**Journal of Al-Turath University College (JATUC)**
Open Access Journal
ISSN: 2074-5621
**DOI:** https://doi.org/10.63964/JATUC.43.2.2026.35
https://jatuc.uoturath.edu.iq/

# AI-powered threats: A Systematic Review of Generative Artificial Intelligence in Phishing Emails and Phishing URL

Ahmed R. AlKarawi [1*], Zied Othman Ahmed [2]

[1,2]Department of Computer science, College of Science, Mustansiriyah University, Baghdad, Iraq.

[*]*Correspondence Email*: Prahmed96@uomustansiriyah.edu.iq

**KEYWORDS**

Phishing; Phishing detection; Social engineering; Generative AI; URL phishing; Email phishing.

**ABSTRACT**

Recent breakthroughs in generative artificial intelligence (Gen-AI) and large language models (LLMs) have improved productivity across many domains but also increased the level of phishing attack realism and targeting. This systematic review examines recent studies on (i) the application of LLMs and other Gen-AI technologies to phishing email composition and phishing URL creation or obfuscation, (ii) phishing email and URL detection, and (iii) human factors in phishing vulnerability. The review considered studies published between 2020 and 2025 and compared their performance using accuracy, recall, precision and F1-score metrics. The reviewed indicates that Machine Learning (ML), Deep Learning (DL), and ensemble learning have been widely used in phishing detection. For email phishing detection, BERT transformer models have shown the highest performance in all studies reviewed. For URL phishing detection, hybrid models that combine structural and lexical information have shown the best performance. The review also emphasizes the role of human factors, such as familiarity of the sender and the fear of missing out (FoMO), in influencing phishing vulnerability. Overall, the findings support combining high-performing detection models with user-centered warnings and training to improve real-world phishing resilience.



## 1. INTRODUCTION

Large language models (LLMSs) are introducing new threats in cybercrime, phishing, and social engineering. Because they can generate highly realistic and tailored content, LLMSs improve the quality and scope of phishing and make it more undetectable. Multimodal generative applications enable malicious actors to use AI-generated images, text, video and audio, increasing attack vectors and creating more credible attacks [1].

The preferred technique for distributing phishing assaults was spamming, which was distinguished by its high-volume, poor-quality methodology. Although phishing has already been shown to be improved by using AI-enhanced approaches, these attacks may readily be progressed with higher levels of complexity and deployed on previously unheard-of scales with the use of LLMSs. These algorithms can produce extremely targeted and persuasive phishing content that can be mistaken for authentic communications, making detection even more challenging. For example, it has been demonstrated that sophisticated models like OpenAI's GPT-3.5-Turbo and GPT4 can produce realistic and customized SpearPhishing emails at scale for pennies, combining the worst features of both generic and highly focused phishing techniques [1].

The Sophos team of AI conducted a study to uncover the threats that AI-generated phishing by using LLMSs that can provide a wealth of knowledge with simple prompts, making it possible for anyone with minimal coding experience to write code to generate a simple scam website and fake images or copy a well-known webpage such as Facebook login page. Thus, introducing URL phishing [2].

The 2024 Cyber Security Breaches Survey conducted by the UK government in 2024, About a third of charities (32%) and half of corporations (50%) say they have been the victim of a cyber-security breach or attack in the past 12 months. For medium-sized enterprises (70%) and large businesses (74%), as well as high-income charities with yearly incomes of £500,000 or more (66%), this is significantly higher.[3] Phishing is by far the most prevalent kind of breach or assault, affecting 84% of enterprises and 83% of charitable organizations. To a far lesser degree, this is followed by people posing as organizations in emails or online (35% of businesses and 37% of charities), followed by malware or viruses (17% of businesses and 14% of charities) [3].

According to the Hoxhunt Phishing Trends Report (Updated for 2025), which offers some important phishing statistics for 2025, phishing has increased by 49% since 2021, when BlackHat AI first appeared. Of these attacks, 65% target companies, while 35% target individuals. Approximately 90% of malicious attachments result in more social engineering, and between 0.7% and 4.7% of phishes created by AI are reported. The top 3 most spoofing companies are Microsoft, DocuSign, and Human Resources [4]. The average yearly cost of phishing increased from $4.45 million to $4.88 million between 2023 and 2024, an almost 10% increase, according to the 2024 IBM/Ponemon Cost of Data Breach research. Since the pandemic, that is the largest jump [4].

Understanding the characteristics of AI-based social engineering attacks and their mechanisms is essential for individuals and organizations to develop proactive and defensive tactics to counteract such threats.

### 1.1 Social Engineering with Gen-AI

Generative AI Social Engineering Framework "provides a high-level overview of AI-enabled SE, but its application to a particular scenario or employment in studies will require adaptation and experimentation. A study by Schmitt and Flechais [5] shows the capabilities of this framework and categorized into:

1. Realistic content generation: Gen-AI creates indistinguishable content that mimic genuine human output, enhancing deception and believability.
2. Automated attack infrastructure: AI accelerates the attack process, enabling simulations of big, scaled attack waves that challenge traditional defensive systems.
3. Advanced targeting and personalization leveraging vast datasets: AI tailors attacks to individual targets, boosting the chance of successful manipulation Also there is another part mentioned the AI-driven attacks as AI amplify the effects of SE and Phishing attacks due to realistic content and personalized targeting which led to concerning industrial-scale attack patterns with significant impact on human computer interaction (HCI) [5].

A study examines two underexplored aspects affecting young people' vulnerability to social media phishing: the user's relationship with the message originator and Fear of Missing Out (FoMO). And the result was as follows [6]:

**Table 1** Frequency table of the coded (non)-susceptible responses and the response options (multiple answers possible) by user-sender relation (unknown, known sender)[6].

| | | ***unknown sender*** | ***known sender*** |
|---|---|---|---|
| **non-susceptible responses** | | 544 | 132 |
| **susceptible responses** | | 31 | 440 |
| **Responses split into response options** | | | |
| **non-susceptible** | ignore the message | 359 | 102 |
| | delete the message | 272 | 39 |
| **susceptible** | open the link | 6 | 257 |
| | share the link | 0 | 11 |
| | respond to the message | 14 | 295 |
| | like the message | 11 | 133 |

This means that if the vulnerable message was sent from an unknown source there is a 19.35% chance that it would be opened but if it was sent from a known source this probability would increase to 58.4% thus using a familiar name would double the chances of successful attack by about 40%.

To deal with such problems there is a need to emphasize human-focused anti-phishing behavior and the need for holistic interventions targeting the complete set of behavioral causes. Drivers which act internally vary from individual traits, personality, prior experience, and biology to company culture, social influence, system design, training, and crises (budget or health) [7].

### 1.2 Summary

The aim of this research is to describe and analyze the most recent trends of LLMS-Powered attacks and detection methods and compare them in accordance with standard evaluation metrics. It wants to understand how attackers would be able to use AI/ML to perform advanced phishing attacks and provide inputs for threat intelligence to build future defenses.

**Methodology and synthesis plan**

This review was conducted as a systematic literature review (SLR) and reported in accordance with the PRISMA 2020 guideline. The review protocol was defined before screening and specified the research questions, search sources, keyword groups, inclusion and exclusion criteria, screening workflow, quality assessment, data extraction, and synthesis plan. The protocol was not formally registered in an external repository. The literature review includes publications from the period between 2020 and 2025. The study focuses on the twin aspects of the role of Generative AI/LLMSs in (i) facilitating phishing attacks (email and URL vectors), and (ii) assessing how modern techniques (ML, DL and transformers) employed in the detection of phishing attacks.

Below is a highlighted review objective, research questions, search strategy, inclusion/exclusion criteria, paper selection and syntheses.

### 1.3 Objectives

- Summarize how Gen-AI/LLMSs are leveraged to create or augment phishing content, URL generation/obfuscation).
- Compare the mechanisms for detecting phishing emails and URLs with ML/DL/transformer/LLMS-based approaches, with a focus on the evaluation metrics used and realism of datasets.
- Identify human factors (e.g., sender, FoMO, behavioral factors) and translate them into deployment recommendations.

### 1.4 Research Questions (RQs)

- RQ1: How do Gen-AI/LLMS create sophisticated, large scale or highly personalized phishing Emails and URLs?
- RQ2: What are the most effective detection model families (ML/DL/transformer/LLMS/Hybrid) for phishing emails and/or phishing URLs, across reviewed studies?
- RQ3: How do dataset characteristics, such as size and feature type, affect model performance (intra URL, content, HTML, processes etc.)
- RQ4: How do human factors play a vital role in the world of phishing and how do they increase susceptibility to such attacks?

### 1.5 Study Selection Criteria

The selection criteria were defined to address the research questions and reduce selection bias.

#### 1.1.1 Inclusion criteria

- Published between (2020-2025)
- Peer-reviewed journal or reputable conference paper.
- Written in English language
- Have direct relevance to

- Gen-AI/LLMS uses in phishing
- Phishing detection using ML/DL/Transformers/LLMSs
- Provide evidence of human factor role in phishing

**1.1.2 Exclusion Criteria**

- Papers focusing on traditional phishing detection methods (e.g., blacklist, rule or signature base) were excluded
- Papers that focus only on LLMS and AI without relation to phishing
- None-research items (News, posts, tutorials, etc.)

**1.6 Search strategy**

The databases and repositories used for this extensive search were IEEE Xplore, ACM Digital Library, Scopus, ResearchGate, and arXiv. For this, the set of queries with keywords used from different concept groups, namely, phishing terminology, generative AI/LLMS terminology, phishing channels, and detection/defense terminology.

This ensured the retrieval of papers on phishing or social engineering attacks (email or URL-based) utilizing generative AI/LLMS or other ML methods, as well as papers on phishing detection. The search was restricted to the 2020-2025 period. The initial search on the above sources retrieved a total of 80.

**1.1.3 Search execution and filters**

The search was conducted in the November 2025 review period across the databases IEEE Xplore, ACM Digital Library, Scopus, Research Gate, and arXiv. The search was restricted to articles published from 2020 to 2025. The research focused on recent developments in GenAI/LLM-based phishing and recent phishing detection techniques. The search was restricted to English-language publications only. News articles, blogs, tutorials, and other non-scholarly writings were excluded.

**1.1.4 Search string design**

Search queries were constructed by combining terms from four concepts: (1) phishing and social-engineering terminology, (2) GenAI/LLM terminology, (3) phishing delivery channel (email and URL/website), and (4) detection and/or defense terminology. Boolean operators (AND/OR) were used to combine synonymous terms and preserve conceptual coverage.

Representative query example: ("phishing" OR "social engineering") AND ("generative artificial intelligence" OR "large language model" OR LLM) AND (email OR URL OR website) AND (detection OR classifier OR defense).

**1.7 Paper Selection and Syntheses**

Following this process, there were 66 unique documents left after removing duplicates (14 documents). The titles and abstracts of these documents were then screened against the inclusion criteria by two reviewers independently. As a result of this process, 15 documents that were obviously irrelevant or out of scope for inclusion were excluded from further review. The full texts of the remaining documents (51 documents) were then assessed for inclusion criteria. At this stage of the process, each document was assessed in-depth, and 3 documents were excluded because they did not meet all the inclusion criteria. At this stage of the process, there remained 48 documents that met all the selection criteria and that were included for qualitative synthesis. Disagreements at any step of this process were solved through discussion and consensus between reviewers. The process of selecting documents for inclusion in this review is represented in the PRISMA flow diagram in Figure 1.

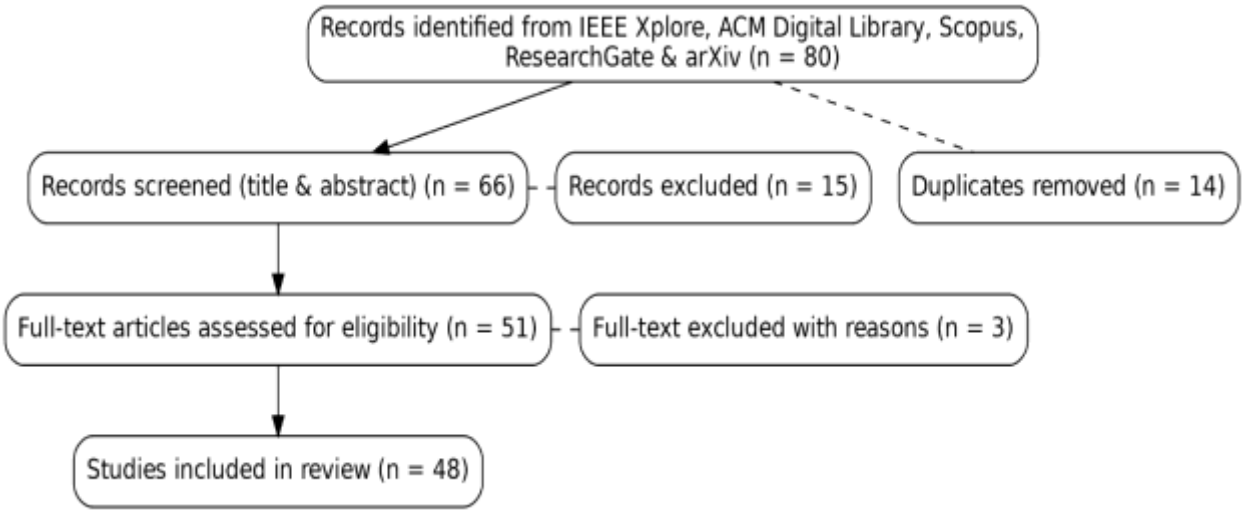


**Fig 1**. PRISMA 2020 flow diagram summarizing the study selection process

To improve transparency and reduce the influence of weak evidence, a structured quality assessment was applied to all included studies. Each paper was evaluated against the following criteria:

- QA1: Clarity of study objective and scope.
- QA2: Clear description of model architecture, features, or detection pipeline.
- QA3: Reporting of evaluation metrics (e.g., accuracy, precision, recall, and F1-score).
- QA4: Discussion of limitations, bias, or generalizability.

Each criterion was scored as 1 (yes), 0.5 (partially), or 0 (no). Quality ratings were used to guide interpretation during synthesis. The inclusion of studies was not based on quality score; however, more inferior studies are interpreted more cautiously when comparing performance claims.

### 1.1.5 Synthesis Method

Because datasets and metrics are heterogeneous, thus performing a narrative synthesis supported by structured tables:

- one synthesis for **email** detection,
- one synthesis for **URL/website** detection,
- one synthesis for **human factors**, translated into deployment recommendations.

## 2. LITERATURE REVIEW AND DISCUSSION

The selected methods and research in this review focus on two main parts the first one how AI and LLMS interfere in phishing and other malicious uses. On the other hand, examine how to detect these phishing attacks and what are the defensive strategies used nowadays that include new methods using ML and DL instead of traditional intrusion detection systems (IDS).

### 2.1 Current Use of Gen-AI Assisted Attacks

Four current uses of Gen-AI assisted attacks include phishing and social engineering, where Gen-AI creates hyper-personalized and convincing messages; malware creation, including the development of novel, evasive code; deepfake generation, which uses realistic audio and video to impersonate trusted individuals; and automated vulnerability discovery and exploitation, where AI rapidly identifies weaknesses in systems for attacks.

A review led by Zhang, Bu and Wen discussed the most famous large language models in recent years and some of the research id conducted to test its ability in offensive cyber security.

### 2.1.1 ChatGPT

One of the most famous Gen-AI nowadays are LLMSs The Chat Generative Pre-Training Transformer (GPT) was selected as an illustrative example and 10 research papers published in recent years were taken containing some use of ChatGPT in offensive cyber-Security.

Alawida M, Abu Shawar B, Abiodun OI, Mehmood A, Omolara AE, Al Hwaitat AK [9] outliened more than 8 applications of ChatGPT for Offensive Security and they are as follows:

1. Malware Creation: Attackers can create ever-changing polymorphic malware that bypasses standard antiviruses.
2. Social Engineering/Phishing: ChatGPT can create realistic messages to be used for such scams.
3. SQL Injection Attacks: It can help attackers develop SQLi payloads and even offer guidance on how to conduct such an attack.
4. Macros and LOLBINs: An attacker could place a link or file in an email and use ChatGPT to create macros that autorun when a spreadsheet is opened.
5. Vulnerability Scanning: Given sufficient input, it can parse and generate human-readable recommendations, assisting to perform scanning tasks to some extent.
6. Reporting Findings: ChatGPT can be used to automate detection/response reports, pretty much creating a quintessentially challenging attack-phase task within reach even for beginner hackers.
7. IT System Vulnerability Assessment: It can be used to scan or find code vulnerabilities.
8. Breach Notifications: Attackers can provide fake breach notifications to victims as part of a SE tactic.

Sharma and Dash [10] highlight the possible advantages and disadvantages of ChatGPT for cybersecurity. They highlighted a variety of cybersecurity concerns, such as phishing, password attacks, and malware attacks. The possible use of ChatGPT in social engineering assaults is also mentioned.

Gupta, Akiri and Aryal [11] worked on the impact of Gen-AI in cybersecurity and privacy and had similar outcomes.

Deng, G. et al [12] created PentestGPT is an automated penetration-testing tool constructed with three modules—inference, generation, and parsing each plays a different role on a pen testing team to more realistically simulate attacks.

Begou, Vinoy and Duda [13] Used An LLMS-driven method automates sophisticated phishing by cloning target sites, altering login forms to capture credentials, obfuscating code, registering domains, and deploying scripts.

Roy, Thota and Naragam [14] Examined models like ChatGPT, GPT-4, Bard (Gemini) and Claude the work shows these models readily produce persuasive phishing webpages and emails, spoofing popular brands and using evasive methods. The team also develops a BERT-driven tool that reliably detects malicious prompts to curb misuse.

Francia, Hansen and Schooley [15] Conducted a study compares GPT-4-written smishing texts to human-written ones and shows that LLMS messages are typically judged more convincing. Participants couldn't tell AI from human authorship or name reliable cues to do so, heightening the risk of personalized, AI-driven social engineering.

Charan, Chunduri and Anand [16] paper argues that LLMSs can be used to craft attack payloads, showing that ChatGPT and Bard can produce runnable code against the 2022 MITRE Top 10 weaknesses—and that these AI-generated payloads are typically more sophisticated and targeted than hand-made ones.

Tann and Sim [17] probe how well existing LLMSs handle CTF challenges by sampling typical categories and assessing GPT-3.5, PaLM2, and Prometheus. They find LLMSs offer partial but not comprehensive help.

Beckerich, Plein and Coronado [18] by abusing ChatGPT as a relay to the attackers' command-and-control, adversaries can operate the victim's machine without a direct connection, complicating attribution.

### 2.1.2 Other types of LLMS studies

Moskal, Laney and Hemberg [19] explore how LLMSs can support network threat testing—helping with threat actions and decision-making. Using virtual machines, they detail how LLMS-guided automated attacks can be launched on networked devices. Although preliminary, their results suggest LLMSs have strong potential for cyber threats. Lin, Cui and Liao [20] provided a study surveys 212 malicious LLMSs ("Malla"), revealing their dissemination channels and operational patterns across the underground economy. It dissects the ecosystem, tool chains, and exploit techniques, measures content-generation effectiveness and offers strategies to counter such cybercrime. Happe and Cito [21] explores high-level planning and hands-on vulnerability hunting in vulnerable VMs, linking LLMS-generated operations with VM feedback so the model can analyze the current state, find issues, and

recommend next steps for exploitation. Huang and Zhu [22] Emphasizing an end-to-end approach, they present PenHeal, a two-phase LLMS framework that both identifies vulnerabilities and guides mitigation, pairing a Pentest Module with a Remediation Module to propose best-practice remedies.

Pratama, Suryanto and Adiputra [23] develop CIPHER (Cybersecurity Intelligent Penetration-testing Helper for Ethical Researchers), which is trained on more than 300 reports, technique guides, and OSS tool documentation. They also add the FARR (Findings–Action–Reasoning–Results) augmentation, creating an automated Pen Test simulation benchmark tailored to LLMSs. Xu, Stokes and McDonald [24] propose Auto Attacker, which uses LLMSs to execute "keystroke-operated" attacks that mimic human behavior. By iteratively coordinating modules for summarization, planning, and choosing actions, it produces targeted commands across varied settings and streamlines what would otherwise be manual operations. Wang, Wang and Jung [25] an additional end-to-end, automated platform for creating and simulating cyberattacks. It can create emulation infrastructure, carry out attack operations, and independently create multi-phase cyberattack strategies based on CTI information. Usman, Upadhyay and Gyawali [26] developed Occupy AI is a tailored, good-tuned LLMS created to automate and perform cyberattacks, capable of outlining attack stages and producing runnable code for threats like phishing, malware injection, and system exploitation.

Happe, Kaplan and Cito [27] their work leverages LLMSs to aid penetration testing, introducing an automated Linux privilege-escalation benchmark to compare model performance and a tool, WinterMute, for rapidly probing how well LLMSs can bootstrap privilege escalation. Prapty, Kundu and Iyengar [28] They investigate automating attack graphs with LLMSs, mapping CVE relationships through preconditions and results, and extracting complete graphs from written threat analyses.

### 2.2 Defensive Methods Against Phishing

Until now, it is seen that the user has become so vulnerable due to phishing. But with proper precautions, one can avoid such scams. Some of the ways to protect users against phishing attacks by using authorized sources, avoid replying to suspicious messages checking the URL and Avoid public Wi-Fi. But it's not convenient for users to check all these things every time they receive an email or access a website, so researchers developed phishing detection tools and systems to do it for them.

One of the most used techniques used today in these tools is Machine learning (ML) and deep learning techniques (DL). An investigation was conducted to see how traditional ML models and transformer models work on the classification task of identifying if an email is phishing or not. The researchers realized that transformer models, in particular distilBERT, BERT, and RoBERTa, had a significantly higher performance compared to traditional models like Logistic Regression, Random Forest, Support Vector Machine, and Naive Bayes [29].

Among the classical models, SVM worked best, recording 0.9876 accuracy with well-balanced precision, recall, and F1 for phishing and non-phishing. All transformer models performed the same, frequently crossing 0.9881 accuracy, suggesting BERT-like models are especially effective in phrasing phishing as distinct from authentic emails [29].

As for URL detection, another study compared eight classification algorithms including K-nearest neighbors (KNN), Random Forest (RF), Support vector machines (SVM), Decision tree (DT), XG Boost, Logistic regression (LR), CNN model, and Deep learning model were assessed for their inherent capabilities [30]. The results were as follows.

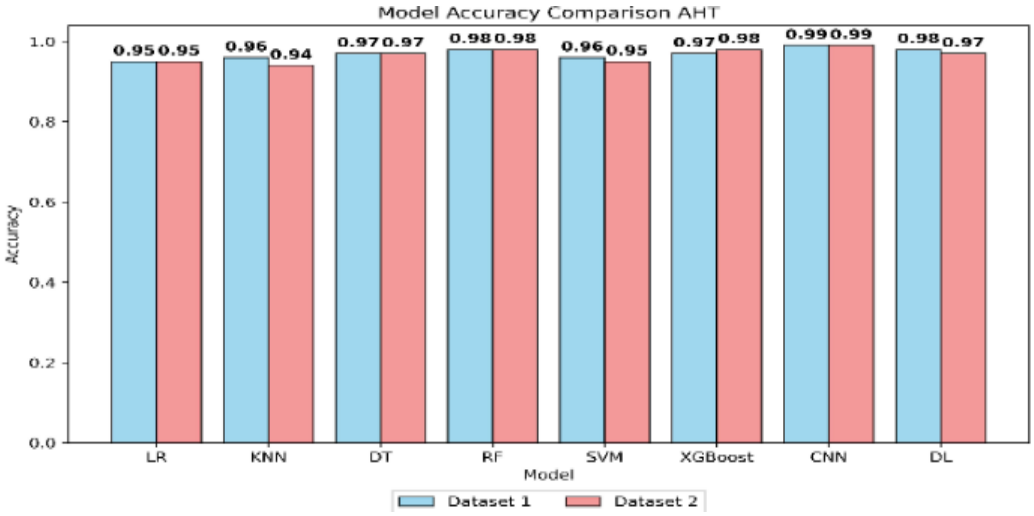


**Fig 2**: Model accuracy comparison AHT [30]

CNN is notably more accurate, highlighting its effectiveness.

Another research uses LLMS themselves to identify the Phishing attacks through benchmarking investigation of 21 cutting-edge open-source LLMSs for phishing URL detection, including Llama 3, Gemma, Qwen, Phi, DeepSeek, and Mistral. The findings show that without fine-tuning, big open-source LLMSs (≥27B parameters) obtain over 90% F1-score, which is comparable to proprietary models [3].

### 2.1.3 Detection of Email phishing

Table 2: Email phishing detection techniques which include a group of 8 studies each study includes the model's name, datasets, the four metrics (Accuracy, Precision, recall and F1 score) and the number of classes.

**Table 2:** Email phishing detection techniques

| No | Model | Dataset | Accuracy | Precision | F1 score | Recall | Class |
|---|---|---|---|---|---|---|---|
| **1** | 1D-CNNPD with Leaky ReLU and Bi-GRU (gated recurrent unit) [32] | 1-Mason, J. The Apache SpamAssassin 2-Public Corpus. 2005 Nazario. Phishing Corpus 2015. | 99.68% | 100% | 99.66% | 99.32% | 2 |
| **2** | MAchine Learning for LanguagE Toolkit (MALLET) with Naive Bayes [33] | DeepAI chatbot 865, The Enron email dataset, dataset of manually crafted regular and spam emails, | 99.3% | 98% | 98.5% | 99% | 4 |
| **3** | Continual Learning-based methods, i.e., EWC method. [34] | 1-PhishTank: A community platform for 2-reporting phishing websites.2022 VirusTotal. Virustotal: A community platform for reporting malicious payloads.2022 | 97.5% | 97.5% | 98.1% | 98.6% | 2 |
| **4** | Semantic watermarking [35] | Francesco Greco | 98.5% | 98.5% | 98.5% | 98.5% | 4 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Human-LLMS generated phishing-legitimate emails 2024 | | | | | |
| **5** | a Deep Q-Network (DQN) [36] | Self-collected from real-world | 95% | 96% | 94.98% | 94% | 2 |
| **6** | Bidirectional Encoder Representations from Transformers (BERT) Real-time [37] | 1-Kaggle-Phishing Email Detection<br>2- Al-Subaiey et al.<br>3-Zenodo. Phishing email<br>4-github: Apaulgithub/oibsip_taskno4 | 99.03% | 99.26% | 99% | 98.75% | 2 |
| **7** | Long Short-Term Memory (LSTM) and the Gated Recurrent Unit (GRU) [38] | Kaggle Enron Spam Data | 90% | 90% | 84% | 84% | 2 |
| **8** | EM-BERT and SPCA-BASED EAI-SC-LSTM [39] | SMS phishing Collection (SSC) Dataset", "Phishing Email Curated (PEC) Dataset", and "Webpage Phishing Detection (WPD) Dataset" | 99.62% | 99.89% | 99.56% | 99.78% | 2 works across email, SMS, URL |

- Altwaijry, Al-Turaiki, Alotaibi, and Alakeel, [32] used the 1D-CNNPD with leaky ReLU and Bi-GRU model the model achieved the high accuracy, however the dataset used is relatively old (2015), which may limits robustness against newer phishing attacks.
- Eze, C. S., & Shamir, L [33] used the MALLET with naive Bayes model performed strongly according to metrics but the test sample was too small (100 emails per class), which limits generalizability.
- Ejaz, Mian, and Manzoor [34] the continual learning is useful for addressing the problem of evolving methods of phishing. Boosting the accuracy a further more would make the model suitable for implementation in real-time.
- Brissett, A., & Wall, J [35] worked on the watermarking to detect of AI and human written phishing emails. The result of the model are excellent and the use of 4 class not just binary classification is also a useful contribution. However the used dataset is small in size and may not support robust generalization. 4000 email is small compared to other researches that used over 100k emails to train and test their models.
- Jabbar, H., & Al-Janabi, S [36] developed a Deep Q-network to enhance detection accuracy but 95% is surpassed by many other models in addition to that the dataset is only 5000 emails which is not preferred in building a generalized model.
- Mun, H., Park, J., Kim, Y., Kim, B., & Kim, J [37] worked on the Real-time BERT model which performed very well on 4 different datasets. And the only drawback to this model is Its incredibly lengthy training and inference timeframes presented efficiency issues for the implementation of real-world systems.
- Saleem, Islam, Hasan, Akbar, Khan, and Ibrar [38] developed a hybrid model that contain the LSTM and GRU but the model achieved relatively low accuracy compared with other reviewed methods.
- Murhej, M., & Nallasivan, G [39] the EM-BERT and SPCA based EAI-SE-LSTM model shows promising results and could be the optimal model since the evaluation metrics are over 99% and the model is tested with 3 different datasets. In addition the model works on Emails, SMS and URLs making it a appropriate overall phishing detection system. However the The authors note that the system does not include TLS/HTTPS-related signal features. However, as a number of phishing websites utilize legitimate TLS certificates, relying on TLS features alone is not a good way to determine a website's legitimacy. A better solution is to use a combination of URL, HTML, and content/behavioral features.

### 2.1.4 Detection of URL phishing

Table3: URL phishing detection techniques which include a group of 9 studies each study includes the model's name, datasets, the four metrics (Accuracy, Precision, recall and F1 score) and the number of classes.

| No | Model | Dataset | Accuracy | Precision | F1 score | Recall | Class |
|---|---|---|---|---|---|---|---|
| **1** | PhishHaven Real-time [40] | 1-Alexa, 2-PhishTank, 3-deepPhish | 98% | 98% | 98.06% | 98.06% | 2 |
| **2** | probabilistic neural network model deterministic neural network model [41] | Phishtank OpenPhish | 95% to 97% | - | - | - | 2 |
| **3** | AU-H-DNN/LSTM [42] | 1-ebbu, 2-Phishstorm, 3-ISCXURL, 4-catchPhish 1D, 5- CatchPhish 2D | 95.41% to 99.76% | 95.87% to 99.79% | 95.30% to 99.75% | 95.81% to 99.75% | 2 |
| **4** | TabNet ensemble model [43] | UNB URL dataset (ISCX-URL2016) | 97.8% | 97.8% | 97.8% | 97.6% | 5 |
| **5** | OFVA with RF [44] | 1-Aalto University's research data (Marchal, 2014<br>2- OpenPhish 2023 | 97.52% | 97.50% | 98% | 97.50% | 2 |
| **6** | Convolutional Neural Network (CNN) deep learning with Adam optimizer [45] | PhishTank PhishStats OpenPhish | 94.79% | 97% | 95.37% | 93.78 | 2 |
| **7** | XGBoost with URL and HTML [46] | PhishTank | 96.76 | 98.28 | 96.38 | 94.56 | 2 |
| **8** | Heterogeneous Machine Learning Ensemble Framework (Adaboost) [47] | Cyber Threats Analysis System (C-TAS) of the Korea Internet & Security Agency (KISA) and open-source intelligence (OSINT). | 95.16% | 95.26% | 95.16% | 95.04% | 2 |
| **9** | EM-BERT and SPCA-BASED EAI-SC-LSTM [39] | SMS Smishing Collection (SSC) Dataset", "Phishing Email Curated (PEC) Dataset", and "Webpage Phishing Detection (WPD) Dataset" | 99.62% | 99.89% | 99.56% | 99.78% | 2 |

- Sameen, M., Han, K., & Hwang [40] Developed PhishHave real-time model which has good accuracy but can detect only AI based Phishing URLs like that of DeepPhish features.
- Ghalechyan, H., Israyelyan, E., Arakelyan, A. et al. [41] worked on Probabilistic and deterministic neural networks models. When focusing on Accuracy, the probabilistic model performs better than the deterministic model by around 4%. However, both models did not reach optimal accuracy.
- Aung, E. S., & Yamana [42] proposed AU-H-DNN/LSTM and tested the model on 5 different datasets giving it an effective cross validation. But the model lacks robustness against attacks that hide malicious behavior beyond the URL text.
- Naseer, M., Ullah, F., Saeed, S., Algarni, F., & Zhao [43] suggested the Tab-Net ensemble model. The model classifies the URLs into 5 categories which is a useful contribution to move away from simple binary classification. And the accuracy achieved is relatively good. However, the dataset used is relatively old (2016) although the research conducted in 2024.

- Tamal, M. A., Islam, M. K., Bhuiyan, T.et al [44] They applied OFVA to distill 41 optimal intra-URL features forming a new dataset consisting of 274,446 raw URLs. The limitation of this study is that it overlooked the textual features beyond the URL in the content of the site itself.
- Loh, P. K. K., Lee, A. Z. Y., & Balachandran [45] used CNN with Adam optimizer as part of exploring the DL models, but the results accuracy less compared to other models and may not suit implementation in a real-world system in comparison to other models' results.
- Aljofey, A., Jiang, Q., Rasool, A., Chen, H. et al [46] developed the XGBoost with URL and HTML features which is an improvement over the use of just URL features in the detection model. However, the model did not achieve high accuracy compared to the other models like Transformer models. And HTTPS does not imply legitimacy; many phishing sites use valid certificates. since it does not depend on the content of the URL.
- Shin, S., Ji, S., & Hong, S [47] developed an Ensemble framework as a heterogeneous ML model using AdaBoost. But the result of the model is only 95% accuracy which is lower than competing methods reported in the reviewed studies.
- And for the Murhej, M., & Nallasivan, G [39] the EM-BERT is discussed earlier in the Email section since it works on both email and URL.

## 3. CONCLUSION

The evidence reviewed converges on three main conclusions. The first one is EM-BERT model with the SPCA-BASED EAI-SC-LSTM [39] by Murhej, M., & Nallasivan It performed strongly when multimodal features were available. This framework has reported the highest accuracy and F1 scores among all the models surveyed. It operates by extracting message content (from SMS and emails), generating word embeddings using EM-BERT, and then applying an EAI-SC-LSTM classifier to decide whether a message is phishing or legitimate. Also applying a dynamic updateable blacklist is a good addition to any system for strengthening the work.

On the URL side, the same approach is used for extracting the features and performing the analysis, and SPCA and a CSOA algorithm are used for feature distillation and selection, respectively, with the page content being embedded with the use of the EM-BERT model. The legitimacy/phishing status of the URL will then be decided by the classifier and the blacklist updated. This multi-channel model demonstrates the potential of the content-based and lexical analysis along with the use of the transformer model [39]. But computational costs of BERT models may limit real-world deployment.

The second point, despite achieving high accuracy on benchmark datasets, model generalizability may be limited. Real-world deployment requires diverse datasets and robust validation.

In addition, many available datasets could make your model reach over 99% accuracy easily, but the final model is not well generalized. To address these limitations, models should be tested on real-world data or sites also preventing leakage in the data preprocessing is always a plus. Some other techniques like regularization are used to prevent overfitting and improve a model's ability to generalize to new unseen data.

Mixing URLs features with HTML features and including some textual content features will produce a more robust model for different kinds of phishing. And can be implemented in a real-world scenario. However, the model could still use the help of continual learning to be able catch up with the evolving phishing threats.

Phishing defenses are two types: preventive (user education) and detective (list-, rule-, similarity-, and ML-based). Because ML relies upon labeled data, lack of labels stalls strong tools. To compensate for this, Tamal, Islam, Bhuiyan, and Sattar suggest a big, labeled URL dataset for phishing detection: 247,950 samples—128,541 phishing and 119,409 legitimate [48] which is a good base to work with.

Using lexical features dataset only like M. A. Tamal, M. K. Islam, T. Bhuiyan, and A. Sattar, "Dataset of suspicious phishing URL detection [48] is very useful, especially when one considers that it does not require downloading website content, a potentially slow and dangerous process, which this dataset is well suited for real-time protection systems.

There is an inherent trade-off between speed and real-time deployment against the complexity of the system to see more than just one type of features (HTML, lexical URL, content, etc.) and ability to have more accurate results but in more time or in non-real-time manner.

The third takeaway is Human factors shift the risk surface. As Gen-AI increases message realism and personalization, foreshadowing signals such as sender familiarity and FoMO increase click intent and reduce suspicion thus phishing success increased from 19% to 58% when the sender was familiar.

Successful defense will integrate technical filters with just-in-time warning and user training that directly addresses these behavioral levers.

**Conflict of interest**

All authors have no actual, potential, or perceived conflicts of interest.

**Consent for publications**

All authors have read and approved the final manuscript for publication.

**Availability of data and material**

All the data supporting the findings of this study are included within the manuscript provided.

**Authors' contributions**

A.R.A conceptualized and investigated the study and performed the data curation and the writing and editing.
Z.O.A proposed the plan and provided reviewing and feedback to enhance the work in addition to oversight and leadership for the research planning.
And both authors read and approved the final version of the study.

**Funding**

Authors did not receive any external funding or financial support for this study

**Acknowledgement:**

The authors thank the CS department in the college of science from Mustansiriyah University for supporting this work.

## 4. REFERENCES

[1] A. Kucharavy, O. Plancherel, V. Mulder, A. Mermoud, and V. Lenders, Large Language Models in Cybersecurity: Threats, Exposure and Mitigation. Springer Nature, 2024.

[2] "The Dark Side of AI: Large-Scale Scam Campaigns Made Possible by Generative AI – Sophos News." Accessed: Oct. 04, 2025. [Online]. Available: https://news.sophos.com/en-us/2023/11/27/the-dark-side-of-ai-large-scale-scam-campaigns-made-possible-by-generative-ai/

[3] "Cyber security breaches survey 2024," GOV.UK. Accessed: Oct. 04, 2025. [Online]. Available: https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2024/cyber-security-breaches-survey-2024

[4] "Phishing Trends Report (Updated for 2025)." Accessed: Oct. 04, 2025. [Online]. Available: https://hoxhunt.com/guide/phishing-trends-report

[5] M. Schmitt and I. Flechais, "Digital deception: generative artificial intelligence in social engineering and phishing," Artif Intell Rev, vol. 57, no. 12, p. 324, Oct. 2024, doi: 10.1007/s10462-024-10973-2.

[6] J. Klütsch, J. Schwab, C. Böffel, V. Zimmermann, and S. J. Schlittmeier, "Friend or phisher: how known senders and fear of missing out affect young adults' phishing susceptibility on social media," Humanit Soc Sci Commun, vol. 11, no. 1, p. 1145, Sept. 2024, doi: 10.1057/s41599-024-03412-8.

[7] R. Jabir, J. Le, and C. Nguyen, "Phishing Attacks in the Age of Generative Artificial Intelligence: A Systematic Review of Human Factors," AI, vol. 6, no. 8, p. 174, Aug. 2025, doi: 10.3390/ai6080174.

[8] J. Zhang et al., "When LLMSs meet cybersecurity: a systematic literature review," Cybersecurity, vol. 8, no. 1, p. 55, Feb. 2025, doi: 10.1186/s42400-025-00361-w.

[9] M. Alawida, B. Abu Shawar, O. I. Abiodun, A. Mehmood, A. E. Omolara, and A. K. Al Hwaitat, "Unveiling the Dark Side of ChatGPT: Exploring Cyberattacks and Enhancing User Awareness," Information, vol. 15, no. 1, p. 27, Jan. 2024, doi: 10.3390/info15010027.

[10] P. Sharma and B. Dash, "Impact of Big Data Analytics and ChatGPT on Cybersecurity," in 2023 4th International Conference on Computing and Communication Systems (I3CS), Shillong, India: IEEE, Mar. 2023, pp. 1–6. doi: 10.1109/I3CS58314.2023.10127411.

[11] M. Gupta, C. Akiri, K. Aryal, E. Parker, and L. Praharaj, "From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy," IEEE Access, vol. 11, pp. 80218–80245, 2023, doi: 10.1109/ACCESS.2023.3300381.

[12] G. Deng et al., "PentestGPT: An LLMS-empowered Automatic Penetration Testing Tool," June 02, 2024, arXiv: arXiv:2308.06782. doi: 10.48550/arXiv.2308.06782.

[13] N. Begou, J. Vinoy, A. Duda, and M. Korczyński, "Exploring the Dark Side of AI: Advanced Phishing Attack Design and Deployment Using ChatGPT," in 2023 IEEE Conference on Communications and Network Security (CNS), Oct. 2023, pp. 1–6. doi: 10.1109/CNS59707.2023.10288940.

[14] S. S. Roy, P. Thota, K. V. Naragam, and S. Nilizadeh, "From Chatbots to PhishBots? -- Preventing Phishing scams created using ChatGPT, Google Bard and Claude," Mar. 10, 2024, arXiv: arXiv:2310.19181. doi: 10.48550/arXiv.2310.19181.

[15] J. Francia, D. Hansen, B. Schooley, M. Taylor, S. Murray, and G. Snow, "Assessing AI vs Human-Authored Spear Phishing SMS Attacks: An Empirical Study," Mar. 19, 2025, arXiv: arXiv:2406.13049. doi: 10.48550/arXiv.2406.13049.

[16] P. V. S. Charan, H. Chunduri, P. M. Anand, and S. K. Shukla, "From Text to MITRE Techniques: Exploring the Malicious Use of Large Language Models for Generating Cyber Attack Payloads," May 24, 2023, arXiv: arXiv:2305.15336. doi: 10.48550/arXiv.2305.15336.

[17] W. Tann, Y. Liu, J. H. Sim, C. M. Seah, and E.-C. Chang, "Using Large Language Models for Cybersecurity Capture-The-Flag Challenges and Certification Questions," Aug. 21, 2023, arXiv: arXiv:2308.10443. doi: 10.48550/arXiv.2308.10443.

[18] M. Beckerich, L. Plein, and S. Coronado, "RatGPT: Turning online LLMSs into Proxies for Malware Attacks," Sept. 07, 2023, arXiv: arXiv:2308.09183. doi: 10.48550/arXiv.2308.09183.

[19] S. Moskal, S. Laney, E. Hemberg, and U.-M. O'Reilly, "LLMSs Killed the Script Kiddie: How Agents Supported by Large Language Models Change the Landscape of Network Threat Testing," Oct. 10, 2023, arXiv: arXiv:2310.06936. doi: 10.48550/arXiv.2310.06936.

[20] Z. Lin, J. Cui, X. Liao, and X. Wang, "Malla: Demystifying Real-world Large Language Model Integrated Malicious Services," Aug. 19, 2024, arXiv: arXiv:2401.03315. doi: 10.48550/arXiv.2401.03315.

[21] A. Happe and J. Cito, "Getting pwn'd by AI: Penetration Testing with Large Language Models," in Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, in ESEC/FSE 2023. New York, NY, USA: Association for Computing Machinery, Nov. 2023, pp. 2082–2086. doi: 10.1145/3611643.3613083.

[22] J. Huang and Q. Zhu, "PenHeal: A Two-Stage LLMS Framework for Automated Pentesting and Optimal Remediation," in Proceedings of the Workshop on Autonomous Cybersecurity, in AutonomousCyber '24. New York, NY, USA: Association for Computing Machinery, Nov. 2024, pp. 11–22. doi: 10.1145/3689933.3690831.

[23] D. Pratama et al., "CIPHER: Cybersecurity Intelligent Penetration-Testing Helper for Ethical Researcher," Sensors, vol. 24, no. 21, p. 6878, Jan. 2024, doi: 10.3390/s24216878.

[24] J. Xu et al., "AutoAttacker: A Large Language Model Guided System to Implement Automatic Cyber-attacks," Mar. 02, 2024, arXiv: arXiv:2403.01038. doi: 10.48550/arXiv.2403.01038.

[25] L. Wang et al., "From Sands to Mansions: Enabling Automatic Full-Lifecycle Cyberattack Construction with LLMS," July 24, 2024, arXiv: arXiv:2407.16928. doi: 10.48550/arXiv.2407.16928.

[26] Y. Usman, A. Upadhyay, P. Gyawali, and R. Chataut, "Is Generative AI the Next Tactical Cyber Weapon for Threat Actors? Unforeseen Implications of AI Generated Cyber Attacks," Aug. 23, 2024, arXiv: arXiv:2408.12806. doi: 10.48550/arXiv.2408.12806.

[27] A. Happe, A. Kaplan, and J. Cito, "LLMSs as Hackers: Autonomous Linux Privilege Escalation Attacks," Feb. 18, 2025, arXiv: arXiv:2310.11409. doi: 10.48550/arXiv.2310.11409.

[28] R. T. Prapty, A. Kundu, and A. Iyengar, "Using Retriever Augmented Large Language Models for Attack Graph Generation," Aug. 11, 2024, arXiv: arXiv:2408.05855. doi: 10.48550/arXiv.2408.05855.

[29] R. Meléndez, M. Ptaszynski, and F. Masui, "Comparative Investigation of Traditional Machine-Learning Models and Transformer Models for Phishing Email Detection," Electronics, vol. 13, no. 24, p. 4877, Jan. 2024, doi: 10.3390/electronics13244877.

[30] N. F. Almujahid, M. A. Haq, and M. Alshehri, "Comparative evaluation of machine learning algorithms for phishing site detection," PeerJ Comput. Sci., vol. 10, p. e2131, June 2024, doi: 10.7717/peerj-cs.2131.

[31] A. H. Nasution, W. Monika, A. Onan, and Y. Murakami, "Benchmarking 21 Open-Source Large Language Models for Phishing Link Detection with Prompt Engineering," Information, vol. 16, no. 5, p. 366, May 2025, doi: 10.3390/info16050366.

[32] N. Altwaijry, I. Al-Turaiki, R. Alotaibi, and F. Alakeel, "Advancing Phishing Email Detection: A Comparative Study of Deep Learning Models," Sensors, vol. 24, no. 7, p. 2077, Jan. 2024, doi: 10.3390/s24072077.

[33] C. S. Eze and L. Shamir, "Analysis and Prevention of AI-Based Phishing Email Attacks," Electronics, vol. 13, no. 10, p. 1839, Jan. 2024, doi: 10.3390/electronics13101839.

[34] A. Ejaz, A. N. Mian, and S. Manzoor, "Life-long phishing attack detection using continual learning," Sci Rep, vol. 13, no. 1, p. 11488, July 2023, doi: 10.1038/s41598-023-37552-9.

[35] A. Brissett and J. Wall, "Machine Learning and Watermarking for Accurate Detection of AI-Generated Phishing Emails," Electronics, vol. 14, no. 13, p. 2611, Jan. 2025, doi: 10.3390/electronics14132611.

[36] H. Jabbar and S. Al-Janabi, "AI-Driven Phishing Detection: Enhancing Cybersecurity with Reinforcement Learning," Journal of Cybersecurity and Privacy, vol. 5, no. 2, p. 26, June 2025, doi: 10.3390/jcp5020026.

[37] H. Mun, J. Park, Y. Kim, B. Kim, and J. Kim, "PhiShield: An AI-Based Personalized Anti-Spam Solution with Third-Party Integration," Electronics, vol. 14, no. 8, p. 1581, Jan. 2025, doi: 10.3390/electronics14081581.

[38] S. Saleem, Z. U. Islam, S. S. U. Hasan, H. Akbar, M. F. Khan, and S. A. Ibrar, "Spam Email Detection Using Long Short-Term Memory and Gated Recurrent Unit," Applied Sciences, vol. 15, no. 13, p. 7407, Jan. 2025, doi: 10.3390/app15137407.

[39] M. Murhej and G. Nallasivan, "Multimodal framework for phishing attack detection and mitigation through behavior analysis using EM-BERT and SPCA-BASED EAI-SC-LSTM," Front. Commun. Netw., vol. 6, July 2025, doi: 10.3389/frcmn.2025.1587654.

[40] M. Sameen, K. Han, and S. O. Hwang, "PhishHaven—An Efficient Real-Time AI Phishing URLs Detection System," IEEE Access, vol. 8, pp. 83425–83443, 2020, doi: 10.1109/ACCESS.2020.2991403.

[41] H. Ghalechyan, E. Israyelyan, A. Arakelyan, G. Hovhannisyan, and A. Davtyan, "Phishing URL detection with neural networks: an empirical study," Sci Rep, vol. 14, no. 1, p. 25134, Oct. 2024, doi: 10.1038/s41598-024-74725-6.

[42] E. S. Aung and H. Yamana, "PhiSN: Phishing URL Detection Using Segmentation and NLP Features," Journal of Information Processing, vol. 32, pp. 973–989, 2024, doi: 10.2197/ipsjjip.32.973.

[43] M. Naseer, F. Ullah, S. Saeed, F. Algarni, and Y. Zhao, "Explainable TabNet ensemble model for identification of obfuscated URLs with features selection to ensure secure web browsing," Sci Rep, vol. 15, no. 1, p. 9496, Mar. 2025, doi: 10.1038/s41598-025-93286-w.

[44] M. A. Tamal, M. K. Islam, T. Bhuiyan, A. Sattar, and N. U. Prince, "Unveiling suspicious phishing attacks: enhancing detection with an optimal feature vectorization algorithm and supervised machine learning," Front. Comput. Sci., vol. 6, July 2024, doi: 10.3389/fcomp.2024.1428013.

[45] P. K. K. Loh, A. Z. Y. Lee, and V. Balachandran, "Towards a Hybrid Security Framework for Phishing Awareness Education and Defense," Future Internet, vol. 16, no. 3, p. 86, Mar. 2024, doi: 10.3390/fi16030086.

[46] A. Aljofey et al., "An effective detection approach for phishing websites using URL and HTML features," Sci Rep, vol. 12, no. 1, p. 8842, May 2022, doi: 10.1038/s41598-022-10841-5.

[47] S.-S. Shin, S.-G. Ji, and S.-S. Hong, "A Heterogeneous Machine Learning Ensemble Framework for Malicious Webpage Detection," Applied Sciences, vol. 12, no. 23, p. 12070, Jan. 2022, doi: 10.3390/app122312070.

[48] M. A. Tamal, M. K. Islam, T. Bhuiyan, and A. Sattar, "Dataset of suspicious phishing URL detection," Front. Comput. Sci., vol. 6, Mar. 2024, doi: 10.3389/fcomp.2024.1308634.

**Journal of Al-Turath University College (JATUC)**
Open Access Journal
ISSN: 2074-5621
**DOI:** https://doi.org/10.63964/JATUC.43.2.2026.35
https://jatuc.uoturath.edu.iq/

# التهديدات المدعومة بالذكاء الاصطناعي: مراجعة منهجية للذكاء الاصطناعي التوليدي في تصيّد البريد الالكتروني الاحتيالي وروابط مواقع التصيّد

احمد رياض الكروي [1*] ، زيد ثمان احمد [2]

[1،2] قسم علوم الحاسوب، كلية العلوم، الجامعة المستنصرية، بغداد، العراق

* *البريد الإلكتروني للتواصل:* Prahmed96@uomustansiriyah.edu.iq

**الملخص**
أدّت التطورات الحديثة في الذكاء الاصطناعي التوليدي (Gen-AI) ونماذج اللغة الكبيرة (LLMs) إلى تحسين الإنتاجية في العديد من المجالات، لكنها في الوقت نفسه زادت من مستوى الواقعية ودقة الاستهداف في هجمات التصيّد الاحتيالي. تستعرض هذه المراجعة المنهجية الدراسات الحديثة المتعلقة بـ: (1) استخدام نماذج اللغة الكبيرة وغيرها من تقنيات الذكاء الاصطناعي التوليدي في صياغة رسائل البريد الإلكتروني الاحتيالية وإنشاء روابط التصيّد الاحتيالي أو تمويهها، و(2) كشف رسائل البريد الإلكتروني الاحتيالية وروابط التصيّد، و(3) العوامل البشرية المرتبطة بقابلية التعرّض للتصيّد الاحتيالي. وقد شملت المراجعة الدراسات المنشورة بين عامي 2020 و2025، وقارنت أداءها باستخدام مقاييس الدقة(Accuracy) ، والاسترجاع (Recall) ، والإحكام (Precision)، ودرجة .F1 وتشير الدراسات المُراجَعة إلى أن تعلّم الآلة (ML) ، والتعلّم العميق (DL) ، والتعلّم التجميعي قد استُخدمت على نطاق واسع في كشف التصيّد الاحتيالي. وفيما يتعلق بكشف رسائل البريد الإلكتروني الاحتيالية، أظهرت نماذج المحوّل من نوع BERT أعلى أداء في جميع الدراسات التي شملتها المراجعة. أما في كشف روابط التصيّد الاحتيالي، فقد أظهرت النماذج الهجينة التي تجمع بين المعلومات البنيوية والمعلومات المعجمية أفضل أداء. كما تؤكد المراجعة دور العوامل البشرية، مثل ألفة المرسِل والخوف من فوات الفرصة (FoMO) ، في التأثير في قابلية التعرّض للتصيّد الاحتيالي. وبصورة عامة، تدعم النتائج الجمع بين نماذج الكشف عالية الأداء والتحذيرات المتمحورة حول المستخدم والتدريب، من أجل تحسين القدرة الواقعية على الصمود أمام هجمات التصيّد الاحتيالي.

**الكلمات المفتاحية**
التصيد الاحتيالي، كشف التصيد الاحتيالي، الهندسة الاجتماعية، الذكاء الاصطناعي التوليدي، التصيد الاحتيالي عبر الروابط.