et al.*, "Zero-day phishing detection using federated continual learning," *IEEE Transactions on Dependable and Secure Computing*, 2024.

[43] R. Johnson *et al.*, "Human-centered xai design for phishing alerts in iot user interfaces," *International Journal of Human-Computer Studies*, vol. 183, pp. 103–120, 2025.

[44] Various, "Evaluating machine learning-based intrusion detection systems with explainable ai: enhancing transparency and interpretability," *Frontiers in Computer Science*, 2025.

[45] A. Holzinger *et al.*, "Towards standardized explainable ai benchmarks for iot security," *Information Fusion*, 2025.

[46] A. Gupta *et al.*, "Quantum machine learning approaches for anomaly detection in iot security," *EPJ Quantum Technology*, 2025.

[47] V. Mothukuri *et al.*, "Privacy-preserving explainable ai for iot intrusion detection," *Journal of Information Security and Applications*, 2024.

[48] F. Tao *et al.*, "Digital twin enhanced proactive phishing threat modeling in iot," *IEEE Internet of Things Journal*, 2025.

[49] J. Smith *et al.*, "Novel explainable ai frameworks for real-time phishing mitigation in iot networks," *Journal of Cybersecurity and Privacy*, vol. 5, no. 2, pp. 234–256, 2025.

[50] M. Alkaeed *et al.*, "Blockchain-based secure phishing detection framework for iot," *Computers Security*, 2024.