Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Short Message Service (SMS) Phishing Attacks and Defenses: A Comprehensive Review

Mir Mehedi A. Pritom
mpritom@tntech.edu
Tennessee Technological University
Seyed Mohammad Sanjari
Tennessee Technological University
Maraz Mia
Tennessee Technological University
Ashfak Md Shibli
Tennessee Technological University
S M Mostaq Hossain
Tennessee Technological University
Muhammad Ismail
Tennessee Technological University
Shouhuai Xu
University of Colorado Colorado Springs

Research Article

Keywords: SMS Phishing · SMS scams · SMS fraud · smishing defense · smishing dataset · mobile threats

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11281923/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Eur. Phys. J. C manuscript No.
(will be inserted by the editor)

Short Message Service (SMS) Phishing Attacks and Defenses: A Comprehensive Review

Mir Mehedi A. Pritom (Corresponding Author) [a,1], Seyed Mohammad Sanjari[b,1], Maraz Mia[c,1], Ashfak Md Shibli[d,1], Mostaq Hossain[e,1], Muhammad Ismail[f,1,2], Shouhuai Xu[g,3]

[1]Department of Computer Science, Tennessee Tech University, Cookeville, TN 38505, USA address
[2]Cybersecurity Education, Research and Outreach Center (CEROC), Tennessee Tech University, Cookeville, TN 38505, USA
[3]Department of Computer Science, University of Colorado - Colorado Springs, Colorado Springs, CO 80918, USA

Received: date / Accepted: date

Abstract SMS Phishing (also known as 'smishing') is a growing deceptive social engineering (SE) attack that leverages mobile SMS to conduct cybercrimes such as stealing sensitive information or spreading malware by tricking users into interacting with attackers' messages (e.g., responding to or clicking URLs). This threat has increased rapidly in recent years, causing $639 million in financial losses for United States users in 2025 alone. This threat is also evolving rapidly, meaning that attackers continually adapt their tactics, reshaping the landscape. There is a significant body of literature on investigating smishing attacks and defenses. However, there is no systematic review that reflects the current attack and defense landscape along with available resources (i.e., relevant datasets). This motivates us to systematize the current smishing research efforts, including the following four research pillars: (a) *user perception and susceptibility*, (b) *attack characterization*, (c) *defense landscape*, and (d) *smishing datasets*. This leads us to propose new research directions to mitigate smishing attacks effectively.

Keywords SMS Phishing · SMS scams · SMS fraud · smishing defense · smishing dataset · mobile threats

1 Introduction

Mobile devices are ubiquitous as we rely on them for daily communications, banking, shopping, and other online activities. However, this reliance has made mobile users a prime target for cybercriminals attempting to steal *personally identifiable information* (PII), financial data and other sensitive information through fraudulent activities. This study focuses on one particular form of these cyber social engineering attacks, known as Short Message Service (SMS) phishing or 'smishing' for short [49].

These attacks often use urgency or fear tactics to deceive recipients into clicking malicious links or revealing personal information, by making incoming fraudulent messages believable to end-users. The threat of smishing attacks is real, as evidenced by the Federal Trade Commission (FTC) report of $639 million in financial losses in 2025 due to text-based scams or smishing (a massive 643% increase in five years from $86M reported in 2020) [43]. Moreover, a recent Anti Phishing Working Group (APWG) report [11] shows that phone-based phishing attacks, including smishing and vishing (voice phishing) have grown by 30% in the first quarter of 2024 when compared with the previous quarter. This makes it imperative to keep up with the rapidly evolving smishing threats, which motivates the present study.

The rapid evolution of smishing attacks can be justified as follows. Firstly, the available commercial anti-smishing tools are not robust enough [153]. Next, defenses such as blacklisting (based on message contents, including URLs within texts) are ineffective as attackers continuously adapt their tactics in employing stealthy techniques [58, 7]. Moreover, the introduction of Artificial Intelligence (AI)-based chatbots, such as ChatGPT, has ushered in a new threat vector at the disposal of smishing attackers [129, 51, 112]. Recent studies on new perspectives of smishing attacks, such as users' perception and susceptibility to smishing attacks

[a]e-mail: mpritom@tntech.edu
[b]e-mail: ssanjarip42@tntech.edu
[c]e-mail: mmia43@tntech.edu
[d]e-mail: ashibli42@tntech.edu
[e]e-mail: shossain42@tntech.edu
[f]e-mail: mismail@tntech.edu
[g]e-mail: sxu@uccs.edu

[104, 108, 145, 121, 29, 32, 6, 15], further highlights the need for a systematic understanding of the smishing threat and defense landscape.

Our Contributions. Despite the impact of these evolving threats and the urgent need for updated defense strategies, the literature lacks systematic reviews of smishing attack tactics, existing defense strategies, and future research directions. To address this gap, we focus on four major areas in this paper: (a) *user perception and susceptibility*, (b) *attack characterization*, (c) *defense landscape*, and (d) *smishing datasets*. First, we aim to understand current users' perception of, and susceptibility to, smishing based on past user studies. Next, we characterize smishing threats and outline a bird's-eye view of the smishing threat landscape. Then, we study existing defense and detection mechanisms and describe their strengths and weaknesses, which helps identify open challenges. Lastly, we collect and analyze all the major smishing and spam SMS datasets available and draw insights through a data-driven investigation, which can help researchers in selecting the right dataset based on their research needs. Our analysis reveals several cross-cutting trends across these four pillars. First, studies on user perception and susceptibility largely focus on demographic and behavioral factors, with limited longitudinal and real-world validation. Second, attack characterization research has primarily emphasized URL-based and theme-based campaign analysis, while emerging LLM-driven and adaptive tactics remain underexplored. Third, the defense literature is dominated by machine learning and deep learning-based detection approaches, with comparatively less emphasis on explainable, user-centric, and infrastructure-level solutions. Finally, existing smishing datasets are often limited in scale, diversity, and temporal coverage. This makes a constraint for reproducibility and long-term evaluation of proposed defenses. To summarize, this study makes the following key contributions:

- Deep-dive into the existing literature on user-study driven smishing research to understand current users' perception, behavior, and susceptibility when exposed to smishing messages.
- Characterize the SMS phishing threat and attack landscape via classification schemes that are applicable to existing smishing messages.
- Critically analyze the existing detection methods and approaches to defend against this evolving threat.
- Analyze the currently available public SMS datasets, their key features/characteristics, and models that leverage those datasets, and characterize their performances. We provide a *Smishing Data Hub* on GitHub for helping future smishing researchers.

Related Work. A brief summary comparison of related survey coverage on smishing research are provided in table 1 to illustrate the need for this study. As depicted, Al Saidat et al. [5] presented an overview of SMS spam detection; Edwards et al. [32] and Clasen et al. [29] studied the human element of smishing by analyzing end-user behaviors; Siddiqi et al. [133] examined the psychological mechanisms underlying social engineering attacks and reviewed existing countermeasures, including smishing-related threats; Sugunaraj et al. [143] provided a short review of cyber fraud targeting U.S. seniors, discussing scam types, communication vectors (including SMS), and preventive resources; Faklaris et al. [35] analyzed the demographic susceptibility of United State's users to smishing; Longtchi et al. [69] provided a systematic examination of internet-based social engineering (SE) attacks, including smishing, through a psychological lens where various psychological factors (PFs) are exploited by attackers using a set of psychological techniques (PTs).

Recently, Rahaman et al. [103] has presented a comprehensive survey of smishing attacks, reviewing machine learning, deep learning, hybrid, and IoT-based detection techniques, along with associated challenges and future research directions.

In addition, several surveys have focused on phishing website detection. For example, Vijayalakshmi et al. [168] presented a taxonomy and performance analysis of URL- and webpage-based detection techniques, while Safi and Singh [113] conducted a systematic review of machine learning and deep learning approaches for phishing websites. However, these studies primarily address web-based attacks rather than SMS-based social engineering threats. These studies touch upon some components of the smishing landscape, but lacking systematic treatment of the smishing landscape. For instance, the studies mentioned above did not cover new developments in the application of Large Language Models (LLMs) and Explainable AI (XAI) to enhance smishing detection and transparency. Moreover, there is no systematic comparison of publicly available smishing datasets, which is crucial to investigating defenses against smishing attacks. By contrast, the present study aims to present the first systematic treatment of smishing attacks and defenses.

Paper Outline. The rest of the paper is organized as follows. Section 2 describes the methodology. Section 3 systematizes user perceptions and susceptibility to smishing attacks. Section 4 characterizes the smishing attack/threat landscape. Section 5 systematizes the smishing defense landscape and existing methodologies. Section 6 assesses existing smishing and SMS datasets

Table 1: Comparison of Related Survey Works Across Key Dimensions

Ref.	User Perception & Susceptibility	Attack	Defense	Dataset Analysis	Regulatory & Policy	PF/PT	Literature Review
[5]	x	x	✓	x	x	x	✓
[69]	✓	✓	✓	x	x	✓	✓
[133]	✓	✓	✓	x	x	✓	✓
[143]	✓	✓	✓	x	✓	x	✓
[103]	x	✓	✓	x	x	x	✓
This Work	✓	✓	✓	✓	✓	✓	✓

available for research. Section 7 discusses future research directions while section 8 concludes the paper.

2 Scope and Methodology

This paper focuses on smishing attacks, defenses, and datasets, meaning that other kinds of cyber social engineering attacks (e.g., Internet-based cyber social engineering attacks, vishing, web phishing) are deemed out of scope.

Figure 1 presents a pictorial summary of the scope of this paper where we systematically study the four pillar areas of smishing research such as the *user perception and susceptibility*, *attack characterization*, *defense landscape*, and *smishing datasets*, which are elaborated below. Accordingly, this work follows a systematic literature review (SLR) methodology and presents a taxonomy-based synthesis supported by a structured search strategy, defined inclusion and exclusion criteria, and a PRISMA-based screening process (Figure 4).

- **User Perception and Susceptibility:** This pertains to user behaviors, perceptions, and especially user susceptibilities when exposed to smishing threats.
- **Attack Characterization:** This includes attack types, attack tactics, emerging threats, and PTs (psychological techniques).
- **Defense Landscape:** This systematizes the strengths, weaknesses, and transparency of existing detection mechanisms reported in the literature, while considering practical defense countermeasures, anti-smish tools, mobile-based solutions, user training, awareness modules, and existing regulatory efforts.
- **Smishing Datasets:** This characterizes publicly available SMS datasets and better summarizes the contents therein. This allows us to create a unified Smishing Data Hub provided at `https://github.com/MarazMia/SMISH_DT`.

Search Strategy and Selection Criteria. Given the defined scope of this study, we conducted a systematic literature review to identify relevant research on smishing across the Google Scholar and DBLP databases. The search queries utilized the following Boolean combinations of keywords: ("smishing" OR "SMS phish-

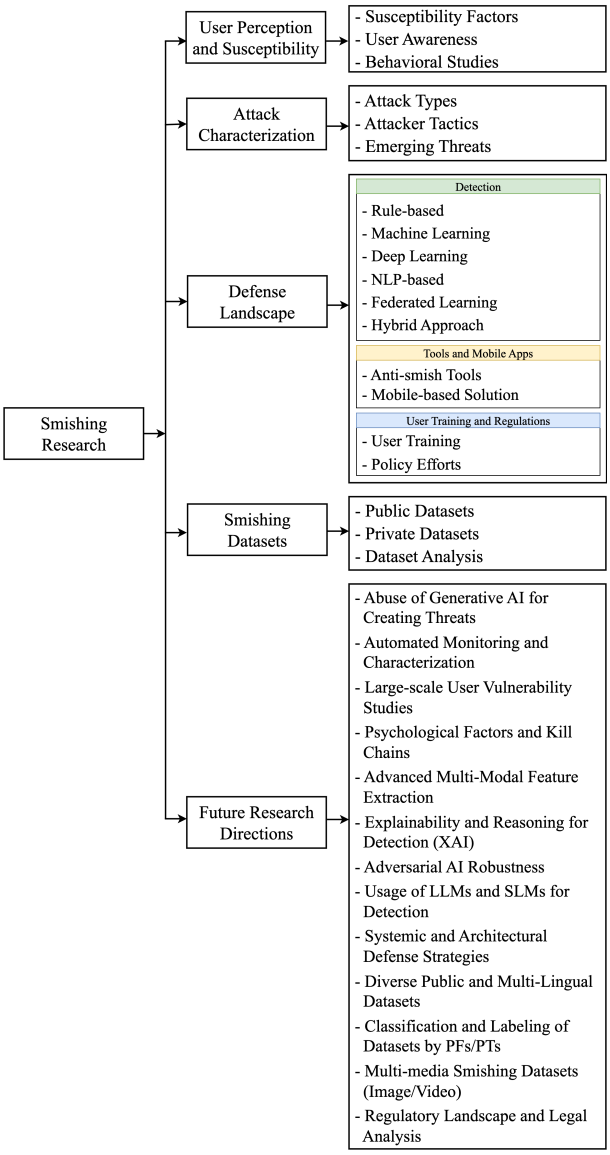


Fig. 1: Scope of the present study.

ing") AND ("detection" OR "defense" OR "dataset" OR "characterization" OR "user susceptibility").

We applied the following Inclusion Criteria (IC) and Exclusion Criteria (EC) to filter the initial search results:

- **IC1 (Relevance):** The study must primarily address SMS-based phishing (smishing). Studies focusing on general phishing or social engineering were included only if they provided essential theoretical frameworks (e.g., psychological factors) that are directly transferable to the smishing domain.
- **IC2 (Language):** Only articles published in English were included.
- **IC3 (Source Type):** Peer-reviewed conference proceedings, journal articles, and reputable technical reports were included.
- **EC1 (Duplicates):** Duplicate entries across databases were removed.
- **EC2 (Scope):** Articles that discussed "mobile security" broadly without a specific focus on social engineering via SMS were excluded.

This process has initially yielded a large set of publications. To manage the volume of results, we limited the screening to the first 10 pages of search results (sorted by relevance) for each keyword combination to ensure the most significant contributions were captured. To ensure the quality and relevance of the selected papers, we deployed a multi-step screening protocol where papers initially were screened based on their titles and abstracts to filter out irrelevant works. Subsequently, we reviewed the full text of these articles to ensure the alignment of these papers with the four research pillars. We conducted an expert review (by *five* coauthors) to verify the selected papers and reduce bias. Any disagreements regarding inclusion or categorization were resolved by unanimous decision among the research group. Finally, a total of 117 papers have been retained after a careful manual screening for relevance based on the defined scope for this paper.

In addition, to examine existing regulatory measures and law enforcement efforts addressing smishing, we conducted a targeted *Grey Literature* search separated from our academic literature review. We queried official government databases (e.g., Federal Register, EUR-Lex) and legal repositories using specific policy-related keywords such as *GDPR*, *TCPA*, *Telecommunications Act*, *smishing penalties*, *caller ID spoofing regulations*. This search targeted documents from major regulatory providers, including the U.S. Federal Communications Commission (FCC), the Federal Trade Commission (FTC), and the European Commission. This search resulted in the identification of 20 legal documents and 18 online reports, which were subsequently analyzed to extract insights into current regulatory frameworks and policy initiatives. Moreover, we have used 2 webpages or APIs (Application Programming Interfaces) and 3 online repositories containing SMS datasets along with 12 other datasets having their individual papers. We also identified 10 applications and anti-smishing tools within the scope of this review. Figure 2 presents a bar graph illustrating the yearly distribution of papers and other available materials (e.g., reports, repositories, apps, legal documents) under our defined categories. Figure 2 shows a significant increase in smishing research starting in 2020, likely due to more mobile communication during the COVID-19 pandemic. Publications on the *Defense Landscape* spiked in 2023 and 2024, in such it covers over 40% of the total paper. This reflects the community's response to new AI-driven threats and the push for more accurate detection tools. It is important to note that some papers and materials are considered under multiple categories (i.e., a paper can be both in *User Perception and Susceptibility* category and *Attack Characterization* category) based on their contents as shown in Figure 3. This figure highlights the interdisciplinary nature of recent research. The large overlap between *Attack Characterization* and *Defense Landscape* suggests that modern defenses are increasingly built on insights into attack tactics rather than in isolation. In contrast, the minimal overlap between *User Training and Regulation* and Technical Defenses (*Defense Landscape (Detection)* and *Defense Landscape (Tools & Mobile Apps)*) reveals a gap between policy and technical implementation. It also shows that a single study often contributes to multiple areas due to the interdisciplinary nature of smishing research. For example, papers that categorize attack characterization often propose detection methods and place them in both *Attack Characterization* and *Defense Landscape*. To capture this, we used a multi-label tagging approach. We assigned papers to multiple categories based on their substantial contributions to pillars, instead of limiting them to a single primary label. It is important to note that, no single paper has contributed to more than 2 pillar categories.

2.1 Methodology

Our methodology is driven by Research Questions (RQs) in association with the user perceptions and susceptibility, smishing attack characterization, drawing smishing defense landscape, and findings insights from smishing datasets. This systematization paves the way for proposing future research directions towards effectively mitigating smishing threats.

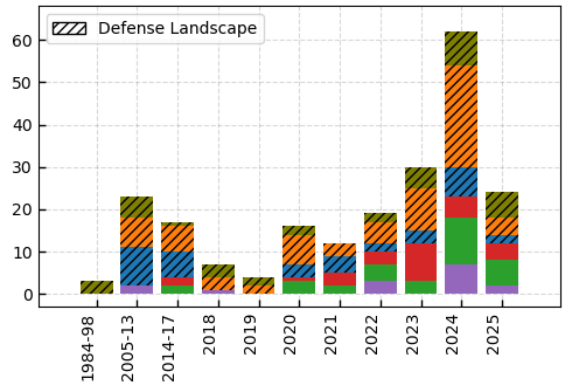


Fig. 2: Selected 117 papers, 20 legal documents, 18 online reports, 10 applications, 3 repositories, and 2 APIs with specific categories and publication years (until December 2025), some items fall under multiple paper categories;

User Perception & Susceptibility (26 papers, 1 report) Attack Characterization (23 papers, 8 reports) Defense Landscape (Detection) (Detection 7 Apps, 65 paper, 1 report) Defense Landscape (Tools & Mobile Apps) (10 Apps, 23 papers, 2 Web/API) Defense Landscape (User Training and Regulation) (23 legal documents, 2 papers, 13 reports) Datasets (12 papers, 3 repositories)

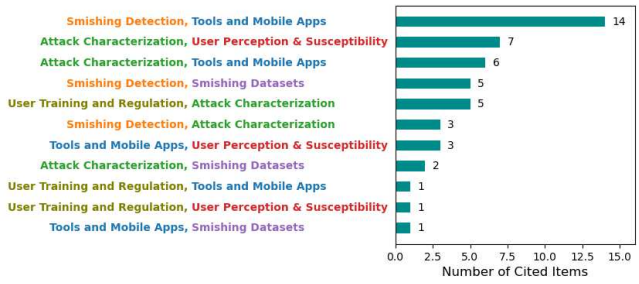


Fig. 3: Count of overlapping papers and materials between different categories

2.1.1 RQs Related to User Perception and Susceptibility

Understanding user perceptions and susceptibility are important to comprehend smishing threats. Hence, we propose the following three RQs to understand user's perception and susceptibility when exposed to smishing attacks based on existing user-studies:

- **RQ-B1:** How much susceptible are users of different backgrounds when encounter smishing messages?
- **RQ-B2:** Does the literature address a diverse user groups when conducting behavioral or quantitative survey studies?
- **RQ-B3:** What are cues used by users to manually identify smishing messages?

2.1.2 RQs Related to Smishing Attack Characterization

Attack characterization is important to completely understand smishing attack classifications, types, themes, and psychological aspects (i.e., techniques, factors). These key findings are essential for creating better user training materials, effective user notification system, and robust smishing detection and defense mechanisms. Thus, we propose the following two RQs:

- **RQ-A1:** What are the current known threat classification and attack types for smishing threats?
- **RQ-A2:** What are the various themes, topics, psychological factors (PFs), and psychological techniques (PTs) in the context of smishing threats?

2.1.3 RQs Related to Smishing Defense Landscape

After understanding the smishing threat landscape, we systematize the existing smishing defense landscape, covering both advanced detection methodologies, practical app and tool based defense solutions, user trainings, and policy-based regulatory efforts. We propose the following eight RQs:

- **RQ-C1:** What types of detection approaches and models are proposed?
- **RQ-C2:** What are the key indicators (or features) use in existing detection methods?
- **RQ-C3:** What are the strengths and weaknesses of existing detection mechanisms?
- **RQ-C4:** Are explainability and transparency considered to enhance trust within existing detection models?
- **RQ-C5:** What is the extent of current studies focusing on the defense capabilities and anti-smish tools?
- **RQ-C6:** How do mobile-based solutions, such as mobile applications and user alert/notification system contribute to preventing smishing threats?
- **RQ-C7:** How effective are user-training and use awareness modules to defend against smishing attacks?
- **RQ-C8:** Are there any existing laws, legal frameworks, and effective regulatory policies around the globe to mitigate smishing threats and what contributions do they have in certain geographical regions?

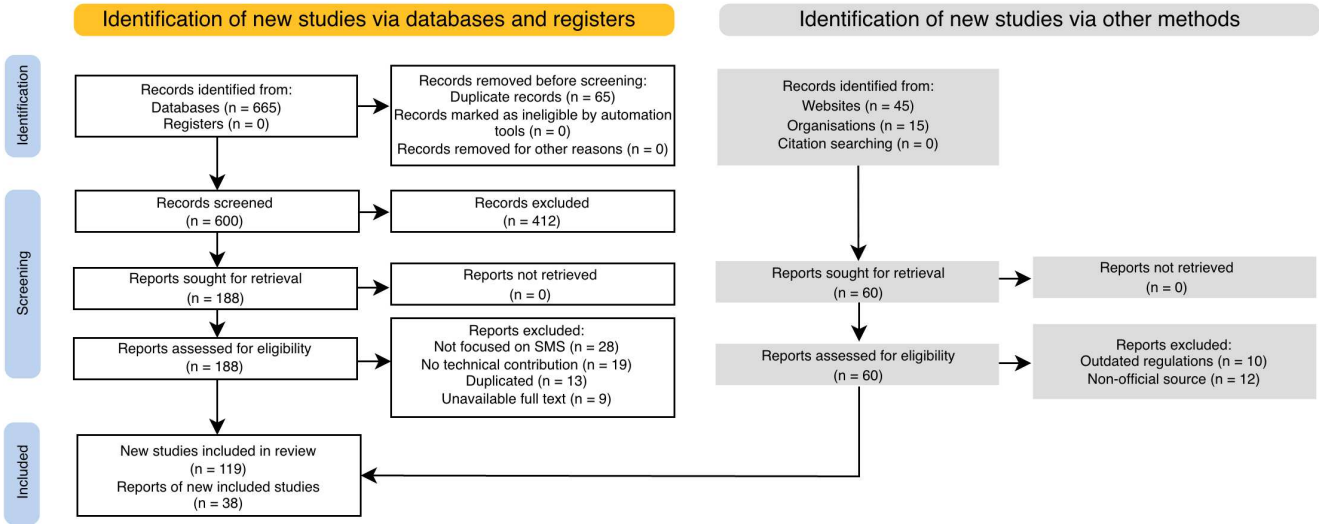


Fig. 4: PRISMA flowchart for a comprehensive review of SMS phishing literature

2.1.4 RQs Related to Investigate Smishing Datasets

The availability of SMS datasets is crucial specifically for understanding current attack tactics as well as machine learning (ML) and AI-driven detection research. We propose the following three RQs to assess existing smishing datasets:

- **RQ-D1:** What are the publicly available datasets for SMS phishing that are leveraged in current research?
- **RQ-D2:** What are the characteristics (e.g., number of instances, message length, containing URLs or not, live websites, etc.) and coverages (e.g., ham, spam, smish) of each of these smishing datasets?
- **RQ-D3:** How these datasets are leveraged in existing literature and what types of detection models are adopted using such datasets?

2.1.5 RQs Related to Future Research

As one of the primary goal of this paper is to guide future research efforts to collectively defend against smishing threats and create a more secure mobile messaging ecosystem for the end-users, we propose the following five RQs in terms of future directions:

- **RQ-E1:** What are the potential new promising avenues of research in attack characterization and landscape of smishing attacks?
- **RQ-E2:** What can we do further to continue assess diverse users' susceptibility and behaviors to understand the needs for tailored defense?
- **RQ-E3:** What are the future of detection and defense landscapes that can safeguard users from smishing attacks?
- **RQ-E4:** What type of large-scale and diverse datasets can further help improve the defense posture in fighting this threat?
- **RQ-E5:** What can we do further to understand and improve current regulatory and policy efforts including law enforcements actions in fighting smishing threats?

3 User Studies On Smishing Susceptibility and Psychological Awareness of Users

3.1 Users' Susceptibility To Smishing Attacks

User behavior and perception are important role in smishing attacks. A lack of knowledge in mobile security makes users easy targets [174], while "security expertise bias" can reduce users' ability to correctly distinguish legitimate and phishing messages. Users also show vulnerabilities when they believe they are account holders of the entity mentioned in the message, which leads them to disregard fraudulent cues [35]. On the other hand, involvement with smishing is sometimes only for curiosity, even past experiences may not influence falling for it again, as the two-round setup of experiment results shown in [104]. Users' routines also influence their susceptibility. Frequent use of social media or banking apps makes related messages appear more trustworthy. For example, Users spending more time on social media are more vulnerable to social media-related smishing [65]. In addition, users often respond to prompts or enter credentials without careful consideration [81]. These findings indicate that susceptibility

is influenced not only by awareness, but also by familiarity, habits, and context.

3.2 Visual Cues To Detect Smishing

Visual cues are elements designed by attackers to manipulate users, but they also help users to detect suspicious messages [29]. These cues can be broadly categorized into metadata-related and content-related features:

3.2.1 Meta-Data-based Cues

Metadata-related cues include *Fake URLs* [152], *inconsistencies in URLs* [29], *Shortened URLs* [29], *Suspicious URLs* [85], and *Spoofed area code* [104] that create a false sense of familiarity. Such inconsistencies are often associated with fraudulent messages and can help users detect phishing attempts.

3.2.2 Content-based Cues

Content-related cues include *Messages containing forms in URLs* and *suspicious links to downloads APK file* [81], *Messages asking for personal information* [57], and the use of call-to-action messages that encourage immediate interaction [152]. Attackers also frequently rely on urgency-related language [57], suspicious keywords [57, 58, 46], poor grammar [29], and special characters to bypass detection systems and influence user behavior [85]. On the other hand, *Entity identification* can serve as both a useful and misleading cue. While recognizing a known organization or brand can help users verify legitimate messages [152], attackers often exploit this by impersonating trusted entities to gain user confidence and increase the likelihood of interaction [104].

3.3 Heterogeneity of Participants in Existing Studies

Here, we discuss various existing user studies that examine how demographic factors impact their susceptibility. While these studies provide useful insights, their findings are often mixed and highlight the complexity of user behavior. Detailed study characteristics and participant distributions are summarized in Table 2 and Table 3.

3.3.1 Ethnic Diversity and Impacts

Existing studies show that ethnicity impact how users perceive and react to smishing messages[104, 151]. Some studies suggest that certain groups tend to be more cautious and demonstrate higher detection ability [151], while others report increased vulnerability in more realistic attack scenarios[104]. These inconsistencies shows that susceptibility varies across contexts and experimental settings. As a result, awareness strategies may need to be adapted to different communities rather than assuming uniform behavior across populations.

3.3.2 Age-related Patterns and Impacts

We have also explored how age affects users' responses to smishing attacks as presented in existing literature. Younger users generally exhibit more cautious behavior and lower interaction rates [151], whereas older users tend to report higher comfort levels with messages and may be more likely to engage in real-world scenarios [104]. The above findings indicate that while age does not significantly alter detection accuracy for distinguishing real SMS from smishing in lab settings, it does influence real-world behavioral responses.

3.3.3 Gender Impacts

Gender-based analysis shows no significant difference on susceptibility[35]. While prior findings reveal females had a slightly higher susceptibility [104], the overall findings suggest that both male and female users are similarly vulnerable to smishing attacks.

3.3.4 Educational Background Related Impacts

Educational background also presents mixed results. Some studies [104] indicate that higher education levels do not necessarily reduce susceptibility and may even increase risk in certain scenarios. Others suggest that experience and familiarity over time can improve resilience [176, 35]. These findings imply that education alone is not a reliable predictor of user behavior in smishing contexts.

4 Characterization of Smishing Threats and Attack Landscapes

In this research pillar, we systematically categorize the smishing threats and introduce various classification schemes based on contents, senders, target-audience, attacker's intent, attack campaigns, attack themes and PTs. Smishing threats or attacks can be classified in several ways based on diverse factors. Figure 5 illustrates our defined taxonomy of SMS phishing threats, which includes content-based, sender information-based, target audience-based, intent-based, message delivery

Table 2: Overview of User-Study-Driven Smishing Research

Reference	Key Finding	User Count	Study Type
[91]	EEG and eye-tracking show users often miss phishing cues but attend to malware warnings.	23	Experimental user study
[29]	Examines how users distinguish legitimate SMS messages from smishing messages.	576	Survey-based quantitative
[16]	Evaluates user vulnerability to SMS phishing attacks.	24	Experimental
[12]	Analyzes phishing susceptibility across groups from European countries with different educational backgrounds using fuzzy clustering.	N/A	Quantitative analysis
[104]	Studies message and participant factors associated with smishing success.	256	Experimental
[32]	Studying user behaviors and susceptibility to smishing attacks when exposed in controlled settings	28	Survey-based Behavioral
[175]	Investigating how demographic factors and personality traits impact phishing susceptibility to selected university undergraduates in Nigeria	480	Quantitative Experimental
[35]	Investigating demographic and contextual factors impacting the vulnerability of user participants to smishing attacks	1,007	Survey-based Demographic Analysis

Table 3: Participant Diversity in User-Study-Driven Smishing Research

Reference	Ethnic/Racial Diversity	Age Diversity	Education Level	Gender Diversity
[91]	N/A	19–22 (44%); 23–26 (20%); 27–30 (16%); 31–34 (8%); 35+ (12%)	Students (72%); working professionals (16%); others (12%)	Male (64%); Female (36%)
[29]	N/A	18–21 (40.5%); 22–25 (27.9%); 26–30 (18%); 31–40 (9.2%); 41+ (4.4%)	N/A	Male (50%); Female (45%); Others (5%)
[16]	N/A	N/A	N/A	N/A
[12]	European countries	N/A	C1 (6 countries) - Low education: 28.33%, Medium education: 40.00%, High education: 53.83%; C2 (18 countries) - Low education: 7.11%, Medium education: 10.17%, High education: 16.06%; C3 (11 countries) - Low education: 20.45%, Medium education: 30.45%, High education: 43.45%	N/A
[104]	White (48.30%); African American (17.35%); Asian American (10.57%); Mexican American (8.67%); American Indian (8.30%); Other Hispanic/Latino (4.90%); Native Hawaiian (1.13%); Other (0.75%)	18–24 (18.11%); 25–34 (54.71%); 35–44 (22.64%); 45+ (4.5%)	<High School (1.88%); High School (9.81%); Associate (25.28%); Some College (18.49%); Bachelor's (32.45%); Master's (9.05%); Doctoral (3.01%)	Male (53.58%); Female (46.42%)
[32]	N/A	18 - 21	High School/GED (53.6%); Associate (3.6%); Some College (42.9%)	Male (32.1%); Female (67.9%)
[175]	Nigerian	Under 21 (21.5%); 21 - 25 (34.5%); 26+ (44.0%)	1Y (36.0%); 2Y (19.8%); 3Y (19.5%); 4Y and above (25.7%)	Male (50.9%); Female (49.1%)
[35]	White (83.4%); Black/African (10.5%); Asian (2.38%); Native American/ Alaska Native (0.9%); Self-Described (1.5%); Prefer not to say (1.3%)	18-24 (23.0%); 25-34 (6.95%); 35-54 (31.0%); 55-64 (17.2%); 65+ (21.7%)	No school (53.37%); In School (No 4Y deg.) (12.82%); 4Y deg. (23.06%); Doctorate (10.73%)	Male (62.6%); Female (36.0%); Nonbinary (1.0%); Self-described (0.2%); Prefer not to say (0.2%)

technique-based, message topic/theme-based, campaign-size based, and user psychology-based classification schemes. This taxonomy provides a structured view of the smishing threat landscape.

4.1 Content-Based Classification

In terms of content diversity, we see smishing messages attached with *malicious URLs* [58], *malicious files* (linked to an *.apk file) [82], *attached image (MMS)*[142], or *QR codes*[101], which are all designed to redirect users to phishing sites or deliver downloadable malware.

4.2 Sender Information-Based Classification

We can also classify smishing threats in terms of sender information such as- *spoofed numbers*[155] that are used imitating legitimate organizations, *short codes* [53] used in SMS marketing, *random numbers* [52] often broadcast to non-targeted audience, and *email addresses* created algorithmically for mimicking a legitimate entity.

4.3 Target Audience-Based Classification

Considering the target audience, smishing attacks can be categorized as *targeted spear-smishing* [52] where specific individuals are targeted based on previous reconnaissance activities and *mass-smishing* where a message is sent to many unknown targets as a message broadcast for opportunistic attacks.

4.4 Intent-Based Classification

Based on the intent (or goal) of the attackers, in this paper, we define the following intent-based smishing classes: *stealing and harvesting information*, *recon for information gathering*, *benefiting financially*, and *conducting identity theft*. These intents may occur in multiple stages, where an initial message is used to identify active victims, followed by subsequent messages designed to extract sensitive information.

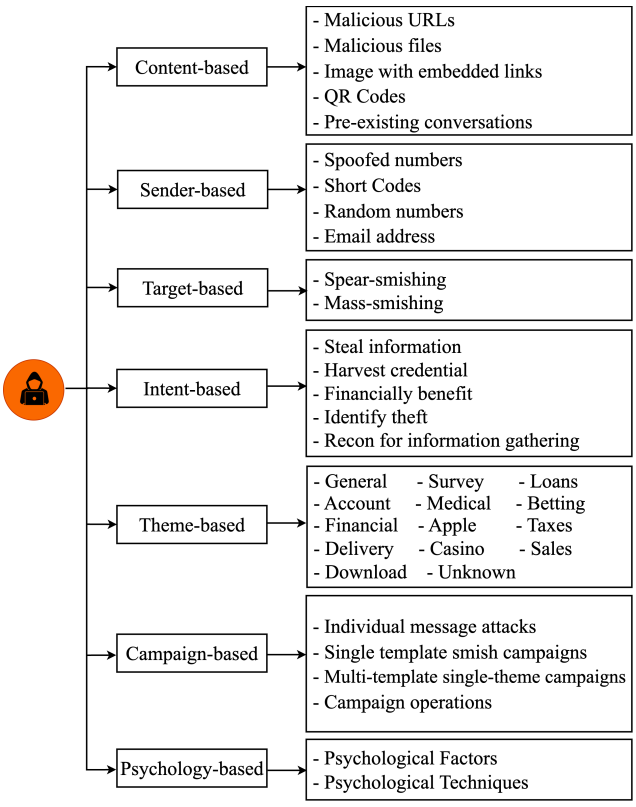


Fig. 5: Taxonomy of SMS phishing threat landscape

4.5 Campaign-Based Classification

Liu et al.'s work [67] on SMS spearphishing also provides valuable insights into smishing campaigns. Their work has identified a total of 11,475 unique campaigns within their dataset. Their campaign-level analysis revealed that a small number of campaigns often dominate large portions of attack activity, indicating structured and repeated attack patterns.

Recent studies, such as *SmishViz* [120] introduced a graph-based visualization approach that enables defenders to monitor and characterize ongoing smishing campaigns and campaign operations by clustering similar messages and connecting them with common web infrastructures used in the message contents as connected tripartite graphs. Moreover, the *SmishViz* system can reveal how large-scale smishing activities can be mapped across topical themes, message templates, and web domains to uncover broader coordinated attack operations. Thus, by considering the structure and reachability of the smishing campaigns introduced in Sanjari et al. [120], we can refer to the following campaign-based smishing classification: (i) individual message attacks, which are one-time smishing attempts that are unique and do not connect to any known themes; (ii) single template smishing campaigns, where attackers reuse the same message template across multiple targets and easier to detect using pattern recognition techniques such as the Ratcliff pattern recognition algorithm [106]; (iii) multi-template single theme campaigns, where multiple templates of similar themes (e.g., IRS tax refund, USPS delivery) send to diverse target population; and (iv) campaign operations, which connect multiple templates or multi-template campaigns of different themes using common shared infrastructures such as domains or URLs to indicate coordinated large-scale attack activities.

4.6 Theme-Based Classification

The theme of a smishing message plays a key role in influencing user interaction. From existing literature [88], common themes include *account-theme*, *financial-theme*, *delivery-theme*, *casino-theme*, and other context-specific scenarios. However, themes can evolve with new events around the globe and be leveraged by attackers. For example, COVID-19 has emerged as a popular theme to carry out cyberattacks [2, 3, 98] in early 2020 while Russia-Ukraine war became a popular theme in the spring of 2022 [79]. In summary, other than regular attack vectors, we may observe more smishing attacks during a global or regional events to increase the attack success rates.

4.7 User Psychology-Based Classification

Alike other SE attacks, smishing attacks also exploit user psychology to succeed [133]. To get a better understanding of user psychology for smishing, it is essential to understand two core concepts in the context of smishing attacks: 1) Psychological Factors (PFs) and 2) Psychological Techniques (PTs) [69].

4.7.1 Psychological Factors (PFs)

In the context of smishing attacks, PFs work as human characteristics and vulnerabilities that attackers can exploit. PFs can be identified into 5 major categories of factors, including– [i] social psychological factors (e.g., authority, social proof, reciprocation), [ii] personality and individual differences factors (e.g., curiosity, impulsivity, agreeableness, self-control), [iii] cognitive factors (e.g., reliance on mental shortcuts such as cognitive miser behavior), [iv] emotional factors (e.g., fear, greed, sympathy, loneliness), and [v] workplace factors (e.g., workload, stress, habituation). Attackers often target multiple PFs simultaneously. For example, authority and trust can be exploited by attackers through the impersonation of legitimate institutions, while urgency and fear can be leveraged to trigger rapid, unverified responses. Similarly, cognitive biases such as reliance on shortcuts can reduce users' ability to critically evaluate suspicious messages.
Overall, PFs provide a useful lens for understanding why users fall victim to smishing attacks and show that susceptibility is not only technical but also deeply behavioral and contextual.

4.7.2 Psychological Techniques (PTs):

Psychological techniques (PTs) are strategies used by attackers to design an attack (i.e., smishing message). Based on prior work [69], we identify the following PTs applicable to smishing scenarios.
[i] Personalization: refers to the customization of messages with personal information to build trust.
[ii] Impersonation: meaning impersonating another identity, such as a legitimate organization, to increase the probability of the victim's compliance.
[iii] Incentive & motivator: refers to offering external rewards or internal satisfaction to encourage a desired behavior.
[iv] Urgency: refers to creating a situation that needs the user's immediate action, with the probability of reducing the chance of recognizing an attack.
[v] Decoy effect: refers to making users believe they are receiving a good deal.
[vi] Loss aversion: means providing a free good or service to build the victim's trust and attachment and later charging an exorbitant amount once the victim is attached to it. It can also involve influencing a victim's decisions by gradually providing them with seemingly legitimate information over time before launching the attack.
[vii] Priming: refers to gradually influencing an individual's subsequent decisions through subtle manipulation.
[viii] Persuasion: is a technique that encourages a particular behavior by exploiting various psychological factors.
[ix] Contextualization: posing as a member of the victim's group to create a sense of shared affiliation.
[x] Pretexting: refers to creating a potentially possible scenario to increase the victim's engagement with the attacker.
Each PT is associated with one or more PFs. For instance, impersonation exploits authority and trust, while incentive-based messages target greed, and emotional responses (e.g., empathy, loneliness).

In summary, the combination of PFs and PTs shows that smishing attacks are not just technical, but strongly influenced by human psychological (cognitive) vulnerabilities [125]. Attackers carefully design messages to trigger quick and emotional responses. This perspective also has several defense implications (i) Detection systems should consider behavioral patterns, not just text features, (ii) User awareness should focus on how messages manipulate people, and (iii) Explainable or reasoning AI models can benefit from linking predictions to PF/PT patterns.

4.8 Summary of Smishing Threat Landscape

The smishing threat landscape presented in Figure 5 is intrinsically multi-dimensional, where a single attack instance can simultaneously occupy multiple categories. For instance, a "Failed Package Delivery" attack may be primarily classified under a *theme-based* scheme, *user psychology-based* (by inducing urgency), *content-based* (via malicious URL embedding), and *intent-based*(targeting credential theft) schemes.

Consequently, the categorization of a smishing campaign is not mutually exclusive but rather multidimensional. A study focusing on "Executive Impersonation" may prioritize the target audience (spear smishing), yet it must also be viewed through the lens of sender information (identity spoofing) and message delivery techniques (personalized social engineering). This overlapping taxonomy ensures that the threat landscape

remains comprehensive, allowing researchers to pivot their analysis between different principal categories while acknowledging the interconnected nature of modern smishing schemes.

5 Evaluating Smishing Defense Landscape

5.1 Detection-based Defense Measures

In this section, we synthesize articles on various smishing detection mechanisms using rule-based, ML-based, DL-based, NLP-based, Federated Learning-based, and hybrid approaches.

5.1.1 Rule-based Detection

Rule-based smishing detection represents a traditional approach that relies on predefined rules to identify suspicious messages [4]. These rules target common smishing characteristics, such as specific keywords associated with phishing scams (e.g., 'urgent' and 'account locked'), suspicious URL patterns, and spoofed sender information. While this method offers a simple and efficient way to detect smishing attempts, its effectiveness can be challenged by evasive tactics and newer attacks.

In literature, Akande et al. [4] combine URL blacklist checks, scam number matching, spam keyword thresholds, symbol patterns, and message style cues (emoticons, abbreviations, self-answer prompts) using a *RIPPER/C4.5* rule-learning pipeline. Their approach detects smishing based on both URL and keyword features, along with additional structural and sender-based rules. However, like other rule-based systems, its adaptability depends on how well the predefined rules capture evolving attack patterns.

5.1.2 Machine Learning (ML) based Detection Approaches

Beyond rule-based methods, ML-based approaches offer a more sophisticated and diverse smishing detection. Table 4 provides a summary of methods and techniques employed in recent literature. In this approach, researchers utilize established algorithms such as Neural Networks (NN) [47, 46, 56, 83, 126, 134, 122, 10, 54, 111, 76], Support Vector Machines (SVMs) [138, 158, 58, 61], Naive Bayes [49, 17, 92, 82, 58], Decision Trees [138, 158, 58, 4, 61], and Random Forests [138, 17, 158, 58] trained on features extracted from message contents, such as keywords, part-of-speech patterns, or sentiment analysis. Some of the important and relevant features from existing literature are highlighted below. We also provide a comparative overview of key works on smishing detection, and highlight their strengths and limitations in Table 5. Clearly, the winners are the language models in terms of accuracy and detection rate as depicted in this table.

Indicators of Detection. A combination of URL-based and/or text-based features serving as indicators of smishing attacks.

(i) Content Features: Oftentimes, the actual content of SMS messages can be used as features. ML algorithms can learn from large datasets of labeled messages (smishing vs. legitimate) where texts are tokenized or grouped as n-grams. The tokenized text is then transformed into numerical representations by applying methods such as bag of words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF), or word to vector (Word2Vec). In addition, the researchers use heuristic-based features such as message length, number of misspelled words, count of special characters, and binary indicators for the presence of emails, URLs, or phone numbers [4, 138].

(ii) URL Features: While analyzing message content is crucial, smishing attacks often rely on malicious URLs to steal user information or deploy malware. In some cases, the message may consist only of a URL. Features extracted from URLs generally fall into two categories:

- ***URL lexical information:*** Lexical information is collected directly from the URL strings and is easily available. Examples include unusual character patterns, excessive length, presence of subdomains, shortened URLs, and a mismatch between domain names and message context.
- ***URL host and web content information:*** These features require querying external sources and include domain reputation, registration details, age, redirection behavior, HTTPS usage, and webpage characteristics[61, 77]. Additional checks using blacklist repositories such as VirusTotal [139] or PhishTank [27] can further support detection [46, 154].

However, the use of traditional content and URL features is not effective against modern and evolving attack models. For example, in conversational scams, such as the *'Hi Mum and Dad'* scam [1], the initial message often bypasses classic cues by avoiding malicious URLs and relying on psychological techniques to move the interaction to another platform. The actual fraud occurs in later stages, highlighting the limitations of feature-based detection at the entry point.

5.1.3 Deep Learning-based Detection Approaches

Deep Learning techniques have also shown promising results [46, 56, 83] in content and URL-based analysis

Table 4: Summary Analysis of Existing Smishing Detection Using Various ML and DL Methods

Reference	Method	Content Analysis		URL Analysis					NLP Techniques			
		ML	DL	APK	SAL	SHRT	BLK	TAG	TN	TKN	WE	TF-IDF
[49]	Naïve Bayes Classifier	✓	.	✓	✓	.	✓	.	✓	✓	.	.
[81]	content-based analysis + URL behavior inspection	✓	.	✓	✓	.	.	✓	✓	✓	.	✓
[47]	CNN-LSTM	.	✓	.	.	.	.	.	✓	✓	✓	✓
[138]	AdaBoost, Random Forest, Decision Tree, and SVM	✓	.	.	.	.	.	.	.	✓	.	.
[82]	Naïve Bayes Classifier	✓	.	✓	✓	✓	✓	✓	✓	✓	.	✓
[17]	Naïve Bayes and Random Forest classifiers	✓	.	.	.	.	.	.	.	✓	.	✓
[46]	Deep Learning (DL) + BERT	.	✓	.	.	.	✓	.	✓	✓	✓	.
[92]	Naïve Bayes Classifier	✓	.	.	.	.	.	.	✓	.	.	.
[158]	SVM, Random Forest, AdaBoost, XGBoost, CART	✓	.	.	.	.	.	.	✓	✓	.	✓
[56]	CNN-LSTM	.	✓	.	.	.	.	.	✓	✓	✓	✓
[58]	XGB, GBDT, RF, BgC, KNN, ETC, DT, LR, AdaBoost, BNB, MNB, SVC, GNB	✓	.	.	.	.	.	.	✓	✓	.	✓
[83]	ANN	.	✓	✓	✓	✓	✓	✓	✓	✓	.	.
[4]	Rule-based + RIPPER + PART C4.5	✓	.	.	✓	.	✓	.	✓	✓	.	.
[61]	Domain Checker (Decision Tree) + NLP Checker(SVC with TF-IDF)	✓	.	.	.	✓	✓	✓	✓	✓	.	✓
[85]	Domain Checking Phase + SMS Classification Phase	.	✓	✓	✓	✓	.	✓	✓	✓	.	.
[126]	FL + LSTM, multi-lingual support	.	✓	.	.	.	.	.	.	.	✓	.
[153]	Comparative study	.	.	.	.	✓	✓	.	.	.	.	.
[89]	Large-scale empirical study	.	.	.	.	✓	✓	.	✓	.	.	.
[134]	FL + DistilBERT	.	✓	.	.	.	.	.	✓	✓	✓	.
[60]	Image-based URL extraction + List-based URL matching	.	.	.	.	.	✓	.	.	.	.	.
[122]	Char-CNN	.	✓	✓	.	.	.	.	✓	✓	✓	.
[10]	FL with Non-IID setup + PhoBERT	.	✓	.	.	.	.	.	.	✓	.	.
[54]	CNN + PCC-PCA	.	✓	.	.	.	.	.	✓	✓	.	✓
[111]	Iterative FL setup + ConvLSTM	.	✓	.	.	.	.	.	.	.	✓	.
[76]	CNN-LSTM	.	✓	✓	.	.	✓	.	✓	✓	✓	✓
[131]	unsupervised learning(K-means, NMF, GMM, PCA) + deep semi-supervised learning (RNN-Flatten, LSTM, Bi-LSTM)	.	✓	.	.	.	.	.	✓	✓	✓	✓
[130]	Persuasion-based prompt learning + transformer-based classification	.	✓	.	.	.	.	.	.	.	.	.
[157]	Fine-tuned RoBERTa + LIME	.	✓	.	.	.	.	.	✓	✓	✓	.

Keys: ML: Machine Learning, DL: Deep Learning, FL✓: Federated Learning, APK: Application Package Analysis(containing download link), SAL: Self-Answering Link, SHRT: URL Shortening Detection, BLK: Blacklist, TAG: HTML Form Tagging, TN: Text Normalization, TKN: Tokenization, WE: Word Embeddings, TF-IDF: Term Frequency-Inverse Document Frequency.

XGB: Extreme Gradient Boosting, GBDT: Gradient Boosted Decision Tree, RF: Random Forest, BgC: Bagging Classifier, KNN: K-Nearest Neighbors, ETC: Extra Trees Classifier, DT: Decision Tree, LR: Logistic Regression, AdaBoost: Adaptive Boosting, BNB: Bernoulli Naïve Bayes, MNB: Multinomial Naïve Bayes, SVC: Support Vector Classifier, GNB: Gaussian Naïve Bayes, ANN: Artificial Neural Network, PCC: Pearson Correlation Coefficient PCA: Principal Component Analysis, LSTM: Long Short-Term Memory, NMF: Non-Negative Matrix Factorization, GMM: Gaussian Mixture Models.

for detecting smishing attempts. For example, Mishra et al. [85] used a leet words check, which indicates if any letters are replaced with proximity numbers and keyword-based indicators within a backpropagation-based model. Convolutional neural networks (CNNs) have been explored for extracting features from SMS text [47], while Long Short-Term Memory (LSTM) networks are widely used due to model sequential dependencies and capture contextual information in messages [56]. In addition, autoencoder-based models have been investigated to learn compact representations of SMS messages and detect anomalies by identifying normal SMS patterns and identifying deviations from normal patterns [76]. However, while these models generally achieve higher detection performance by automatically learning relevant features, their increased complexity reduces transparency and explainability, which remains a challenge for real-world deployment.

5.1.4 Natural Language Processing-based Detection Approaches

Different strategies that integrate NLP techniques with ML algorithms for detecting smishing and fraudulent

Table 5: Comparative Analysis of Smishing Detection Literature (Categorized by Rule-based, ML, DL, and LM approaches)

Reference	Key Idea	Strengths	Weaknesses
Rule-based and Heuristic Approaches			
[49]	multi-step logics, Normalizes SMS, applies NB classifier	Multi-feature coverage (URLs, APKs, sender, text cues)	No performance metrics
[81]	Combines text and URL for filtering rules with OneVsRestClassifier	Hybrid and comprehensive detection approach	Relies on keywords, Imbalanced dataset, No performance metrics
[92]	User notification and content filtering with NB classifier	Accuracy of 94%, real-time user notification	Reliance on keyword filtering, lacks comparison metrics
[4]	Rule-based detection using PART (C4.5) model, content filtering and user alerts	High accuracy (98.42%), real-time performance	Relies on pre-defined rules, limited adaptability to novel smishing tactics
[153]	Evaluation of commercial anti-smishing tools, comprehensive benchmark of anti-smishing tools	Characterizing smishing attack through a qualitative analysis	Infeasible full testing, time-limited experiments (3 days)
[89]	Content and infrastructure-based clustering	Large public dataset, evaluation through blacklists	Potential bias in the Data source due to the use of public gateways
[60]	Cross-device analysis for smishing defense using augmented reality (AR) glasses	Accuracy of 85.19%, high System Usability Scale score (74.4)	Simple URL blacklist matching
Machine Learning (ML) Approaches			
[82]	4-part detector for text, URL, code, apk with NB classifier	Covers diverse attacks, 96.3% accuracy	No malware analysis of downloaded APK
[138]	Ranks features for ensemble models (AdaBoost)	High accuracy (98.4%), feature focus	Lacks feature/data details
[17]	Detects via term importance and complaints using RF	98.15% accuracy, adapts to vishing/phishing	Low recall, many false negatives
[158]	Advanced feature selection via ANOVA test, SVM model for detection	98.39% accuracy	Weak adaptability to language evolution
[58]	Text + URL spam detection using rare words with voting (RF, KNN, ETC)	99% accuracy, clean dataset	Ignores URL-less SMSes
[61]	Hybrid domain and NLP feature detection with RF	99.4% accuracy, F1-score ¿ 99% tested on publicly avaiable data	No independent real-time system, higher processing time complexity
Deep Learning (DL) Approaches			
[47]	Bilingual spam detection with CNN-LSTM	Handles Arabic/English, 98.4% accuracy	Limited feature explanation
[56]	CNN-LSTM for spam detection with embeddings	98.4% accuracy, 98% F1-score	Low recall (89%) for minority spam class
[83]	Feature-focused ANN Backpropagation for smishing	97.4% accuracy, domain + message dual checking	Potential bias for Paytm themed URLs
[85]	URL-based detection using authenticity with MLP model	97.93% accuracy, consistent results	Limited features, narrow comparison
[54]	Time-efficient feature extraction with Pearson correlation coefficient and Principal Component Analysis (PCA) and CNN model	Competitive accuracy rate (99.8%) with robust model setup	Intricacy to re-generate features for new sms
[122]	Model resilient to text evasion	On-device, 99% accuracy on 1.2M messages	EVA tool generalization unverified
2024[131]	Optical Character Recognition + hybrid model for detecting spam from screenshots	94.13% Accuracy, user-adaptive	Limited scalability, language constraints
[111]	Federated ConvLSTM model	99.19% Accuracy, Scalable and adaptable for other phishing scenarios	Low interpretability, domain-limited
Language Model (LM) Approaches			
[46]	App with multilingual smishing detection BERT-FC model	99.63% accuracy, cross-language support	Lack of mobile implementation details causes privacy concerns
[134]	Protecting users' privacy using an on-device detection method with DistilBERT model	Privacy-preserving method and high accuracy (98%)	No user-study validation
[157]	Explainable SMS spam detection RoBERTa model	99.84% detection rate, balanced data	lack of robust explainable approach like SHAP or gradient explainers
[10]	Privacy-preserving SMS spam detection with PhoBERT model	99.38% accuracy, privacy-preserving via FL, dual language support	Subjective data labeling for non-spam entries, synthetic evaluation data
[76]	Use CNN-LSTM model for detection	99.74% Accuracy, strong evaluation	Overfitting, low data diversity
[130]	Few-shot RoBERTa with augmentation	Cost-effective learning, 97.8% accuracy with RoBERTa model in Ten-fold augmented data	Needs human checks, attacker misuse risk
[172]	Agentic AI , LLMs detection , user-friendly explanation	Uses external context (URLs, WHOIS, HTML, screenshots, brand info) , multi-modal LLM, 98.8 % accuracy	Potential high computational cost of LLMs, external tool dependency

content are widely observed in the literature. Text-based representations such as *TF-IDF* and topic modeling methods like *LDA* (Latent Dirichlet Allocation) have demonstrated strong performance, indicating that linguistic analysis is highly smishing detection [117]. Combining NLP with ML provides substantial ground for addressing smishing attacks [158].

Several studies focus on extracting semantic and structural information from SMS messages. For example, Liu et al. [67] proposed a NLP-based detection approach with entity recognition (NER), regular expressions for PII, and syntactic parsing that achieved a high precision in detecting spearphishing messages. Similarly, transformer-based models such as BERT [46] and RoBERTa [130] have been widely adopted due to their ability to capture contextual relationships in text. These models have been further enhanced using data augmentation techniques and prompt-based methods to improve detection performance [157].

Additionally, recent approaches use LLMs for both detection and explanation. For instance, *SmishX* [172] integrates external contextual information such as domain data and web content to improve reasoning and generate evidence-based explanations, and achieved high accuracy on real-world datasets. Similarly, recent work on LLM-based frameworks for classification and explanation demonstrates improvements over earlier models [66].

While these approaches significantly improve detection accuracy, they often rely on additional contextual information and computational resources. Furthermore, some methods designed for related domains, such as Online Social Network (OSN)-based *impersonation* detection [69], may not directly apply to smishing due to the limited context in SMS messages. To address this limitation, systems such as SmishViz [120] focus on identifying patterns across messages and shared infrastructures, enabling more effective detection of impersonation-based attacks.

5.1.5 Federated Learning-Based Detection Approaches

Researchers are also deploying federated learning (FL) to address the limitations associated with centralized data collection. This approach enhances privacy and security by eliminating the need to transmit raw data to a central server. Instead, models are trained locally, and only updated weights are shared and aggregated to improve a global model in an iterative manner.

Sidhpura et al. [134] leveraged a periodic training schedule on recent SMS data where local models predict and store labels on client devices while a central DistilBERT-based model aggregates updates from distributed clients. Similarly, [10] proposed an FL framework using the PhoBERT model for Vietnamese SMS spam detection and achieved competitive performance compared to centralized approaches while preserving data privacy. Moreover, Shanto et al. [126] developed an LSTM-based global model, while, Remmide et al. [107] applied a BiLSTM model with *FedAvg* aggregation, but in both cases they suffered from low accuracy in the federated setup. In contrast, Rose et al. [111] used a *ConvLSTM* model with *FedAvg* and achieved improved accuracy after multiple federated iterations.

While FL-based approaches effectively address privacy concerns in SMS data, they introduce challenges such as heterogeneity in smishing and legitimate message distributions across users, communication overhead, and variations in model performance compared to centralized training.

5.1.6 Hybrid Detection Approaches

Hybrid methods for smishing detection fully exploit the strengths offered by multiple ML, DL algorithms, and rule-based or a combination of these, instead of a single method, thereby reducing the possibility of false positives. A hybrid algorithm comprising CNN and LSTM gives an accuracy score of 99.74%, which efficiently identifies significant features from text messages [76]. Another hybrid method combines Bidirectional Gated Recurrent Units (Bi-GRUs) and CNN for enhancing smishing classification, using text embeddings and Word2Vec trained on preprocessed text [72].

5.1.7 Explainability and Transparency of Detection Approaches

Explainability remains a critical yet often overlooked challenge in smishing detection, despite the widespread adoption of ML and DL models. The black-box nature of architectures such as CNNs and LSTMs makes it difficult to interpret classification decisions, thereby limiting trust, regulatory compliance, and broader adoption.

Recent research has explored explainability techniques such as SHAP (SHapley Additive exPlanations) [70] and LIME (Local Interpretable Model-agnostic Explanations) [109], which have been applied in various cybersecurity [77] scenarios to identify key features influencing predictions. However, their application to smishing detection remains under-studied. In current literature, Shendkar et al. [128] show that their methods improve interpretability by clarifying model behavior for analysts, while Uddin et al. explored attention mechanisms within transformer models such as BERT to enhance the interpretability of smishing detection [157],

but these models are often computationally expensive. In addition, hybrid approaches that combine ML models with interpretable rule-based techniques are proposed to balance accuracy and transparency [86]. Despite these efforts, many existing approaches rely on technical outputs, such as feature relevance scores or attention weights, which are typically comprehensible to domain experts but lack the intuitive clarity required for a non-expert user.

More recent work shifts toward user-friendly explanations where Lee and Han [66] employ fine-tuned KULLM to generate structured explanations. Compared to feature-based methods, LLM-based approaches provide more coherent and context-aware reasoning. Notably, Wang et al. [172] propose a system that generates evidence-based explanations for lay users by augmenting chain-of-thought LLM reasoning with external context (e.g., domain info, brand info, URL redirection, and web screenshots), producing concise outputs that include detection results, key reasons, and actionable advice.

5.2 Anti-Smishing Tools and Mobile-based Defense Measures

5.2.1 Third-party Anti-Smishing Tools

In literature, Timko et al. [153] provide a comprehensive benchmark of anti-smishing tools and a qualitative analysis of smishing attacks. Third-party anti-smishing solutions serve as supplementary protection alongside carrier-level filtering mechanisms. While mobile carriers can filter or block suspicious messages at the network level, these applications provide an additional layer of detection to identify phishing messages that bypass such filters [153]. These apps mostly work as filters and classifiers to categorize messages into spam or phishing. Content-based filtering approaches, including heuristic and machine learning methods, are widely used due to their ability to detect zero-day smishing attacks, unlike traditional list-based filtering. The effectiveness of these tools varies across platforms. For example, applications such as Robokiller [110] and Textkiller [149], showed higher smishing hit rates. The study tested various anti-smishing apps, including Key Messages, Anti Nuisance, Call Control, and Calls Blacklist for Android, and SpamHound [140], Robokiller, Malwarebytes [73], and NomoRobo [93] for iOS. Among these tools, Textkiller had a significant smishing hit rate, and Robokiller demonstrated the highest smishing hit rate among the tested apps while blocking a high percentage of benign messages. NomoRobo also achieved strong performance with lower false positives. There are some other tools called Bulk Messaging Services, designed for legitimate mass communication, but can be exploited by attackers. These tools attempt to filter out malicious content. Among them, SimpleTexting [135] and Text-Em-All [148] achieved higher smishing hit rates than other services, for example, Twilio [156], SlickText [137], and TextSpot [150]. This shows a better impact in identifying and blocking phishing messages [153]. Another proposed solution for mitigating smishing attacks on mobile platforms, (MMSAMP) [92], introduces an Anti-Smishing Application (ASAPP) that authenticates sender IDs, preprocesses data, filters smishing content, and classifies messages as safe or unsafe. The system also uniquely informs users of potentially harmful content, a feature not found in existing solutions, while achieving 94% efficiency in detecting and classification.

5.2.2 Mobile-based Solutions and User Alerts

Mobile-app-based solutions and user alerts are crucial defense mechanisms against smishing attacks. These solutions address the unique vulnerabilities of mobile devices. As highlighted by Shahriar et al. [123], small screen size makes it difficult for users to identify legitimate websites from spoofed ones. To prevent this, mobile security solutions often employ content-based filtering to analyze message context and embedded URLs. Additionally, user awareness remains important, as users are often less familiar with mobile security features compared to desktop environments. Best practices such as installing official applications, using secure browsers, and bookmarking trusted sites can reduce the risk of falling victim to smishing. Furthermore, [123] stricter app store controls and mobile security applications can help limit malicious applications.

Researchers have explored the use of security features on mobile devices, such as two-factor authentication. While 2FA adds an additional layer of protection, Hossein Siadati et al. [132] demonstrates that SMS-based mechanisms remain vulnerable to social engineering attacks, specifically Verification Code Forwarding Attacks (VCFA), where users are tricked into sharing authentication codes. Moreover, the authors discussed two more strategies to decrease smishing attacks: (1) *malicious request detection:* which involves profiling users using ML with location data from SMS providers to limit attackers, (2) *sender verification:* enhance user interface usability to highlight sender information to SMS recipients. Furthermore, effective warning messages are also emphasized, such as: (1) *abuse-proof:* meaning the content of the warning should be carefully mentioned to prevent attackers from abusing it to their advantage

Table 6: Critical Analysis of Existing Anti-Smishing Tools and Defense Measures

Defense Strategy	Reference	Countermeasure Approach	Performance Metrics	Strengths	Weaknesses
	[123]	SMS/URL filtering with blacklists and trusted number whitelist	-	Detects various mobile phishing types, low false positives, and computationally lightweight	Limited detection methods, reliance on user awareness, and operative system or device dependency
	[92]	Sender authentication + con- tent filtering + user awareness	-	Focuses on user awareness, combined filtering	Naive Bayes reliance lacks evaluation details
Mobile-Based	[64]	Neural Networks (CNN, RNN, Simple NN) + Translator API + Floating Bubble UI	99.68% ACC	multilingual detection without language- specific datasets, real-time monitoring, user-friendly interface for social media platforms	Potential struggle with inten- tional misspellings/grammar errors in fraudulent messages, challenges with hybrid languages (e.g., Singlish, Tanglish)
	[4]	Mobile app for SMS spam, user-controlled filtering	98.42% ACC	Handled missing data well	Limited features, struggles with missing data
	[171]	Call-to-verify system with dynamic Message ID and static Secret Words	SUS: 79.1/100	Inclusive (works on feature phones without apps), resilient to omission-based spoofing attacks	High user effort (requires dialing), Scalability issues for multiple services
	[48]	SMS slang normalization + Naive Bayes classifier	96.2% ACC, 97.14% TPR, 96.12% TNR	Effectively handles informal SMS language.	English-only slang dictionary, content-only (no URL/app behavior yet)
	[132]	Improving the message verification mechanisms with warnings	8% Success Rate (Attack)	Effective multi-factor authentication security (e.g., 2FA) and scalable tools	only focus on SMS 2FA with limited use-case scenarios
User Awareness	[16]	Phishing awareness campaign using simulation	Smishing campaign success rate: 4.17%	Comprehensive platform comparison, user behavior analysis	Small sample size, static secu- rity, single campaign
	[29]	Phishing detection from URLs using ML techniques	73.4% OA (87.6% phishing vs. 59.2% genuine)	User-centric focus, cue-based behavioral analysis	Demographic bias, age imbalance, artificial survey environment
Law Enforcement	[178]	ACAEC law: SMS abuse reporting system + senderID/CallerID spoofing controls + operator blocking of fraudulent texts/domains	-	Direct legal basis for prosecution, compels industry action, provides public reporting mechanism	Implementation and compliance burden; relies on user reporting only
Infrastructure-Based	[89]	Analysis of SMS phishing campaigns, infrastructure, and illicit markets (derived from public gateway data) to inform investigations and disruption efforts against cyber-criminals	Smishing dataset with 68K records	Offers detailed insights into phishing, infrastructure, and supply chains to aid targeted disruption and actor identification.	Data source bias (e.g., Western focus, no MMS), reliance on limited external detection tools, and reveals current regulatory gaps leading to abuse shift to unmonitored channels.
Cross-Device	[60]	Cross-device approach using AR and URL analysis	SUS: 74.4/100	Reduced device vulnerability, user engagement	Relies on AR tech, camera limitations
Social Media Analysis	[146]	SpamHunter pipeline (image detection, tweet classification, text recognition)	95% Prec, 87% TPR	Up-to-date, multi-lingual corpus, surfaces spam URLs days–weeks before VirusTotal	Relies on voluntary Twitter reports/Twitter API, sampling bias, no direct SMS blocking
Detection and Filtering	[153]	Crowd-sourced smishing data collection	55 zero-day samples, 0%–61.6% Detection Rate	Used real smishing data, novel dataset approaches	Small sample size (20 tested) and selection bias (user-reported samples only)

Keys: ACC: Accuracy, SUS: System Usability Scale, TPR: True Positive Rate, TNR: True Negative Rate, OA: Overall Accuracy, Prec: Precision,

for social engineering attacks, (2) *worry-free:* meaning the warnings should include the risk alerts by telling users that no immediate action is required if they have not requested the code, reducing unnecessary anxiety, (3) *actionable and practical:* it should provide clear instructions on how to avoid the hazard, and (4) *concise and clear:* it should be brief, easily understandable, and avoid technical jargon. It is interesting to highlight that the study mentions risk alerts stating "Please ignore this message if you did not request a code" before the verification code significantly reduced the attack success rate [132].

Another approach explores cross-device solutions for smishing detection. Kanaoka et al. [60] propose a system using external devices, such as augmented reality (AR) glasses, tto capture and analyze URLs from screens and provide real-time risk indicators. While this approach improves user decision-making, with an 85.19% correct rate with the prototype compared to 46.3% without the glasses, its reliance on additional hardware limits its practicality for widespread adoption.

Mobile-based detection applications have also been proposed in literature. *SMSPROTECT* [4] intercepts incoming messages and applies ML-based classification to determine whether messages are legitimate or malicious, and allows users to take appropriate action. Similarly, ML and NLP models can be integrated into mobile applications or on-device systems to detect abnormal language patterns and generate user alerts [133]. However, these approaches require continuous model updates to remain effective against evolving smishing techniques.

Another solution, *VeriSMS* [171], introduces a verification mechanism using message identifiers and user-specific secret words to validate message authenticity. While this approach improves trust, it relies heavily on user participation and introduces usability challenges, such as manual verification steps and potential input errors. In addition, its effectiveness may be reduced if users do not consistently follow verification procedures.

There are other proposed methods that incorporate mobile applications for smishing detection by analyzing incoming messages for suspicious message contents, links, or sender information [48, 131, 24, 25]. These systems can leverage real-time threat intelligence to identify suspicious messages. However, the effectiveness depends on user adoption, model updates, and the ability to address privacy concerns when analyzing personal SMS data.

5.3 User-Training, Awareness, and Regulation-based Defense Measures

5.3.1 User Training and Awareness Practices

User awareness emphasizes the importance of educating users to identify various red flags while interacting with SMS messages [62]. Clasen et al. [29] used an online questionnaire to examine which features help users detect smishing, which shows that awareness training plays a vital role in enabling individuals to recognize and avoid such attacks. Training programs can highlight common smishing indicators, including shortened URLs, inconsistent message content, and urgency style language [29]. This study also suggests that interactive simulations and real-world scenarios can enhance user engagement and retention of defensive strategies [29]. Blancaflor et al. [16] further emphasize the role of user awareness through a phishing campaign that included smishing. Although their smishing campaign had a low success rate (4.17%), their findings indicate that trust remains a key factor influencing whether users interact with malicious links. This suggests that even with existing security measures, human factors such as trust and curiosity continue to be primary vulnerabilities. Finally, we find evidence from literature that user awareness training, combined with technological solutions, supports a multi-layered defense against smishing attacks.

5.3.2 Social Media Analysis and Awareness

People often report smishing messages on social media to warn others, seek help, or notify impersonated services. This behavior creates an opportunity to use platforms like X (previously Twitter) as a real-time and collective user reporting crowd source, which enables early detection and rapid response against ongoing smishing campaigns. For instance, researchers have proposed a solution that allows users to verify smishing attempts from social media platforms such as Facebook and WhatsApp using ML-based English translated content analysis [64]. Furthermore, SpamHunter enables researchers to collect spam, scam, and smishing messages derived from X (previously Twitter) posts where users shared screenshots of these messages [146]. The authors highlighted the strength of this approach, evidenced by the fact that more than 15% of the dataset's reported URLs were subsequently flagged by threat platforms (e.g., VirusTotal).

5.3.3 Policy Enforcement and Regulatory Efforts

As smishing continues to rise, policymakers must enforce coordinated measures through regulation and mon-

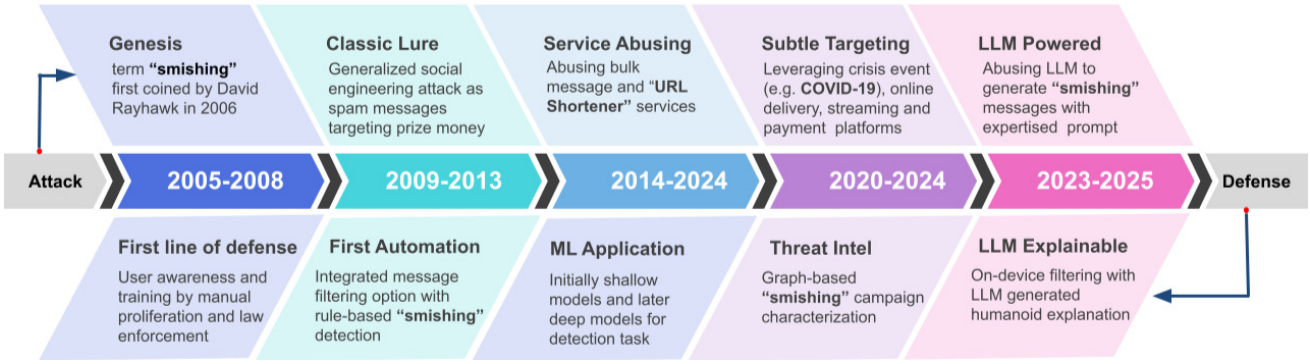


Fig. 6: Evolution in smishing attack (upper boxes) and defense (lower boxes) strategies over the years

itoring. Law enforcement agencies are increasingly tasked with tracking cybercriminals across jurisdictions, while regulatory bodies aim to define and mitigate these threats within existing legal frameworks [178, 143]. In the United States, several federal and non-profit organizations support prevention, reporting, and recovery efforts. The *Federal Trade Commission* (FTC) enables consumers to report fraud and provides guidance on mitigating financial loss, while the *Internet Crime Complaint Center* (IC3) collects incident data to support law enforcement investigations [143]. Additional organizations, such as the *Office for Victims of Crime* (OVC), *AARP Fraud Watch Network*, and the *National Center on Elder Abuse* (NCEA), provide support services, awareness tools, and resources for vulnerable populations [33, 143].

Despite these efforts, challenges remain. Recent research [89] highlights a significant increase in SMS and voice phishing attacks, indicating that regulatory measures alone are insufficient. For example, actions by the Federal Communications Commission (FCC) to block robotext messages have targeted application-to-person (A2P) messaging channels. While these measures reduce abuse in A2P traffic, they have also shifted attackers toward exploiting less regulated person-to-person (P2P) communication channels. The continued presence of illegal bulk messaging services on public platforms further demonstrates the limitations of current enforcement mechanisms.

From a legal perspective, different regions adopt distinct approaches. The European Union emphasizes centralized and proactive regulation through frameworks such as the General Data Protection Regulation (GDPR), which governs all electronic communications in the EU, including SMS marketing, and enforces strict consent requirements and data usage transparency [170, 136]..

The *General Data Protection Regulation* (GDPR) governs all electronic communications in the EU, including SMS marketing. It requires businesses to get clear permission from users, explain how their data will be used, and give them an easy way to stop receiving messages, which helps prevent fraudsters and unapproved marketers from misusing SMS services [170, 136]. The *EU NIS2 Directive* further strengthens cybersecurity obligations by requiring communication providers to implement risk management and incident reporting mechanisms, indirectly addressing large-scale smishing threats [34].

Beyond the EU, the United Kingdom addresses smishing through the *Fraud Act 2006*, particularly under Fraud by False Representation, which enables direct criminal prosecution of deceptive communication practices [99]. Similarly, Poland took a more specific approach by introducing the *Act on Combating Abuses in Electronic Communication* (ACAEC) in 2023, which provides a legal definition of smishing and permits regulatory authorities to block malicious messages and penalize caller ID spoofing [178]. In the Asia-Pacific region, hybrid frameworks such as Singapore's PDPA [50] and Thailand's PDPA [90] combine privacy protection with telecommunications controls and enforce penalties for misuse of messaging systems. In contrast, the United States relies on sector-specific state laws. The *California Consumer Privacy Act(2018)* (CCPA)[19] provides a state-level framework that gives consumers rights to access, delete, and restrict the use of their personal data, particularly in marketing communications [100, 59].

While these frameworks provide strong consumer privacy protections, they do not directly address smishing as a distinct threat (especially in the US agencies or laws), limiting early intervention capabilities. Tables 7 and 8 further illustrate these differences and highlight the contrast between the EU's unified regulatory approach and the fragmented enforcement landscape in the US.

Real-World Cases with Policy Enforcements

United States: Despite the robust legal frameworks, a gap remains between policy implementation and ef-

Table 7: Comparison of Common Features Across Regulatory Frameworks in the US and EU Regions

Feature	US (TCPA & State Laws) [59]	California (CCPA) [100, 59, 22]	Poland (ACAEC) [178]	EU (GDPR) [136]
Scope	Governs telemarketing & SMS as calls	Privacy & data selling restrictions for CA residents	Focused on smishing & telecom abuse	Comprehensive data privacy regulation
Consent	Prior express consent needed	Consent for data use; opt-out for data selling	Not main focus, targets fraud	Explicit opt-in required [136]
Penalties	State & federal fines (up to $20K/msg)	Up to $7,500 per violation, lawsuits allowed	Criminal (3 months–5 years)	Up to EUR 20M or 4% revenue
Reporting	FCC + state rules	Complaint & lawsuit mechanisms	Toll-free reporting (8080)	General breach notification
Data Handling	Disclosure of identity & opt-out options	Must disclose data practices & third-party sharing	Allows telecom data sharing to combat fraud	Strict transparency & privacy policies
Enforcement	Federal, state laws	CA Attorney General	National body (CSIRT NASK)	National supervisory authorities
Cross-Border Coordination	Complex US compliance	Limited to California only	Minimal, national focus	EU-wide harmonization
Consumer Empowerment	Do-Not-Call registry, opt-out	Rights to know, delete, opt-out	Reporting channel	Broad data rights (access, erase)

Table 8: Legal and Policy Categories Related to Smishing

Category of Law/Policy	Primary Governing Body & Relevant Acts	How It Applies to Smishing	Typical Penalties
Consumer Protection	Federal Trade Commission (FTC): Enforces consumer protection laws that prevent fraudulent, deceptive, and unfair business practices[42]. The CAN-SPAM Act (2003)[20] also applies to some commercial texts.	FTC investigates and penalizes smishing activities that fall under its broad mission to protect users. They focus on how these messages trick people and are used to commit fraud. They can use the CAN-SPAM Act (2003) to stop and punish these commercial text scams [45, 20].	FTC mentions about civil penalties and fines of up to $50,120 per violation [45].
Unsolicited Communications	The Federal Communications Commission (FCC) regulates interstate and international communication[36]. The Telephone Consumer Protection Act (1991) [37] restricts the use of auto-dialed calls and text messages [59].	The FCC fights smishing by mandating that mobile providers should block illegal robotexts, especially those from invalid or spoofed numbers. Under the Telephone Consumer Protection Act (1991) [37], the FCC requires companies to get consumer consent before sending marketing texts. The agency also uses consumer complaints on spam and reports forwarded to 7726 to inform its enforcement actions and to help providers block illegal messages [40, 39].	FCC mentions that civil penalties and fines could be up to $10,000 for each violation of illegal spoofing [39]. Moreover, The Telephone Consumer Protection Act (1991) imposes statutory damages of $500 to $1,500 per violation [30].
Fraud & Identity Theft	Department of Justice (DOJ) and Federal Bureau of Investigation (FBI) are the primary law enforcement and legal bodies for fraud and identity theft protection [38, 41]. Further details can be found at Title 18, U.S. Code, Section 1343 (Wire Fraud Statutes, 2008) [159, 161] and Title 18, U.S. Code, Section 1028 (Identity Theft and Assumption Deterrence Act, 1998) [163, 160].	Smishing is often just the initial step in a larger criminal campaign. The Wire Fraud statute (2008) [159] is a powerful policy tool since it criminalizes the use of an electronic medium, such as a text message, to execute a plan to defraud [161, 31]. The Identity Theft and Assumption Deterrence Act (1998)[163] criminalizes the knowing transfer or use of another person's identifying information to commit any unlawful activity.	In case of wire fraud, fines and/or imprisonment for up to 20 years. Penalties can increase to 30 years for scams targeting financial institutions [31]. In case of identity theft, fines and/or imprisonment for up to 15 years, with higher penalties for aggravated identity theft [162].
State-Level Action	State Attorneys General and various state laws. For example, California's Anti-Phishing Act (2005)– specifically targets phishing and spoofing [18]. Virginia's Computer Crimes Act (1984) contains provisions that can be applied to smishing [169].	States create specific legislation to complement federal laws and prosecute crimes that occur in their areas. These laws extend the definitions of computer and identity theft to include smishing and allow state prosecutors to pursue cases independently of federal agencies. [173, 169]	Penalties by state and often include major fines and felony charges, which can lead to imprisonment. For example, Virginia's laws can result in felony charges with up to 10 years in prison [169].

fective enforcement. While the Department of Justice pursues large-scale criminal cases, such as coordinated smishing campaigns targeting corporate employees [164], general consumer-level smishing continues to result in significant financial losses, suggesting that current measures are insufficient [44].

United Kingdom: In contrast, the UK demonstrates more direct enforcement. The Information Commissioner's Office (ICO) has issued substantial penalties for nuisance calls and texts, and law enforcement agencies have taken active measures, such as arresting individuals operating unauthorized SMS transmission devices [28]. These cases reflect a more immediate enforcement approach compared to reliance on financial penalties alone.

5.3.4 Summary of Smishing Defense Landscape

In summary, the evolution of smishing attack and defense strategies is a sophisticated "arms race" between cybercriminals and security researchers, transitioning from crude bulk spam to highly personalized, AI-driven deception as shown in figure 6. In its infancy (2005–2008), smishing was a novelty defined by manual outreach and basic law enforcement intervention [141]. As mobile adoption surged through 2013, attackers pivoted to "classic lures," such as lottery or prize winning scams, which necessitated the first wave of automated, rule-based filtering. The subsequent decade (2014–2024) saw an escalation where bad actors weaponized bulk-messaging infrastructure and URL shorteners [58, 146]. Defenders countered by integrating machine learning, moving from shallow statistical models to complex deep learning architectures to identify malicious patterns at scale [10, 56, 157, 81, 49].

In recent years, the landscape has shifted toward extreme contextualization and technical stealth. From 2020 to 2024, attackers capitalized on global crises (e.g. Covid-19 [98]) and the rise of the "delivery economy" [154] to craft hyper-relevant messages that bypassed traditional skepticism. During this same window, defenders adopted threat intelligence frameworks, utilizing graph-based analysis to characterize entire smishing campaigns rather than treating messages as isolated incidents [120]. Now entering a frontier (2023–2025) which is dominated by Large Language Models (LLMs). While attackers use generative AI to automate flawless, expert-level social engineering [129, 51], the defensive community is responding with on-device LLMs that provide humanoid, explainable insights, thus transforming mobile security from a passive filter into an interactive user-centered defense and awareness tool [172, 119].

6 Evaluating Existing Smishing Datasets

6.1 SMS Phishing Datasets For Research

We analyzed 14 SMS datasets and 1 email spam dataset, widely used in prior work (Table 9). Among these, 11 datasets include labeled data, such as *Enron Email Spam* [166], *British English SMS* [94], *UCI Machine Learning Repository SMS Collection* [9], *Mendeley Smishing* [116], Super Dataset [84], *Korean Smishing*, *RevisedIndian* [127], *Bengali SMS Spam* [26], *Bangla Barta* [124], and MOZ-Smishing [8]. The remaining four datasets—*NUS SMS Corpus* [23], *SMS Gateways* [89], *Spam Hunter* [146], and *Smishtank* [154], lack explicit labels, though some contain only spam or smishing messages by design; labeling inconsistencies across studies were also noted.

Table 9 summarizes dataset size, message length, and label information. Table 10 further reports URL-related metrics, including counts of messages with URLs, unique FQDNs, and unique live and parked websites during our analysis period. FQDN counts are based on original URLs (excluding expanded shortened links). URLs were extracted using a custom rule-based method. When multiple URLs appeared (e.g., in Enron, Super, and Spam Hunter), only the first was considered for live website analysis. For reproducibility, we release both refined and original datasets in a unified Smishing Data Hub at `https://github.com/MarazMia/SMISH_DT`. Further dataset-specific details are discussed below.

6.1.1 Public Datasets:

(i) Enron Email Spam [166]: Although originally curated from spam emails, this dataset has been used for SMS phishing detection due to similar textual characteristics. Published in 2006, it contains 33,715 instances labeled as `ham` or `spam`, with a near-balanced distribution. The average message length (1446.84 characters) is significantly higher than in typical smishing datasets. As shown in Table 10, 33.08% of messages contain URLs, with 80.19% of those appearing in spam. The dataset includes 3,433 unique FQDNs and 3,456 unique links, of which 564 are live and 193 are parked. Notably, spam messages contain more links than ham messages.

(ii) English SMS [94]: This dataset combines two corpora with manually selected `ham` and `spam` messages, totaling 875 instances. The average message length is 106.89 characters overall, with `ham` messages shorter (78.89) than `spam` (136.4). Owing to its small size, it has the fewest URL-containing messages (64), all within `spam`. No live websites appear in `ham`; among 41 unique FQDNs, 10 websites are live and 7 are parked.

(iii) UCI ML Repo [94]: This dataset merges three sources: the *English SMS dataset* (full), the *National University of Singapore SMS dataset* [23] (3,375 manually selected messages), and the obsolete *SMS Spam Corpus v.0.1 Big dataset*. It contains 5,574 instances (4,827 `ham`, 747 `spam`) with similar average lengths for `ham` (71) and `spam` (139). Only 1.94% of messages include URLs, of which 98.14% are `spam`. A total of 54 unique domains are identified, with 13 live websites (10 from `spam`); notably, 10 of these live sites are parked, predominantly from `spam` (9/10).

(iv) ExAIS SMS Spam [97]: This dataset, collected at the Federal University of Agriculture, Abeokuta, Nigeria, comprises SMS messages from 20 volunteers, in-

Table 9: Summary Information of Publicly Available SMS/Smishing Datasets (labels are defined as: **ham** - legitimate message, **spam** - unsolicited and unwanted message, **smish** - phishing message containing malicious URL)

Dataset	Language	Format	Labeled	# Instances				Avg. Message Length			
				full	ham	spam	smish	full	ham	spam	smish
Enron Email spam [166]	English	.txt	yes	33,715	16,545	17,170	-	1446.84	1651.67	1249.47	-
English SMS [94]	English	.doc	yes	875	450	425	-	106.83	78.89	136.4	-
UCI ML Repo [9]	English	.txt	yes	5,574	4,827	747	-	80.48	71.47	138.68	-
ExAIS SMS Spam[97]	English	.csv	yes	5,240	2,890	2,350	-	136.88	132.42	142.374	-
RevisedIndian[127]	Multiple	.xls	yes	4,567	3,356	1,211	-	140.64	140.86	140.04	-
Mendeley Smishing [84]	English	.csv	yes	5,971	4,844	489	638	83.24	70.70	133.90	139.59
Super Dataset [116]	English	.csv	yes	67,008	40,830	26,178	-	96.09	65.19	144.30	-
Kor-Smishing [63]	Korean	.csv	yes	43,209	42,594	-	615	44.83	41.20	-	296.1
Bengali SMS Spam[26]	Bengali	.csv	yes	2,602	1,199	1,403	-	91.00	100.39	82.97	-
Bangla Barta[124]	Bengali	.csv	yes	2,772	1,848	-	924	75.31	55.59	-	114.75
MOZ-Smishing[8]	Portuguese	.csv	yes	2,561	2,009	-	552	102.68	104.57	-	95.79
NUS Corpus [23]	English	.json	no	55,835	-	-	-	52.01	-	-	-
Spam Hunter [146]	English	.txt	no	25,826	-	25,826	-	147.36	-	147.36	-
SMS Gateways Smishing [89]	English	.csv	no	68,029	-	-	68,029	112.45	-	-	112.45
SmishTank [154]	English	.csv	no	1,091	-	-	1,091	176.96	-	-	176.96

cluding 2,350 `spam` and 2,890 `ham`. The average message lengths are similar across classes (132.42 for `ham`, 142.37 for `spam`). As shown in Table 9, 193 messages (3.99%) contain URLs, mostly in `spam` (138). From Table 10, 115 unique links are identified, with 79 live (59 from `spam`) and 35 parked websites, 22 of which belong to `spam`.

(v) RevisedIndian [127]: This dataset contains 4,567 real-world SMS messages in English, Hindi, and Telugu, though it is predominantly English (91.24%). It is labeled as `ham` and `spam`, with 73.48% (3,356) `ham`. The average message length is about 140 characters. Analysis shows 1,441 messages contain URLs, of which 867 (60.17%) are `spam`. Among 720 unique websites, 450 are live, including 249 from `spam`. Additionally, 180 live websites are parked, nearly half (96) associated with `spam`.

(vi) Mendeley Smishing [84]: The dataset is a partial clone of the UCI dataset, containing 5,971 instances (4,844 *ham*, 489 *spam*, and 638 *smish*). It shares 4,753 entries with UCI (4,199 `ham`, 554 `spam`), while `smish` messages were sourced from Pinterest images and converted to text. Published in 2022, it is the first to explicitly use the `smish` label. Message lengths are similar to UCI, with `smish` (139.59) closer to `spam` (133.90). The dataset includes 233 URL-containing messages, all from `spam` and `smish`, with 125 unique websites and 39 live ones; no `ham` messages contain URLs.

(vii) Super Dataset [116]: *Salman et al.* compiled a large and diverse dataset of 67,008 SMS messages by aggregating existing sources (UCI, NUS, Spamhunter, and three multilingual datasets) and adding 4,904 new messages from platforms such as Twitter and Scamwatch. This includes 40,837 (60.9%) `ham` and 26,181 (39.1%) `spam`, with 2,920 newly contributed spam samples. The average message length is ∼96 characters overall and ∼144 for `spam`. We identified 10,492 (15.66%) messages containing URLs, of which 10,472 (99.81%) are `spam`. Among 7,759 unique URLs, 1,645 are live websites, almost all from `spam` (1,641). Additionally, 903 websites are parked, with nearly all (902; 54.83% of live sites) appearing in `spam` messages.

(viii) Kor-Smishing [63]: This is the only public dataset with Korean messages labeled as 1 (`smish`) and 0 (`ham`). It is highly imbalanced, containing 42,594 `ham` and 615 `smish` instances. Due to the language, average message lengths differ significantly (41.20 for `ham` vs. 296.1 for `smish`). Despite its size, only 0.31% of messages include URLs, among the lowest across datasets. A total of 120 unique URLs and 75 websites are identified, with 66 live (63 from `smish`) and 23 parked. Only 6 messages contain non-Korean (mostly English) text.

(ix) Bengali SMS Spam [26]: This Kaggle-sourced dataset contains 2,602 Bengali SMS messages, with only 0.58% (15) fully in English. It is relatively balanced, comprising 46.08% `ham` and 53.92% `spam`. The average message length is 91 characters, with `ham` messages longer (100.39) than `spam` (83.97). Only 6.88% (179) of messages contain URLs, nearly half of which (49.72%) are `spam`; URLs in `ham` are mainly government or carrier-related links. A total of 58 unique websites are identified, with 55 live (38 from `spam`) and 23 parked, mostly associated with `spam` (20).

(x) Bangla Barta [124]: This Bengali SMS dataset contains three categories: `normal`, `promo`, and `smish`. We merge `normal` and `promo` into a single `ham` class. It is balanced with 2,772 total instances (924 per category). Only 14 messages (0.51%) are fully English, often in mixed *Banglish*. The average message length is higher for `smish` (114.75) than for `ham` (55.59). A total of 410 messages contain URLs, with 272 (66.34%) belonging to `smish`. As in other datasets, `ham` URLs are mostly government or carrier-related. We identify 151 unique websites, 128 of which are live and 119 from

`smish`. Additionally, 26 websites are parked, with only 2 associated with `ham`.

(xi) MOZ-Smishing [8]: This Mozambican Portuguese SMS dataset contains 2,561 messages, with 1,816 (71%) written using English alphabets. It is labeled as `ham` (legitimate) and `smishing`. The average message length is 102.68 characters, with `smishing` messages slightly shorter on average. Only 31 messages (1.21%) contain URLs, and none of these appear in `smishing`. In contrast, 768 messages (29.99%) include phone numbers, of which 324 (42.18%) are `smishing`, indicating phone-based attacks are more prevalent than URL-based ones in this context. Among 9 unique links identified, 8 are live and 4 are parked websites.

(xii) NUS Corpus [23]: The original dataset [23] was submitted to arXiv in 2011 and published in 2013, with a later version in 2015. It includes English and Chinese corpora provided in XML, SQL, and JSON formats, and both are unlabeled. The Chinese dataset contains 31,465 instances with mostly descriptive text and few URLs, while the English dataset has 55,835 messages with an average length of 52 characters. Only 0.04% of messages include URLs (the lowest among all datasets). These URLs correspond to 18 unique domains, of which 10 are live and 3 are parked.

(xiii) Spam Hunter [146]: This 2022 dataset was constructed from Twitter API data (2018–2021), targeting spam-related content including *spam, phishing, malicious, smish, fraud*, and *scam* messages with associated screenshots. The authors re-verified images using object detection and classified messages via a neural sentiment model (91% accuracy), yielding 21,918 curated spam messages. The released corpus contains 25,826 instances with an average length of 147.36 (Table 9). Overall, 60.14% of messages include URLs, corresponding to 8,615 unique domains, with 7,269 live and 4,052 parked websites (Table 10).

(xiv) SMS Gateways Smishing [89]: Released in 2023, this is the largest SMS dataset, containing 68,029 messages collected from 11 SMS gateways via web crawling. All messages are confirmed smishing using VirusTotal, APWG, and Google Safe Browsing. The original study [89] provides two files: one with all phishing messages (68,029 instances) and another with 35,128 phishing campaigns including URLs, timestamps, and repeated exposures (same message content with varying identifiers such as sender or links). The average message length is 112.45 characters, and 63.64% of messages include URLs (Tables 9 and 10). Despite 43,297 unique links, only 857 unique FQDNs are observed. Even after confirmation as phishing, 6,945 websites remain live and 572 are parked.

(xv) SmishTank [154]: SmishTank is an active system and one of the most recent smishing datasets (2024), continuously collecting user-submitted messages (images and text) via its website. It verifies inputs using VirusTotal and automatically extracts metadata such as URLs, sender information, and brand mentions. The current CSV snapshot contains 1,091 entries with no confirmed labels, though messages are likely smishing by design. The average message length is 177 characters. About 86% of messages contain URLs, with 740 unique FQDNs among 860 unique links. Of these, 193 websites are live and 82 are parked.

6.1.2 Private Datasets:

As shown in Table 11, prior studies use private smishing datasets either exclusively [17, 92, 122] or in combination with public data [47, 46, 58, 61]. The KISA dataset [122], managed by KISA, focuses on Korean smishing messages and retains only URL-containing samples. Although augmented with public normal messages from three open datasets for generalization, the final dataset remains non-public. Using a character-level CNN, it achieves 99.60% accuracy and a 99.44% F1-score [122]. In contrast, Gourabi et al. [47, 46] use a hybrid English–Arabic dataset integrated into a mobile environment. These studies report similar performance, achieving 99.63% accuracy and 95.58% F1-score using BERT.

6.2 Performance of Various Detection Models Using Datasets

Table 11 summarizes smishing detection studies using private, public, or hybrid datasets. The feature space varies from standard text vectorization to custom features extracted from SMS text, URL lexical patterns, and domain-level website content [61, 138, 85, 4]. On the KISA dataset [122], a character-level CNN achieves 0.996 accuracy and 0.9944 F1-score. Similarly, Gourabi et al. combine UCI and a private Arabic dataset for multilingual detection. Their CNN-LSTM model yields 0.9837 accuracy but lower F1 (0.9148), while later enhancements incorporating VirusTotal URL checks, regex-based pattern detection, and BERT-FC improve performance to 0.9963 accuracy and 0.9558 F1-score [47, 46]. Karhani et al. [61] achieve 0.994 accuracy and 0.9901 F1-score using a hybrid Random Forest–SVC model on NUS and private telecom data. The ExAIS system [97] reports 99% accuracy using an artificial immune system with rule-based mechanisms. Among public datasets, UCI ML Repo (2011) remains widely used, with RoBERTa achieving state-of-the-art performance

Table 10: Analysis of Live and Parked Websites Among Corresponding Datasets (Queried Between August 2025 and October 2025)

Dataset	#Messages with URL	#Spam Messages with URL	#Smish Messages with URL	#Unique Link	#Unique FQDN	#Live Websites				#Parked Websites			
						full	ham	spam	smish	full	ham	spam	smish
Enron Email spam	11,152 (33.08%)	8,943 (54.05%)	-	3,456	3,433	564	198	366	-	193	30	163	-
English SMS	64 (7.31%)	64 (15.01%)	-	45	41	10	0	10	-	7	0	7	-
UCI ML Repo	108 (1.94%)	106 (14.19%)	-	59	54	13	1	12	-	10	1	9	-
ExAIS SMS Spam	209 (3.99%)	138 (5.87%)	-	115	66	79	20	59	-	35	13	22	-
RevisedIndian	1,441 (31.55%)	565 (46.66%)	-	720	113	450	201	249	-	180	84	96	-
Mendeley Smishing	223 (3.73%)	93 (19.01%)	130 (15.20%)	125	109	39	0	20	19	23	0	14	9
Super Dataset	10,492 (15.66%)	10,472 (40.00%)	-	7,759	5,610	1,645	4	1,641	-	903	1	902	-
Kor-Smishing	134 (0.31%)	-	127 (20.65%)	120	75	66	3	-	63	23	2	-	21
Bengali SMS Spam	179 (6.88%)	89 (6.34%)	-	88	58	55	17	38	-	23	3	20	-
Bangla Barta	410 (14.79%)	-	272 (29.44%)	151	42	128	9	-	119	26	2	-	24
MOZ-Smishing	31 (1.21%)	-	0 (0.00%)	9	5	8	8	-	-	4	4	-	-
NUS Corpus	21 (0.04%)	-	-	18	18	10	-	-	-	3	-	-	-
Spam Hunter	15,533 (60.14%)	-	-	15,531	8,615	7,269	-	-	7,269	4,052	-	-	4,052
SMS Gateways Smishing	43,297 (63.64%)	-	-	43,297	857	6,945	-	-	6,945	572	-	-	572
SmishTank	938 (85.98%)	-	-	860	740	193	-	-	193	82	-	-	82

Table 11: Evaluation of Existing Detection Models Using Both Public and Private Datasets

Data Availability	Reference	Datasets Used	Feature Space	Best Algorithm Model	Accuracy	F1 Score
Private	[17]	Private	TF-IDF	RF	0.9815	0.9257
	[92]	Private	Word Vectorizer	NB	0.9375	0.9333
	[74]	Private (Swahili)	TF-IDF	RF	0.9986	0.9986
	[122]	Private (KISA)	Word2Vec	Char-CNN	0.9960	0.9944
	[66]	Private (Korean)	Tokenized Text	KUBERT	0.9900	0.9900
	[10]	Private (English + Vietnamese)	Tokenized Text	PhoBERT (FL)	0.9938	-
	[131]	Private	Tokenized Text	RNN	0.9413	0.9400
Public + Private	[47]	UCI ML Repo + Private (Arabic)	Word Embedding	CNN-LSTM	0.9837	0.9148
	[46]	UCI ML Repo + Private (Arabic)	Tokenized Text	BERT-FC	0.9963	0.9558
	[58]	UCI ML Repo + Private (URL Phishing)	TF-IDF	Voting (RF, KNN, ETC)	0.9903	-
	[61]	NUS Corpus + Private (TELUS)	Heuristic + TF-IDF	RF + SVC	0.9940	0.9901
	[130]	Mendeley + GPT-3.5-Turbo Augment	Tokenized Text	RoBERTa	-	0.9820
Public	[138]	UCI ML Repo	BoW + Heuristic	AdaBoost	0.9867	0.9486
	[82]	Mendeley Smishing	TF-IDF	NB	0.9629	0.9200
	[83]	Mendeley Smishing	Word2Vec (ANN filter)	ANN	0.9740	0.8672
	[85]	Mendeley Smishing	Heuristic	MLP Back Propagation	0.9793	0.8150
	[107]	Mendeley Smishing	GloVe	BiLSTM (FL)	0.8878	0.8700
	[111]	UCI ML Repo	Word Embedding	ConvLSTM (FL)	0.9919	-
	[158]	UCI ML Repo	TF-IDF	SVM	0.9839	0.9637
	[56]	UCI ML Repo	Word Embedding	CNN-LSTM	0.9840	0.9800
	[134]	Enron + UCI ML Repo + English SMS	Tokenized Text	DistilBERT (FL)	0.9800	-
	[4]	UCI ML Repo	Heuristic	PART (C4.5)	0.9842	0.9840
	[157]	UCI ML Repo	Tokenized Text	RoBERTa	0.9984	0.9984
	[116]	Super Dataset	Word2Vec	SVM	0.9780	0.9900

(0.9984 accuracy and F1-score) [157]. Overall, deep learning models consistently outperform traditional ML methods, particularly transformer-based approaches such as BERT and its variants across most datasets. However, federated learning setups show lower robustness in some cases, with Remmide et al. [107] reporting the lowest accuracy (0.8878). For the Super dataset, Salman et al. [116] report 97.8% accuracy and 99% F1-score using SVM, while RoBERTa-base achieves 97.4% accuracy and 98% F1-score.

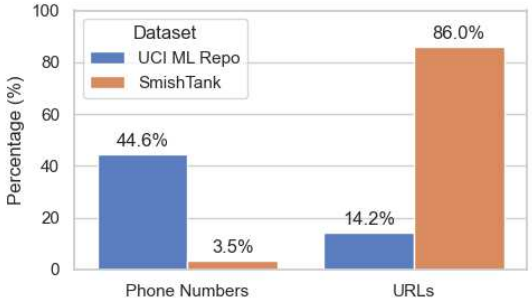


Fig. 7: Percentage of phone numbers and URLs inside spam/smish messages in UCI ML Repo (2011) and SmishTank (2024) datasets

6.3 Data Drift Analysis Within Existing English Datasets Based on Time and Geolocation of Data

Out of 14 candidate datasets, we select UCI SMS Spam Collection (2011) and SmishTank (2024) for data drift analysis to maximize temporal and geographical contrast (UK vs. US). This setup ensures observed lexical shifts reflect long-term evolution in smishing strategies rather than short-term variation. To measure drift without model dependence, we apply the Chi-square (χ^2) test of independence [75] on the top 500 features,

retaining only terms with $p < 0.05$ (95% confidence). This isolates statistically discriminative words across the two corpora. The 2011 UCI dataset is characterized by keywords {'landline', 'draw', 'win', 'won', 'free', 'awarded', 'box', 'tone', 'ur', 'txt', 'urgent', 'claim', 'mobile', 'send', 'text', 'cash', 'prize', 'po', 'guaranteed', 'nokia'}, reflecting premium-rate SMS scams prevalent in the UK. In contrast, the 2024 SmishTank dataset shows a shift toward credential and service-based phishing, dominated by {'safari', 'account', 'confirm', 'browser', 'delivered', 'package', 'link', 'tax', 'arrived', 'information', 'warehouse', 'refund', 'address', 'incomplete', 'open', 'provide', 'usps', 'click', 'irs', 'copy'}, indicating US-centric administrative and logistics impersonation (e.g., USPS, IRS). This highlights strong temporal and geographic influence on lexical patterns.

To ensure reproducibility, preprocessing removes URLs via regex, filters tokens with digits, and discards non-alphanumeric punctuation before tokenization. For χ^2 analysis, we use `CountVectorizer` with binary features (presence/absence). Each term is evaluated using a 2×2 contingency table (2011 vs. 2024, present vs. absent) with Yates' continuity correction applied.

6.4 Structural and Linguistic Evolution

Our structural analysis further reveals a profound shift in message composition between the 2011 and 2024 corpora. The average punctuation density increased from 4.10% in the legacy UCI dataset to 6.76% in the modern SmishTank collection ($p < 0.001$). This significant rise is primarily attributed to the transition from text-based social engineering to link-centric phishing, where URL delimiters (e.g., '/', '.', '?') contribute to a higher punctuation-to-character ratio. Conversely, a slight but statistically significant decrease in Shannon Entropy [165] was observed ($\mu_{2011_corpus} = 4.69$ vs. $\mu_{2024_corpus} = 4.64$; $p < 0.001$). This downward drift suggests a move away from the high-entropy, 'noisy' linguistic patterns of legacy SMS spam which is often characterized by non-standard abbreviations and irregular capitalization, toward a more professionalized and low-entropy style that mimics legitimate institutional correspondence. These structural comparisons were verified using an independent two-sample t-test, assuming unequal variances where applicable, with a significance threshold set at $\alpha = 0.05$. This finding is also backed by our manual analysis as shown in figure 7 with a radical inversion in call-to-action modalities where legacy 2011 messages relied heavily on telephony-based triggers (44.6% phone density), and 2024 smishing has almost entirely transitioned to a URL-centric architecture (85.98%). This shift signifies a technical evolution from premium-rate text fraud to digital credential harvesting via malicious web links. We have also generated word clouds in English texts for each datasets which are provided in the Github portal.

6.5 Critical Assessment of Dataset Limitations

Beyond the temporal drift analyzed above, our review identifies distinct structural limitations in the current smishing dataset landscape:

6.5.1 URL Decay:

A major challenge for researchers is URL Decay. Phishing domains are short-lived, often surviving less than 24 hours before being blocked [95]. Furthermore, sinkholing, where security vendors redirect malicious traffic to a safe server, can pollute datasets, as researchers may inadvertently analyze a sinkhole page, thinking it is the attack payload. In addition to a population study involving 50 randomly selected websites in total, restricted to a maximum of five per dataset, we conducted a secondary validation of the currently active domains to evaluate their status. Our findings from manual website queries, the Wayback Machine, and WHOIS lookup reveal that nearly all selected websites are now benign, despite their original inclusion in smishing or spam messages. In rare instances, WHOIS reports indicated that a domain was registered several years after the initial dataset was compiled; this discrepancy likely stems from different versions of the dataset aggregating messages from various time periods, with our current version containing all such entries. Given the highly ephemeral nature of these websites, we recommend that future researchers avoid relying on these datasets for URL or website analysis in their studies.

6.5.2 Anglocentric Bias and Multilingual Scarcity

As detailed in Section 7.4, there is a severe anglocentric bias in public repositories. While detection efforts exist for languages like Arabic [47], Vietnamese [10] and English-Chinese [23], high-quality public benchmarks remain predominantly English.

6.5.3 Benign-to-Malicious Imbalance

Public datasets rarely reflect real-world traffic distributions. In the real-world, the number of legitimate ones (ham) is more than smishing. However, repositories like SmishTank contain only smishing messages. This prevents the training of robust binary classifiers that can distinguish attacks from benign notifications (e.g., real delivery alerts vs. fake ones).

7 Research Gaps and Future Directions

With this comprehensive study of recent smishing research, we aim to systematically understand the current research trends, gaps in existing approaches or methods, and direct future researchers to enhance the trustworthiness and security of the SMS communication system. We envision the following major areas for future exploration: (A) smishing attacks, (B) user perception and susceptibility studies, (C) smishing defenses, (D) smishing datasets, and (E) Law enforcement, policy, and regulatory studies. We have listed the specific directions and open problems under these major areas below.

7.1 Smishing Attack Vectors

7.1.1 Abuse of GenAI for Smishing Threats

With the ongoing rapid development of LLMs and large foundation models, LLMs can be abused by attackers through jail-breaking and prompt injections to create innovative smishing campaigns that can themselves be a new attack vector in this area [129, 51, 112]. The ease of generating these diverse attack templates and propagating them through an anonymous bulk medium has made it more concerning to the defender. We envision conducting more in-depth analysis on various generative models to assess the potential abuse cases for smishing threats.

7.1.2 Automated Monitoring and Characterization of Smishing Attacks

Establishing systems and methods for *real-time threat intelligence* generation and sharing among parties (e.g., service providers, security agencies) can be beneficial. Collaborative efforts can enhance the industry's ability to respond promptly to emerging smishing campaigns and adapt countermeasures accordingly. Brands can also benefit from this kind of monitoring and tracking, where they can easily warn their users to safeguard them with awareness instructions. Research efforts like the SmishViz system/tool [120] for monitoring and tracking large smishing campaign-operations to dismantle coordinated attacks will be valuable. We envision that continuous and collective security hardening efforts, like large-scale campaign monitoring, are needed for a more secure SMS messaging ecosystem. Additionally, further usability research is required to understand how analysts can effectively use these kinds of tools for defense.

7.2 User Perception and Susceptibility Studies

In literature, even though there has been some steady progress in systematically understanding users' susceptibility and behaviors when encountering smishing messages, further continuous studies in this direction would help develop effective defense mechanisms that are relevant to a diverse population. For instance, as mentioned in the 2022 FBI Elder Fraud report [96], victims over the age of 60 experienced a total loss of $3.1 billion to fraud in 2022, reflecting an 84% overall increase in losses from 2021, with an average loss of $35,101 per victim. This highlights a critical vulnerability in older demographics. Specifically, large-scale user studies are needed to understand vulnerabilities in populations like users of niche/emerging communication platforms (e.g., Discord, Telegram, Threads, Twitch, Mastodon, BeReal), individuals in high-risk professions, and those with lower digital literacy. Hence, we recommend that researchers explore these open problems and ensure practical personalized defense mechanisms, including evaluating advanced user alert systems through comprehensive user studies among various population groups. For example, a recent work has successfully designed and evaluated visual trust indicators that help users rapidly assess message credibility [177]. Furthermore, we recommend that the community should investigate the psychological factors and sophistication of these attacks [147, 69, 68] and develop cyber social engineering kill chains [87] for various types of smishing threats. While existing studies have begun to explore psychological techniques (PTs) used in smishing messages [71], these efforts mainly focus on visible attack tactics rather than psychological factors that affect the user behavior. This creates a gap in understanding how and why users respond to smishing attacks. To address this limitation, future research should explore psychological awareness frameworks that model how users perceive and react to attacks, rather than only how attacks are constructed. In this context, developing a structured Cognitive Attack Taxonomy (CAT) [21] that maps observable attack techniques (e.g., impersonation, urgency cues) to underlying psychological factors can provide a more comprehensive representation of smishing behavior, enable more explainable detection, improve user susceptibility modeling, and support more realistic evaluation of smishing defenses. This investigation should be directly linked to the development of standardized taxonomies and labeling methods that can categorize smishing messages by their psychological techniques, which enable us to conduct more granular and realistic user susceptibility studies.

7.3 Smishing Defenses

7.3.1 Effective and Natural Reasoning-based Smishing Detection

[A] Advanced Multi-Modal Techniques for Feature Extraction and Detection: In the current literature, most of the SMS phishing detection models extract features based on human intelligence or NLP-based features. These limit the existing research to text only smishing problem. However, with the current developments of large image and video models, in the coming days, smishing messages with multimedia attachments will increase. Thus, to effectively defend against those threat vectors, we need to collect feature data from not only texts but also other multimedia items (e.g., images, audio, video). More research efforts to automatically extract these important multi-model features can be an interesting research avenue in the coming days.

[B] Explainability and Reasoning for Detection: Explainable AI is rapidly replacing traditional AI approaches due to its added benefits of transparency and trustworthiness. Like many other detection engineering problems, we envision a series of ongoing research where explainability would be an important aspect for detecting smishing and explaining why certain messages are smishing rather than benign. Additionally, XAI should be integrated even with the usage of LLMs to ensure users can trust the model outcomes and rely on a decision-making process with certain reasoning. Drawing from these concepts, Wang et al. [172] have demonstrated the value of LLM-based explanations in improving users' understanding and efficacy in detecting phishing SMS. Existing studies show that focusing on user-facing design is essential. For example, visual trust indicators successfully enhanced user confidence and decision-making because they clearly communicated security cues generated by the detection system [177]. This approach of providing evidence-based explanations has proven effective and highly usable for end-users. Future research should focus on practical ways to improve these explanations. This includes making them more personalized to individual users' needs and understanding, developing interactive tools that allow users to explore the AI's reasoning, and finding better ways to communicate the AI's confidence in its decisions, especially for ambiguous messages. The goal is to ensure users can reliably and securely depend on these explanations without becoming overly dependent on the AI [78, 80].

[C] Adversarial AI Robustness for Detection Models & Methods: Future research should also focus on comprehensive *adversarial robustness* testing for smishing detection models using ML, DL, NLP, smaller language models (SLMs) and LLM-based systems. Developing standardized evaluation metrics and benchmarks for assessing models' resilience against adversarial attacks will contribute to more secure and reliable defense solutions [118]. These directions of research are more applied here in the context of smishing defense, given that the field of adversarial robustness for ML and AI models is also advancing.

[D] Usage of LLMs and SLMs for Detection: LLMs and other large models can be utilized for defensive actions, such as reasoning-based detection of smishing messages within a user device. However, the main concerns with these larger models are that they need extensive computational capacity, which is generally better for cloud computation rather than in-device (i.e., mobile device) computation. Thus, researchers can explore smaller language models (SLMs) that are compressed to effectively run within the mobile device's storage constraints and evaluate their detection and reasoning performances [119, 118]. The in-device computation ensures that the user's private message data remains in the end-device and does not go to the cloud at any time for decision-making. This would be a promising research direction given the faster development of smaller language models that are intelligent and showing reasoning capabilities.

7.3.2 Architectural and Infrastructural Smishing Defenses

[A] Cyber Deception Strategies for Smishing Defense: SMS phishing messages mostly contain URLs through which users can interact. Oftentimes, these links can lead to mobile malware rather than just a phishing website. Thus, we envision that future research should be conducted on the feasibility of analyzing dynamic mobile malware [102] through file analysis attached to any spam or suspicious message links. We also envision that mobile cyber deception research [115, 114] can be an exciting and challenging direction to explore, where we can find malware types by analyzing the attached contents and send back fake data towards the attacker, as it expects to deceive them.

[B] User and Data Privacy in Smishing Defense: We also envision a rise in the adoption of privacy-preserving ML/AI approaches, such as federated learning, to address the concerns of privacy while analyzing private incoming messages [167, 107]. Moreover, the Homomorphic Encryption mechanism [144] can enable collaborative model training without compromising user privacy, and thus, researchers can explore the feasibility of such approaches to solve the privacy concerns.

[C] Blockchain for Smishing Defense: Exploring the integration of secure and attack-resilient proof-of-stake *blockchain* technology [14, 13] for message authentication and integrity verification is a promising research avenue. A decentralized and tamper-evident ledger can provide an additional layer of trust in SMS communications. Moreover, blockchain can be leveraged while sharing common cyber threats across multiple parties to enable trust and ensure tamper-resistant feeds [55].

7.4 Enhancing Coverage of Smishing Datasets

Expanding research to include more *diverse public datasets* while conducting cross-cultural and cross-lingual analyses can enhance the understanding of regional variations in smishing tactics. This approach contributes to the development of context-aware detection models with improved accuracy. Currently, the majority of public datasets are in the English language, with limited representation from other languages, despite some detection efforts in Arabic [47], Bengali [124, 26], Swahili [74], Portuguese [8], Hindi [127], and Chinese [23]. Thus, a critical research gap is the extensive lack of high-quality and labeled smishing datasets for non-English languages, particularly for widely spoken but underserved languages (e.g., Mandarin, Spanish, French, Persian, etc.). We believe smishing datasets from more languages might enrich the current detection research (alike multi-lingual SPAM detection research [105]) to assess the effectiveness of existing defense in these regions. Specifically, while existing literature discusses psychological factors and phishing techniques, current available public datasets lack a standardized method for automatically classifying and labeling messages based on these PFs/PTs. Thus, future research should focus on developing systematic classification and automated data labeling methods for messages. Moreover, current public datasets only contain text modality of data for messages, which may not be sufficient as attacks can happen through other modality of data such as multi-media (e.g., images, videos, audios in MMS). Specifically, with the deepfake video generation capabilities, in the future, we may experience more fake video lures by attackers attached to the messages to convince users. Thus, we recommend publishing datasets containing multi-modal data types (e.g., images, videos) for these messages. In addition, we believe future datasets with longer conversational data rather than one individual message, would give better a much basis for understanding attacker's strategies and psychological factors they target for successfully running various scam campaigns propagated through SMS messages. These types of heterogeneous datasets for smishing would enrich the future smishing attack understanding and defense research.

7.5 Law Enforcement and Regulation for Smishing

Currently, there is a lack of uniform and comprehensive research dedicated to the legal, regulatory, and law enforcement aspects of smishing. While automated monitoring and threat intelligence sharing are valuable for technical defense, as mentioned in Section VII-A-2, the community requires a deeper understanding and enforces more rigorous regulations on these categories of fraud and scams. We find a significant lack of academic research that focuses on the penalties and specific laws related to smishing. We suggest that future work in this particular area should follow several key thematic recommendations. First, one primary theme is the need for a systematic regulatory landscape analysis with a focus on existing federal and state laws related to mobile communication fraud and their specific applicability in smishing attacks, such as impersonation and/or financial theft. Future work can explore clearly identifying organizational structures and hierarchies that are mandated to fight and penalize these kinds of threats. Second, researchers can collect, review, and analyze previous court cases and legal precedents related to smishing with comprehensive case studies to provide concrete evidence of current legal effectiveness, jurisdictional challenges, and the enforcing penalties. Third, we recommend exploring international and comparative law research focused on smishing regulation around the globe. This would compare the different regulations and enforcement against smishing between regions (like Europe and Asia) by looking at diverse legal frameworks (e.g, GDPR, cybersecurity laws). This comparison would identify best practices to develop a stronger and more coordinated global response to find a uniform and coordinated action against smishing threats.

8 Conclusion

This paper comprehensively investigates the recent advancements in user susceptibility, attack landscape, available defenses, and datasets for smishing. Specifically, the paper evaluates the existing literature to systematically understand the current attack taxonomy and defense capabilities, including policy and regulations, to identify gaps in existing maps and provide future direction for defending againsts these threats. We also conduct data analysis on all the available datasets, report summary statistics, and provide a cleaned version of these data in a unified GitHub repository to aid reproducible future research. Overall, this review suggests

some promising future research directions to protect users from sharply rising smishing threats.

Data Availability All analysis code, results and curated datasets are made available via a public Github repo located at `https://github.com/MarazMia/SMISH_DT`.

Author Contributions

Mir Pritom contributed in conceptualization, methodology, study design, supervision, interpretation of results, writing original draft, reviewing, and editing. ***Seyed Sanjari*** contributed in conducting the Data collection, data curation, investigation, formal analysis, visualization, writing original draft, and editing (Section 3, Section 4, Section 5, Section 7). ***Maraz Mia*** contributed primarily in existing SMS dataset collection, processing, and analysis, visualization, and writing the original draft (Section 6). ***Ashfak Shibli*** contributed in creating the smishing threat landscape taxonomy and writing original draft (Section 4). ***Mostaq Hossain*** contributed in evaluating the smishing defense landscape and writing the original draft (Section 5). ***Muhmammad Ismail*** contributed in methodology, validation, manuscript organization, and Writing (reviewing and editing). ***Shouhuai Xu*** contributed in refinement of methodology, manuscript organization, validation, and writing (reviewing and editing).

Declarations

Conflict of interest. There are no potential conflict of interest with respect to the research, authorship, or publication of this paper.

Ethical approval. As this is a comprehensive review (SoK style) paper and all data were obtained from the publicly available domain, no ethical approval or permission was required for the conduct of this study.

Funding Statement. The authors received no research funding to conduct this study.

References

1. S. Agarwal, E. Harvey, E. Mariconti, G. Suarez-Tangil, M. Vasek, et al. 'hey mum, i dropped my phone down the toilet': Investigating hi mum and dad sms scams in the united kingdom. In *Usenix Security Symposium*, 2025.
2. M. M. Ahsan Pritom, K. M. Schweitzer, R. M. Bateman, M. Xu, and S. Xu. Characterizing the landscape of covid-19 themed cyberattacks and defenses. In *2020 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 1–6, 2020.
3. M. M. Ahsan Pritom, K. M. Schweitzer, R. M. Bateman, M. Xu, and S. Xu. Data-driven characterization and detection of covid-19 themed malicious websites. In *2020 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 1–6, 2020.
4. O. N. Akande, O. Gbenle, O. C. Abikoye, R. G. Jimoh, H. B. Akande, A. O. Balogun, and A. Fatokun. Smsprotect: An automatic smishing detection mobile application. *ICT Express*, 9(2):168–176, 2023.
5. M. R. Al Saidat, S. Y. Yerima, and K. Shaalan. Advancements of sms spam detection: A comprehensive survey of nlp and ml techniques. *Procedia Computer Science*, 244:248–259, 2024.
6. R. Alabdan. Phishing attacks survey: types, vectors, and technical approaches. *Future Internet*, 12(10):168, 2020.
7. A. Aleroud and L. Zhou. Phishing environments, techniques, and countermeasures: A survey. *Computers & Security*, 68:160–196, 2017.
8. F. D. Ali, H. L. Cardoso, R. Sousa-Silva, et al. Moz-smishing: A benchmark dataset for detecting mobile money frauds. In *Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025)*, pages 158–166, 2025.
9. T. A. Almeida, J. M. G. Hidalgo, and A. Yamakami. Contributions to the study of sms spam filtering: new collection and results. In *Proceedings of the 11th ACM Symposium on Document Engineering*, DocEng '11, page 259–262, New York, NY, USA, 2011. Association for Computing Machinery.
10. H. Q. Anh, P. T. Anh, P. S. Nguyen, and P. D. Hung. Federated learning for vietnamese sms spam detection using pre-trained phobert. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 254–264. Springer, 2024.
11. APWG. Phishing activity trends report 1st quarter 2024. https://apwg.org/trendsreports/, 2024. Accessed: August 16, 2024.
12. M. P. Bach, T. Kamenjarska, and B. Žmuk. Targets of phishing attacks: The bigger fish to fry. *Procedia Computer Science*, 204:448–455, 2022.
13. F. H. Bappy, T. Islam, K. Hasan, M. S. I. Sajid, and M. M. A. Pritom. Securing proof of stake blockchains: Leveraging multi-agent reinforcement learning for detecting and mitigating malicious nodes. *arXiv preprint arXiv:2407.20983*, 2024.

14. F. H. Bappy, T. Islam, T. S. Zaman, M. S. I. Sajid, and M. M. A. Pritom. Conchain: A scheme for contention-free and attack resilient blockchain. In *2024 IEEE 21st Consumer Communications and Networking Conference (CCNC)*, pages 1054–1055. IEEE, 2024.
15. E. Blancaflor, M. A. Romero, I. Nacu, and D. R. Golosinda. A case study on smishing: An assessment of threats against mobile devices. In *Proceedings of the 2023 9th International Conference on Computer Technology Applications*, pages 172–178, 2023.
16. E. B. Blancaflor, A. B. Alfonso, K. Banganay, G. D. Cruz, K. Fernandez, and S. Santos. Let's go phishing: A phishing awareness campaign using smishing, email phishing, and social media phishing tools. In *Proceedings of the international conference on industrial engineering and operations management*, 2021.
17. B. E. Boukari, A. Ravi, and M. Msahli. Machine learning detection for smishing frauds. In *2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC)*, pages 1–2. IEEE, 2021-1-9 2021.
18. California State Legislature. California business and professions code §§ 22948-22948.3. `https://law.justia.com/codes/california/2005/bpc/22948-22948.3.html`, 2005. Accessed on Justia's website.
19. California State Legislature. California civil code, title 1.81.5: California consumer privacy act of 2018. `https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5`, 2018. Accessed: October 8, 2025.
20. CAN SPAM ACT. CAN-SPAM Act: A Compliance Guide for Business — ftc.gov. `https://www.ftc.gov/business-guidance/resources/can-spam-act-compliance-guide-business`, 2023. [Accessed 31-08-2025].
21. CAT. The cognitive attack taxonomy (cat). https://cognitiveattacktaxonomy.org/, 2026. Accessed: April 29, 2026.
22. CCPA. CCPA Enforcement Case Examples — oag.ca.gov. `https://oag.ca.gov/privacy/ccpa`, 2018. [Accessed 29-07-2025].
23. T. Chen and M.-Y. Kan. Creating a live, public short message service corpus: the nus sms corpus. *Language Resources and Evaluation*, 47:299–335, 2013.
24. A. E. Chichwadia and N. Mpekoa. Detecting smishing and vishing attacks using machine learning. 2024.
25. R. Chithambaramani, M. Sankar, D. Manivannan, P. Sivaprakash, P. Mohan, et al. Sms guard: Android application to prevent smishing attacks. In *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*, pages 69–74. IEEE, 2024.
26. F. S. Chowdhury. Bengali SMS Spam Detection Dataset — kaggle.com. `https://www.kaggle.com/datasets/fahimvj/bengali-sms-spam-detection-dataset`, 2024. [Accessed 06-10-2025].
27. Cisco Talos Intelligence Group. Phishtank: Join the fight against phishing. `https://www.phishtank.com`.
28. City of London Police. Two people arrested in connection with investigation into homemade mobile antenna used to send thousands of smishing text messages to the public, June 2024. Press Release.
29. M. Clasen, F. Li, and D. Williams. Friend or foe: An investigation into recipient identification of sms-based phishing. In *Human Aspects of Information Security and Assurance: 15th IFIP WG 11.12 International Symposium, HAISA 2021, Virtual Event, July 7–9, 2021, Proceedings 15*, pages 148–163. Springer, 2021.
30. Cote, Alexandra. Tcpa fines, violations, and penalties for text messages. *SimpleTexting*, 2024.
31. J. Easton. Federal wire fraud explained: Key elements and risks. `https://leppardlaw.com/federal/fraud/federal-wire-fraud-explained-key-elements-and-risks/`, 2024.
32. M. Edwards, T. Morris, J. Chen, and J. Still. Smishing attack vector: Surveying end-user behavior, experience, and knowledge. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 67, pages 1911–1915. SAGE Publications Sage CA: Los Angeles, CA, 2023.
33. Elder Fraud and Abuse. OVC Elder Fraud & Abuse Initiatives — Office for Victims of Crime — ovc.ojp.gov. `https://ovc.ojp.gov/program/elder-fraud-abuse/ovc-elder-fraud-abuse-initiatives`, 2020. [Accessed 09-09-2025].
34. European Parliament and Council of the European Union. Directive (eu) 2022/2555 of the european parliament and of the council of 14 december 2022 on measures for a high common level of cybersecurity across the union, amending regulation (eu) no 910/2014 and directive (eu) 2018/1972, and repealing directive (eu) 2016/1148 (nis 2 directive). *Official Journal of the European Union*, L 333:80–152, 2022.
35. C. Faklaris, H. R. Lipford, and S. Tabassum. Preliminary results from a us demographic anal-

ysis of smish susceptibility. *arXiv preprint arXiv:2309.06322*, 2023.

36. FCC. About the FCC — fcc.gov. `https://www.fcc.gov/about-fcc/what-we-do`. [Accessed 01-09-2025].
37. F. C. C. (FCC). Rules and regulations implementing the telephone consumer protection act of 1991 (47 u.s.c. § 227). `https://www.fcc.gov/sites/default/files/tcpa-rules.pdf`, 1991. Accessed on October 7, 2025.
38. Federal Bureau of Investigation. Spoofing and phishing. `https://www.fbi.gov/how-we-can-help-you/scams-and-safety/common-frauds-and-scams/spoofing-and-phishing`.
39. Federal Communications Commission. Stop unwanted robocalls and texts. `https://www.fcc.gov/consumers/guides/stop-unwanted-robocalls-and-texts`. Published March 3, 2025.
40. Federal Communications Commission. Targeting and eliminating unlawful text messages report and order and further notice of proposed rulemaking. `https://www.fcc.gov/document/fcc-adopts-its-first-rules-focused-scam-texting-0`. Released March 17, 2023.
41. Federal Trade Commission. Ftc, justice department halt identity theft scam. `https://www.ftc.gov/news-events/news/press-releases/2004/03/ftc-justice-department-halt-identity-theft-scam`. Published March 2004.
42. FTC. About the FTC — ftc.gov. `https://www.ftc.gov/about-ftc`. [Accessed 01-09-2025].
43. FTC. Consumer protection data spotlight 2026. `https://www.ftc.gov/system/files/ftc_gov/pdf/social-media-spotlight.pdf`. Last accessed: 2026-09-23.
44. FTC. New ftc data show top text message scams of 2024, overall losses to text scams hit $470 million. `https://www.ftc.gov/news-events/news/press-releases/2025/04/`. [Accessed 07/22/2025].
45. FTC. Notices of Penalty Offenses — ftc.gov. `https://www.ftc.gov/enforcement/penalty-offenses`. [Accessed 31-08-2025].
46. A. Ghourabi. Sm-detector: A security model based on bert to detect smishing messages in mobile environments. *Concurrency and Computation*, 33(24), 2021.
47. A. Ghourabi, M. Mahmood, and Q. Alzubi. A hybrid cnn-lstm model for sms spam detection in arabic and english messages. future internet 12 (9): 156, 2020.
48. D. Goel, H. Ahmad, A. K. Jain, and N. K. Goel. Machine learning driven smishing detection framework for mobile security. *arXiv preprint arXiv:2412.09641*, 2024.
49. D. Goel and A. K. Jain. Smishing-classifier: a novel framework for detection of smishing attack in mobile environment. In *Smart and Innovative Trends in Next Generation Computing Technologies: Third International Conference, NGCT 2017, Dehradun, India, October 30-31, 2017, Revised Selected Papers, Part II 3*, pages 502–512. Springer, 2018.
50. Government of Singapore. Personal data protection act 2012, 2012. Act No. 26 of 2012, 2020 Revised Edition.
51. 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*, 11:80218–80245, 2023.
52. S. Gupta, P. Gupta, M. Ahamad, and P. Kumaraguru. Exploiting phone numbers and cross-application features in targeted mobile attacks. In *Proceedings of the 6th Workshop on Security and Privacy in Smartphones and Mobile Devices*, pages 73–82, 2016.
53. M. N. B. Haizam et al. Analysing the impact of smishing attack in public announcement system on mobile phone. *Procedia Computer Science*, 245:1165–1174, 2024.
54. D. S. Hapase and L. V. Patil. Telecommunication fraud resilient framework for efficient and accurate detection of sms phishing using artificial intelligence techniques. *Multimedia Tools and Applications*, pages 1–23, 2024.
55. S. He, E. Ficke, M. M. A. Pritom, H. Chen, Q. Tang, Q. Chen, M. Pendleton, L. Njilla, and S. Xu. Blockchain-based automated and robust cyber security management. *Journal of Parallel and Distributed Computing*, 163:62–82, 2022.
56. S. M. M. Hossain, J. A. Sumon, A. Sen, M. I. Alam, K. M. A. Kamal, H. Alqahtani, and I. H. Sarker. *Spam Filtering of Mobile SMS Using CNN–LSTM Based Deep Learning Model*, volume 420 of *Hybrid Intelligent Systems*, pages 106–116. Springer International Publishing, Cham, 2022.
57. A. K. Jain and B. B. Gupta. Feature based approach for detection of smishing messages in the mobile environment. *Journal of Information Technology Research (JITR)*, 12(2):17–35, 2019.
58. A. K. Jain, B. B. Gupta, K. Kaur, P. Bhutani, W. Alhalabi, and A. Almomani. A content and url analysis-based efficient approach to detect smishing sms in intelligent systems. *International Journal of Intelligent Systems*, 37(12):11117–11141, 2022.

59. H. K. Jovanovska. The Essential Guide to SMS Marketing Compliance in the US - GDPR Local — gdprlocal.com. `https://gdprlocal.com/the-essential-guide-to-sms-marketing-compliance-in-the-us/`, 2023. [Accessed 07-04-2025].
60. A. Kanaoka and T. Isohara. Beyond mobile devices: A cross-device solution for smishing detection and prevention. In *USENIX Symposium on Usable Privacy and Security*, pages 6–8, 2023.
61. H. E. Karhani, R. A. Jamal, Y. B. Samra, I. H. Elhajj, and A. Kayssi. Phishing and Smishing Detection Using Machine Learning. In *2023 IEEE International Conference on Cyber Security and Resilience (CSR)*, pages 206–211, Venice, Italy, July 2023. IEEE.
62. E. A. Katsarakes, M. Edwards, and J. D. Still. Where do users look when deciding if a text message is safe or malicious? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 68, pages 221–225. SAGE Publications Sage CA: Los Angeles, CA, 2024.
63. S. Kim, J. Park, H. Ahn, and Y. Lee. Detection of korean phishing messages using biased discriminant analysis under extreme class imbalance problem. *Information*, 15(5), 2024.
64. R. Kohilan, H. E. Warakagoda, T. T. Kitulgoda, N. Skandhakumar, and N. Kuruwitaarachchi. A machine learning-based approach for detecting smishing attacks at end-user level. In *2023 IEEE International Conference on e-Business Engineering (ICEBE)*, pages 149–154. IEEE, 2023.
65. P. Kumarasinghe, D. Dissanayake, P. Gamage, and G. U. Ganegoda. User behavior analysis in determining the vulnerable category of vishing and smishing. In *2023 5th International Conference on Advancements in Computing (ICAC)*, pages 35–40. IEEE, 2023.
66. Y. Lee and D. Han. Korsmishing explainer: A korean-centric llm-based framework for smishing detection and explanation generation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 642–656, 2024.
67. M. Liu, Y. Zhang, B. Liu, Z. Li, H. Duan, and D. Sun. Detecting and characterizing sms spearphishing attacks. In *Proceedings of the 37th Annual Computer Security Applications Conference*, pages 930–943, 2021.
68. T. Longtchi and S. Xu. Characterizing the evolution of psychological tactics and techniques exploited by malicious emails. In *Science of Cyber Security: 6th International Conference, SciSec 2024, Copenhagen, Denmark, August 14–16, 2024, Proceedings*, page 97–117, Berlin, Heidelberg, 2025. Springer-Verlag.
69. T. T. Longtchi, R. M. Rodriguez, L. Al-Shawaf, A. Atyabi, and S. Xu. Internet-based social engineering psychology, attacks, and defenses: A survey. *Proceedings of the IEEE*, 112(3):210–246, 2024.
70. S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
71. S. Ma, T. Ma, J. Liu, W. Song, Z. Liang, X. Xiao, and Y. Ye. Psyscam: a benchmark for psychological techniques in real-world scams. *arXiv preprint arXiv:2505.15017*, 2025.
72. T. Mahmud, M. A. H. Prince, M. H. Ali, M. S. Hossain, and K. Andersson. Enhancing cybersecurity: Hybrid deep learning approaches to smishing attack detection. *Systems*, 12(11):490, 2024.
73. Malwarebytes. Malwarebytes - Mobile Security — apps.apple.com. `https://apps.apple.com/us/app/malwarebytes-mobile-security/id1327105431?mt=8`. [Accessed 05-08-2025].
74. I. S. Mambina, J. D. Ndibwile, and K. F. Michael. Classifying swahili smishing attacks for mobile money users: A machine-learning approach. *IEEE Access*, 10:83061–83074, 2022.
75. M. L. McHugh. The chi-square test of independence. *Biochemia medica*, 23(2):143–149, 2013.
76. M. K. Mehmood, H. Arshad, M. Alawida, and A. Mehmood. Enhancing smishing detection: A deep learning approach for improved accuracy and reduced false positives. *IEEE Access*, 2024.
77. M. Mia, D. Derakhshan, and M. M. A. Pritom. Can features for phishing url detection be trusted across diverse datasets? a case study with explainable ai. In *Proceedings of the 11th International Conference on Networking, Systems, and Security*, pages 137–145, 2024.
78. M. Mia and M. M. A. Pritom. Explainable but vulnerable: Adversarial attacks on xai explanation in cybersecurity applications. In *2025 IEEE 7th International Conference on Trust, Privacy and Security in Intelligent Systems, and Applications (TPS-ISA)*, pages 101–110, 2025.
79. M. Mia, M. M. A. Pritom, T. Islam, and S. Xu. Characterizing event-themed malicious web campaigns: A case study on war-themed websites. *22nd Annual International Conference on Privacy, Security, and Trust (PST)*, 2025.

80. M. Mia, S. Roy, M. M. A. Pritom, and M. Gupta. Explainguard: A zero trust framework for post-hoc explanation integrity guarantees in blackbox xai models. *arXiv preprint arXiv:2608.21803*, 2026.
81. S. Mishra and D. Soni. A content-based approach for detecting smishing in mobile environment. In *Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur-India*, 2019.
82. S. Mishra and D. Soni. Smishing detector: A security model to detect smishing through sms content analysis and url behavior analysis. *Future Generation Computer Systems*, 108:803–815, 2020.
83. S. Mishra and D. Soni. Implementation of 'smishing detector': An efficient model for smishing detection using neural network. *SN COMPUT.SCI.*, 3(3), 2022.
84. S. Mishra and D. Soni. Sms phishing dataset for machine learning and pattern recognition. In *International Conference on Soft Computing and Pattern Recognition*, pages 597–604. Springer, 2022.
85. S. Mishra and D. Soni. Dsmishsms-a system to detect smishing sms. *Neural Comput and Applic*, 35(7):4975–4992, 2023.
86. R. Mohite and L. Ouarbya. Interpretable anomaly detection: A hybrid approach using rule-based and machine learning techniques. In *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, pages 1–10. IEEE, 2024.
87. R. Montanez Rodriguez and S. Xu. Cyber social engineering kill chain. In *International Conference on Science of Cyber Security*, pages 487–504. Springer, 2022.
88. A. Nahapetyan, K. Childs, S. Prasad, B. Reaves, and A. Kapravelos. Poster: Looking at the landscape of sms phishing. *The Network and Distributed System Security Symposium (NDSS), San Diego, CA*, pages 1–2, 2023.
89. A. Nahapetyan, S. Prasad, K. Childs, A. Oest, Y. Ladwig, A. Kapravelos, and B. Reaves. On sms phishing tactics and infrastructure. In *2024 IEEE Symposium on Security and Privacy (SP)*, pages 1–16, 2024.
90. National Legislative Assembly of Thailand. Personal data protection act, B.E. 2562 (2019). *Government Gazette*, 136(69 A):52–95, May 2019. Unofficial translation.
91. A. Neupane, M. L. Rahman, N. Saxena, and L. Hirshfield. A multi-modal neuro-physiological study of phishing detection and malware warnings. In *CCS'15: The 22nd ACM Conference on Computer and Communications Security*, pages 479–491. ACM, 2015-10-12 2015.
92. D. Njuguna, J. Kamau, and D. Kaburu. Model for mitigating smishing attacks on mobile platforms. In *2021 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, pages 1–6. IEEE, 2021.
93. Nomorobo . Nomorobo - Robocall Blocker — apps.apple.com. `https://apps.apple.com/us/app/nomorobo-robocall-blocker/id1134727588`. [Accessed 05-08-2025].
94. M. T. Nuruzzaman, C. Lee, and D. Choi. Independent and personal sms spam filtering. In *2011 IEEE 11th International Conference on Computer and Information Technology*, pages 429–435, 2011.
95. A. Oest, Y. Safaei, P. Zhang, B. Wardman, K. Tyers, Y. Shoshitaishvili, and A. Doupé. {PhishTime}: Continuous longitudinal measurement of the effectiveness of anti-phishing blacklists. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 379–396, 2020.
96. F. B. of Investigation. 2022 elder fraud report. `https://www.ic3.gov/AnnualReport/Reports/2022_IC3ElderFraudReport.pdf`, 2022. [Accessed 19-07-2025].
97. A. S. Onashoga, O. O. Abayomi-Alli, A. S. Sodiya, and D. A. Ojo. An adaptive and collaborative server-side sms spam filtering scheme using artificial immune system. *Information Security Journal: A Global Perspective*, 24(4-6):133–145, 2015.
98. B. Ousat, M. A. Tofighi, and A. Kharraz. An end-to-end analysis of covid-themed scams in the wild. In *Proceedings of the 2023 ACM Asia Conference on Computer and Communications Security*, ASIA CCS '23, page 509–523, New York, NY, USA, 2023. Association for Computing Machinery.
99. Parliament of the United Kingdom. Fraud act 2006, 2006. Chapter 35.
100. V. Philbrick. All You Need To Know About The California Consumer Privacy Act (CCPA) — Yotpo — yotpo.com. `https://www.yotpo.com/blog/all-you-need-to-know-about-the-california-consumer-privacy-act-ccpa/`, 2024. [Accessed 07-04-2025].
101. Puig, Alvaro. Scammers hide harmful links in QR codes to steal your information — consumer.ftc.gov. `https://consumer.ftc.gov/consumer-alerts/2023/12/scammers-hide-harmful-links-qr-codes-steal-your-information`, 2023. [Accessed 01-08-2025].
102. M. A. Qbeitah and M. Aldwairi. Dynamic malware analysis of phishing emails. In *2018 9th In-*

ternational Conference on Information and Communication Systems (ICICS), pages 18–24. IEEE, 2018.

103. M. Rahaman, N. Cajes, B. B. Gupta, K. T. Chui, and N. Nedjah. Defending against smishing attacks: State-of-the-art techniques, challenges, limitations, and future directions. *Computer Networks*, page 111758, 2025.
104. M. L. Rahman, D. Timko, H. Wali, and A. Neupane. Users really do respond to smishing. In *Proceedings of the Thirteenth ACM Conference on Data and Application Security and Privacy*, CODASPY '23, page 49–60, New York, NY, USA, 2023. Association for Computing Machinery.
105. E. Ramanujam, K. Shankar, and A. Sharma. A review on artificial intelligence techniques for multilingual sms spam detection. In *International Conference on Computer Vision, High-Performance Computing, Smart Devices, and Networks*, pages 525–536. Springer, 2022.
106. J. W. Ratcliff, D. E. Metzener, et al. Pattern matching: The gestalt approach. *Dr. Dobb's Journal*, 13(7):46, 1988.
107. M. A. Remmide, F. Boumahdi, B. Ilhem, and N. Boustia. A privacy-preserving approach for detecting smishing attacks using federated deep learning. *International Journal of Information Technology*, pages 1–7, 2024.
108. L. Ribeiro, I. S. Guedes, and C. S. Cardoso. Which factors predict susceptibility to phishing? an empirical study. *Computers & Security*, 136:103558, 2024.
109. M. T. Ribeiro, S. Singh, and C. Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.
110. Robokiller. Robokiller: Spam Call Blocker — apps.apple.com. `https://apps.apple.com/us/app/robokiller-block-spam-calls/id1022831885`. [Accessed 05-08-2025].
111. J. D. Rose et al. Next-gen phishing detection system based on federated learning integrated cnn-lstm for sms communication. In *2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, pages 367–372. IEEE, 2024.
112. S. S. Roy, P. Thota, K. V. Naragam, and S. Nilizadeh. From chatbots to phishbots?: Phishing scam generation in commercial large language models. In *2024 IEEE Symposium on Security and Privacy (SP)*, pages 221–221. IEEE Computer Society, 2024.
113. A. Safi and S. Singh. A systematic literature review on phishing website detection techniques. *Journal of King Saud University-Computer and Information Sciences*, 35(2):590–611, 2023.
114. M. S. I. Sajid, J. Wei, B. Abdeen, E. Al-Shaer, M. M. Islam, W. Diong, and L. Khan. Soda: A system for cyber deception orchestration and automation. In *Proceedings of the 37th Annual Computer Security Applications Conference*, ACSAC '21, page 675–689, New York, NY, USA, 2021. Association for Computing Machinery.
115. M. S. I. Sajid, J. Wei, M. R. Alam, E. Aghaei, and E. Al-Shaer. Dodgetron: Towards autonomous cyber deception using dynamic hybrid analysis of malware. In *2020 IEEE Conference on Communications and Network Security (CNS)*, pages 1–9, 2020.
116. M. Salman, M. Ikram, and M. A. Kaafar. Investigating evasive techniques in sms spam filtering: A comparative analysis of machine learning models. *IEEE Access*, 2024.
117. S. R. A. Samad, P. Ganesan, J. Rajasekaran, M. Radhakrishnan, H. Ammaippan, and V. Ramamurthy. Smishguard: Leveraging machine learning and natural language processing for smishing detection. *International Journal of Advanced Computer Science & Applications*, 14(11), 2023.
118. S. M. Sanjari, M. M. A. Pritom, and M. Gupta. On the robustness of small language models for smishing detection against adversarial text perturbations. In *2026 IEEE 8th International Conference on Trust, Privacy and Security in Intelligent Systems, and Applications (TPS-ISA)*, pages 1–10, 2026.
119. S. M. Sanjari, J. Roberts, and M. M. A. Pritom. Poster: A few-shot learning method for sms phishing detection and explanation using slms. In *2025 IEEE Symposium on Security and Privacy (SP)*, San Francisco, CA, USA, May 2025. IEEE.
120. S. M. Sanjari, A. M. Shibli, M. Mia, M. Gupta, and M. M. A. Pritom. Smishviz: Towards a graph-based visualization system for monitoring and characterizing ongoing smishing threats. In *Proceedings of the Fifteenth ACM Conference on Data and Application Security and Privacy*, CODASPY '25, page 257–268, New York, NY, USA, 2025. Association for Computing Machinery.
121. D. M. Sarno and J. Black. Who gets caught in the web of lies?: Understanding susceptibility to phishing emails, fake news headlines, and scam text messages. *Human Factors*, 66(6):1742–1753, 2024.

122. J. W. Seo, J. S. Lee, H. Kim, J. Lee, S. Han, J. Cho, and C.-H. Lee. On-device smishing classifier resistant to text evasion attack. *IEEE Access*, 2024.
123. H. Shahriar, T. Klintic, and V. Clincy. Mobile phishing attacks and mitigation techniques. *Journal of Information Security*, 6(3):206–212, 2015.
124. M. F. Shahriyar and G. Tanbhir. Bangalabarta : A spam / smishing sms dataset bangla, 2025. Version 2.
125. N. Shamy, , W. Xie, and C. Zhong. Smishing: Exploring how different persuasion techniques influence users emotions, cognitions, and identification accuracy. In *Lecture Notes in Information Systems and Organisation (NeuroIS)*, volume 66. Springer, 2024.
126. M. J. A. Shanto, E. A. Tuli, R. Akter, D.-S. Kim, and T. Jun. Federated learning empowered spam message detection for multilingual short message service (sms). *KICS Summer Conference 2023*, pages 1310–1311, 2023.
127. S. Shashank. GitHub - shshnk158/Multilingual-SMS-spam-detection-using-RNN — github.com. `https://github.com/shshnk158/Multilingual-SMS-spam-detection-using-RNN/tree/master/`, 2018. [Accessed 06-10-2025].
128. B. D. Shendkar, P. R. Chandre, S. S. Madachane, N. Kulkarni, and S. Deshmukh. Enhancing phishing attack detection using explainable ai: Trends and innovations. *ASEAN Journal on Science and Technology for Development*, 42(1):8, 2024.
129. A. M. Shibli, M. M. A. Pritom, and M. Gupta. Abusegpt: Abuse of generative ai chatbots to create smishing campaigns. In *IEEE 12th International Symposium on Digital Forensics and Security (ISDFS)*, 2024.
130. H. S. Shim, H. Park, K. Lee, J.-S. Park, and S. Kang. A persuasion-based prompt learning approach to improve smishing detection through data augmentation. *arXiv preprint arXiv:2411.02403*, 2024.
131. A. Shinde, E. Q. Shahra, S. Basurra, F. Saeed, A. A. AlSewari, and W. A. Jabbar. Sms scam detection application based on optical character recognition for image data using unsupervised and deep semi-supervised learning. *Sensors*, 24(18):6084, 2024.
132. H. Siadati, T. Nguyen, P. Gupta, M. Jakobsson, and N. Memon. Mind your smses: Mitigating social engineering in second factor authentication. *Computers & Security*, 65:14–28, 2017.
133. M. A. Siddiqi, W. Pak, and M. A. Siddiqi. A study on the psychology of social engineering-based cyberattacks and existing countermeasures. *Applied Sciences*, 12(12):6042, 2022.
134. J. Sidhpura, P. Shah, R. Veerkhare, and A. Godbole. *FedSpam: Privacy Preserving SMS Spam Prediction*, volume 1793 of *Neural Information Processing*, pages 52–63. Springer Nature Singapore, Singapore, 2023.
135. SimpleTexting. SimpleTexting — apps.apple.com. `https://apps.apple.com/us/app/simpletexting/id1214066522`. [Accessed 05-08-2025].
136. J. Singh. SMS Data Privacy Under GDPR: Best Practices Unveiled — idtexpress.com, 2024. [Accessed 07-04-2025].
137. SlickText. SlickText - Apps on Google Play — play.google.com. `https://play.google.com/store/apps/details?id=io.gonative.android.njldwy&hl=en`. [Accessed 05-08-2025].
138. G. Sonowal. Detecting phishing sms based on multiple correlation algorithms. *SN computer science*, 1(6):361, 2020.
139. G. Sood. virustotal: R client for the virustotal api. `https://www.virustotal.com/`, 2021. R package version 0.2.2.
140. SpamHound. SpamHound SMS Spam Filter — apps.apple.com. `https://apps.apple.com/us/app/spamhound-sms-spam-filter/id1263185195`. [Accessed 05-08-2025].
141. A. Steele. Smishing.
142. W. J. Stuart and A. McNeil. The Growing Threat of MMS Scam Messages — Proofpoint US. `https://www.proofpoint.com/us/blog/email-and-cloud-threats/growing-threat-mms-scam-messages`, 2024. [Accessed 01-08-2025].
143. N. Sugunaraj, A. R. Ramchandra, and P. Ranganathan. Cyber fraud economics, scam types, and potential measures to protect us seniors: A short review. In *2022 IEEE International Conference on Electro Information Technology (eIT)*, pages 623–627. IEEE, 2022.
144. X. Sun, P. Zhang, J. K. Liu, J. Yu, and W. Xie. Private machine learning classification based on fully homomorphic encryption. *IEEE Transactions on Emerging Topics in Computing*, 8(2):352–364, 2018.
145. S. Tabassum, C. Faklaris, and H. R. Lipford. What drives {SMiShing} susceptibility? a {US}. interview study of how and why mobile phone users judge text messages to be real or fake. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*, pages 393–411, 2024.
146. S. Tang, X. Mi, Y. Li, X. Wang, and K. Chen. Clues in tweets: Twitter-guided discovery and analysis of sms spam. In *Proceedings of the*

2022 ACM SIGSAC Conference on Computer and Communications Security, CCS '22, page 2751–2764, New York, NY, USA, 2022. Association for Computing Machinery.

147. T. Tangie Longtchi, R. Montañez Rodriguez, K. Gwartney, E. Ear, D. P. Azari, C. P. Kelley, and S. Xu. Quantifying psychological sophistication of malicious emails. *IEEE Access*, 12:187512–187535, 2024.
148. Text-Em-All. Text-Em-All — apps.apple.com. `https://apps.apple.com/us/app/text-em-all/id1522359622`. [Accessed 05-08-2025].
149. Textkiller. Textkiller - Spam Text Blocker — apps.apple.com. `https://apps.apple.com/us/app/textkiller-spam-text-blocker/id1514005355`. [Accessed 05-08-2025].
150. TextSpot. TextSpot — Mass Text Messaging Software - Start for Free — textspot.io. `https://textspot.io/`. [Accessed 05-08-2025].
151. D. Timko, D. H. Castillo, and M. L. Rahman. Understanding influences on sms phishing detection: User behavior, demographics, and message attributes.
152. D. Timko, D. H. Castillo, and M. L. Rahman. A quantitative study of sms phishing detection. *arXiv preprint arXiv:2311.06911*, 2023.
153. D. Timko and M. L. Rahman. Commercial anti-smishing tools and their comparative effectiveness against modern threats. In *WiSec '23: 16th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, pages 1–12. ACM, 2023-05-29 2023.
154. D. Timko and M. L. Rahman. Smishing dataset i: Phishing sms dataset from smishtank. com. *arXiv preprint arXiv:2402.18430*, 2024.
155. A. Tsunoda. Demonstrating spoofability of an originating number when sending an sms using smpp. *Digital Threats: Research and Practice*, 5(1):1–13, 2024.
156. Twilio. Communications APIs with AI and data for SMS, Voice, Email — Twilio — twilio.com. `https://www.twilio.com/en-us`. [Accessed 05-08-2025].
157. M. A. Uddin, M. N. Islam, L. Maglaras, H. Janicke, and I. H. Sarker. Explainabledetector: Exploring transformer-based language modeling approach for sms spam detection with explainability analysis. *arXiv preprint arXiv:2405.08026*, 2024.
158. R. E. Ulfath, I. H. Sarker, M. J. M. Chowdhury, and M. Hammoudeh. Detecting smishing attacks using feature extraction and classification techniques. In *Proceedings of the International Conference on Big Data, IoT, and Machine Learning: BIM 2021*, pages 677–689. Springer, 2022.
159. United States Congress. Fraud by wire, radio, or television - title 18, u.s. code, § 1343. `https://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title18-section1343&num=0&edition=prelim`, 2008. Reflects text as last amended by Public Law 110-179 in Jan. 2008. Accessed: October 8, 2025.
160. United States Department of Justice. 1512. prohibited acts—18 u.s.c. 1028. `https://www.justice.gov/archives/jm/criminal-resource-manual-1512-prohibited-acts-18-usc-1028`, 2018. Archived content from the U.S. Department of Justice website.
161. United States Department of Justice. 941. 18 u.s.c. 1343—elements of wire fraud. `https://www.justice.gov/archives/jm/criminal-resource-manual-941-18-usc-1343-elements-wire-fraud`, 2018. Archived content from the U.S. Department of Justice website.
162. United States Department of Justice. Identity theft and identity fraud. `https://www.justice.gov/criminal/criminal-fraud/identity-theft/identity-theft-and-identity-fraud`, 2023.
163. United States House of Representatives. Title 18, united states code, section 1028: Fraud and related activity in connection with identification documents, authentication features, and information. `https://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title18-section1028&num=0&edition=prelim`, 1998. Accessed: October 8, 2025.
164. U.S. Attorney's Office, Central District of California. 5 defendants charged federally with running scheme that targeted victim companies via phishing text messages, Nov. 2024. Press Release 24-289.
165. S. Vajapeyam. Understanding shannon's entropy metric for information. *arXiv preprint arXiv:1405.2061*, 2014.
166. M. Vangelis, A. Ion, and P. Geogios. Spam filtering with naive bayes-which naive bayes? In *Third conference on email and anti-spam (CEAS)*, 2006.
167. S. Vats, S. Shastri, and S. Mehta. Federated learning for sms spam detection: A privacy-focused approach. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–5. IEEE, 2024.
168. M. Vijayalakshmi, S. Mercy Shalinie, M. H. Yang, and R. M. U. Web phishing detection techniques: a survey on the state-of-the-art, taxonomy and future directions. *Iet Networks*, 9(5):235–246, 2020.

169. Virginia General Assembly. Code of virginia, title 18.2, chapter 5, article 7.1. https://law.lis.virginia.gov/vacodefull/title18.2/chapter5/article7.1/, 1984. Accessed via the Legislative Information System (LIS) website,[Accessed 10-08-2025].
170. J. Walker. GDPR and SMS Marketing - What You Need to Know in 2023. https://www.courier.com/guides/gdpr-sms-marketing. [Accessed 07-04-2025].
171. C. Wang, Z. Jia, H. Benkraouda, C. Zevnik, N. Heuermann, R. Foulger, J. A. Handler, and G. Wang. Verisms: A message verification system for inclusive patient outreach against phishing attacks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2024.
172. Y. Wang, H. Zhai, C. Wang, Q. Hao, N. A. Cohen, R. Foulger, J. A. Handler, and G. Wang. Can you walk me through it? explainable sms phishing detection using llm-based agents.
173. Werksman Jackson & Quinn LLP. All about california's anti-phishing act of 2005. https://werksmanjackson.com/blog/all-about-californias-anti-phishing-act-of-2005/, 2024.
174. R. Yan. Awareness, knowledge, and ability of mobile security among young mobile phone users. In *Research Anthology on Securing Mobile Technologies and Applications*, pages 517–526. IGI Global, 2021.
175. R. E. Yoro, F. O. Aghware, M. I. Akazue, A. E. Ibor, and A. A. Ojugo. Evidence of personality traits on phishing attack menace among selected university undergraduates in nigerian. *International Journal of Electrical and Computer Engineering*, 13(2):1943, 2023.
176. R. E. Yoro, F. O. Aghware, B. O. Malasowe, O. Nwankwo, and A. A. Ojugo. Assessing contributor features to phishing susceptibility amongst students of petroleum resources varsity in nigeria. *Int. J. Electr. Comput. Eng*, 13(2):1922, 2023.
177. N. Zare, C. Faklaris, S. Tabassum, and H. R. Lipford. Improving mobile security with visual trust indicators for smishing detection. In *2025 IEEE World AI IoT Congress (AIIoT)*, pages 0212–0221. IEEE, 2025.
178. S. Zieliński. Evolving threats, emerging laws: Poland's 2023 answer to the smishing challenge. *Computer Law & Security Review*, 54:106013, 2024.