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.

Artificial Intelligence-Based Illegal Online Gambling Detection: A Systematic Literature Review of Adversarial Evasion and Countermeasure Methods

Abdul Azzam Ajhari
University of Indonesia
Rizal Fathoni Aji
rizal@cs.ui.ac.id
University of Indonesia
Aprinaldi Jasa Mantau
University of Indonesia
Mahardhika Pratama
University of South Australia
Wisnu Jatmiko
University of Indonesia

Research Article

Keywords: artificial intelligence, illegal online gambling, AI-based detection, adversarial evasion, machine learning, deep learning, multimodal fusion, systematic literature review

Posted Date: June 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9845547/v1

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

Additional Declarations: No competing interests reported.

Artificial Intelligence-Based Illegal Online Gambling Detection: A Systematic Literature Review of Adversarial Evasion and Countermeasure Methods

Abdul Azzam Ajhari[1,2†], Rizal Fathoni Aji[1*†], Aprinaldi Jasa Mantau[1†], Wisnu Jatmiko[1†], Mahardhika Pratama[1†]

[1*]Faculty of Computer Science, Universitas Indonesia, Depok City, 16424, Indonesia.
[2]The National Cyber and Crypto Agency (BSSN) of the Republic of Indonesia.
[3]STEM, University of South Australia, Adelaide, SA, 5095, Australia.

*Corresponding author(s). E-mail(s): rizal@cs.ui.ac.id;
Contributing authors: abdul.azzam51@ui.ac.id; aprinaldi12@cs.ui.ac.id; wisnuj@cs.ui.ac.id; dhika.pratama@adelaide.edu.au;
†These authors contributed equally to this work.

Abstract

Artificial intelligence (AI)-based illegal online gambling (IOG) detection systems can support cybersecurity practitioners, regulators, and digital platforms in identifying prohibited gambling activities distributed across websites, social media, mobile applications, hidden pages, payment infrastructures, and blockchain environments. However, adaptive adversarial behaviours, including textual obfuscation, visual concealment, domain rotation, traffic camouflage, mobile-app masking, payment obfuscation, and blockchain role hiding, challenge existing AI and hybrid detection pipelines. This systematic literature review examines how AI-enabled detection methods, adversarial evasion techniques, and computational countermeasures have been addressed in IOG detection systems. It synthesises 53 primary studies retrieved from Scopus, ScienceDirect, IEEE Xplore, and ACM Digital Library using a PICOC framework, PRISMA-style selection, predefined eligibility criteria, quality assessment, and structured data

extraction. The studies are categorised by AI model family, evidence surface, evasion mechanism, feature representation, computational architecture, dataset evidence, evaluation metric, baseline comparison, robustness evidence, scalability evidence, and reported limitation. The novelty lies in a modality–evasion–AI countermeasure synthesis, adversarial evasion taxonomy, and evidence-maturity assessment for resilient, scalable, and operationally deployable AI-based IOG detection systems.

Keywords: artificial intelligence, illegal online gambling, AI-based detection, adversarial evasion, machine learning, deep learning, multimodal fusion, systematic literature review

1 Introduction

Illegal online gambling (IOG) is increasingly an artificial intelligence (AI) and information-security problem rather than only a regulatory or public-health issue. IOG services abuse legitimate digital infrastructures, including websites, social media platforms, mobile application ecosystems, payment intermediaries, newly registered domains, compromised servers, and blockchain networks. These heterogeneous traces require AI-based detection systems that can learn from textual, visual, behavioural, network, mobile, financial, and graph-structured evidence. In technical terms, the problem involves adversarial misuse of digital platforms, abuse-infrastructure detection, fraud and financial-flow monitoring, content-security enforcement, web and network security, digital forensics, and operational response. Regulated gambling research links responsible-gaming duties to neural-network risk prediction Buttigieg et al. (2022), longitudinal player tracking and early-risk indicators Auer and Griffiths (2024), and account-based behavioral monitoring Haefeli et al. (2011). However, unlicensed or prohibited gambling ecosystems introduce additional AI challenges because they combine adaptive content changes with covert distribution, payment-service abuse, and rapid infrastructure churn Fiedler (2018); Gao et al. (2021); Guo et al. (2026).

The detection surface of IOG is broad and heterogeneous, making it suitable for an AI-oriented systematic review. Website and multimodal studies show that gambling pages can be detected from OCR-derived screenshot text in co-training Wang et al. (2022a), from visual–textual decision fusion Chen et al. (2020), and from hybrid multimodal data fusion over screenshots and semantic features Wang et al. (2022b). Social-media studies examine semi-supervised IndoBERT detection of Indonesian YouTube promotions Santika et al. (2025), IndoBERT-based performance on gambling-promotion comments Kamdan et al. (2025), and multilingual/Indonesian BERT detection of disguised promotional comments Prakoso et al. (2025). Mobile and traffic studies cover UI-transition graph similarity for underground economy apps Chen et al. (2023), mobile IOG service detection from SMS/MMS evidence Min et al. (2022), POST-message behaviour learned from gambling website templates Feng et al. (2025), and honeypot-based SMS reporting of IOG URLs Min et al. (2021). Blockchain and financial-flow studies show that IOG detection can involve smart-contract and

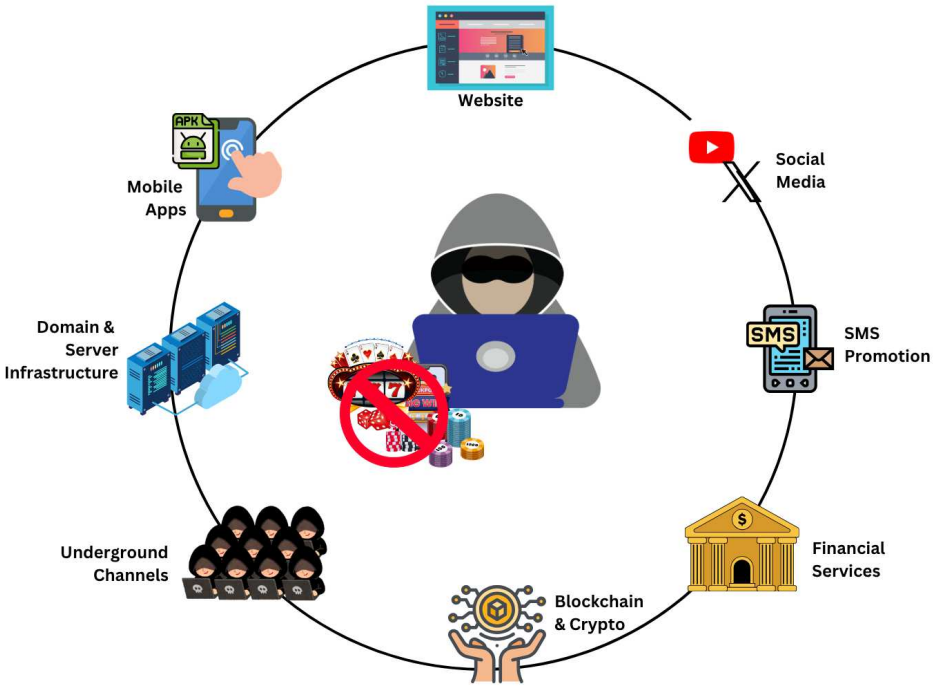


Fig. 1 AI-oriented multichannel ecosystem of illegal online gambling activities across websites, mobile applications, social media, SMS promotion, domain and server infrastructure, underground channels, financial services, and blockchain or crypto-based environments.

address classification Huang et al. (2022), hidden participant-role inference in crypto casinos Wang et al. (2025b), and AI-assisted betting-anomaly detection for illegal-gambling prevention Kim et al. (2024). This breadth makes IOG detection relevant to *Discover Artificial Intelligence* because it connects AI algorithms, multimodal representation learning, graph analytics, adversarial robustness, and real-world cyber-abuse detection.

Figure 1 summarizes the multichannel nature of IOG operations. The figure is included in the Introduction because it motivates the core AI argument of this review: a detector that learns from only one channel, such as keywords or URLs, may miss adversarial activity that has moved into images, app interfaces, traffic templates, hidden links, payment relationships, or blockchain graphs. This motivates AI-based multichannel representation learning rather than isolated keyword or blacklist matching.

A major challenge is adversarial evasion. IOG actors can mutate text through font or emoji manipulation in YouTube promotions Anderson et al. (2026), obfuscated language and special characters in Shorts comments Santosa et al. (2025), and slang or semantic-context evasion in social-media advertisements Muzakir et al. (2025a). They can also distribute gambling cues through OCR/ASR-derived multimodal advertisements Muzakir et al. (2025b), modality-level camouflage in illegal websites Wen et al. (2024), and visual gambling imagery or heuristic-visual detection settings Sánchez

et al. (2026). At the infrastructure layer, evasion includes newly registered domain churn Guo et al. (2026), JavaScript-driven proactive web defacement and cloaking Zhao (2023), and hidden-link threats that affect crawler or LLM-based inspection Xiong et al. (2025). At the mobile and blockchain layers, gambling services may use covert mobile-app distribution and payment-service abuse Gao et al. (2021), hidden delegatee roles in crypto casinos Wang et al. (2025b), and WebView-based concealment or missing malware payloads in illicit apps Lu et al. (2025). These patterns show that IOG detection must be evaluated as an adversarial security problem, not merely as a static classification problem.

Although many studies report high accuracy, precision, recall, or F1-score, such results do not automatically imply adversarial resilience. A system may perform well on a fixed dataset but fail when attackers change text, redesign visual content, rotate domains, alter traffic templates, repackage applications, or move financial activity across intermediaries. This distinction between *task-specific detectability* and *adversarial resilience* is central to information-security evaluation. It is also the main reason this review synthesizes not only detection methods, but also evasion evidence, robustness evidence, scalability evidence, dataset transparency, and real-world applicability.

This SLR contributes to the body of knowledge in four ways. First, it provides an AI-oriented adversarial evasion taxonomy for IOG across textual, visual, web, mobile, operational, financial-flow, and blockchain surfaces. Second, it synthesizes AI and computational countermeasure architectures, including machine learning, deep learning, transformer models, multimodal fusion, graph learning, anomaly detection, honeypot workflows, filtering, intrusion detection, and forensic monitoring. Third, it distinguishes ordinary AI model performance from adversarial robustness, scalability, reproducibility, and operational readiness through an evidence-maturity analysis. Fourth, it derives a research agenda for resilient AI-based IOG detection systems, emphasizing adversarial validation, cross-platform generalization, transparent datasets, abuse-infrastructure monitoring, human–AI operational response, and adaptive model updating.

The manuscript is organized as follows. Section 2 reviews prior gambling-related reviews, AI-based detection studies, adjacent technical studies, information-security foundations, the scope of this SLR, and the analytical taxonomy. Section 3 describes the review protocol, search process, study selection, quality assessment, data extraction, and evidence-maturity assessment. Section 4 reports the synthesis for RQ1–RQ4 and derives the proposed resilient AI-based detection framework. Section 5 discusses the findings from an information-security perspective, including implications and limitations. Section 6 concludes the review. Appendix A compares related work, Appendix B provides the technical evidence matrix, Appendix C lists the primary studies with matched citation keys, and Appendix D reports published and repository-linked datasets. For ease of reference, Table 1 summarizes the common abbreviations used throughout this manuscript.

Table 1 List of common abbreviations used in the manuscript.

Abbreviation	Description
IOG	Illegal online gambling
AI	Artificial intelligence
ML	Machine learning
DL	Deep learning
LLM	Large language model
OGDS	Online gambling detection system
SLR	Systematic literature review
PICOC	Population, Intervention, Comparison, Outcome, and Context
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
OCR	Optical character recognition
NLP	Natural language processing
IDS	Intrusion detection system
APK	Android application package
GNN	Graph neural network
EMS	Evidence maturity score

2 Related work

This section is expanded to separate the background material from the Introduction and to position the review more clearly within artificial intelligence and information security. Prior studies relevant to this review can be grouped into four streams: gambling-oriented reviews and behavioral studies, AI-based and computational detection studies, adjacent information-security abuse studies, and methodological SLR guidance. These streams are useful, but none of them provides a dedicated synthesis of how AI-based models, adversarial evasion techniques, and countermeasure methods interact in IOG detection systems.

2.1 Gambling-oriented reviews and behavioral monitoring

Existing gambling research explains why online gambling requires monitoring, regulation, and early-risk identification. Studies on problematic gambling and high-risk user behavior have used neural-network risk prediction Buttigieg et al. (2022), account-based tracking and longitudinal early-risk indicators Auer and Griffiths (2024); Haefeli et al. (2011), gambling-behavior detection or monitoring evidence Andersson et al. (2025), lottery-oriented segmentation Challet-Bouju et al. (2020), and database-level feature extraction Stechschulte et al. (2024). These studies are important because they show that gambling behavior can be represented computationally. However, their main emphasis is responsible gambling, behavioral risk, or player protection rather than adversarial evasion by illegal operators.

Other studies clarify the social, regulatory, and conceptual boundaries of online gambling. Regulation-oriented work by Fiedler Fiedler (2018) explains why online gambling supervision depends on jurisdiction and enforcement capacity. Studies on gaming–gambling convergence show how gambling-like digital experiences blur entertainment and wagering boundaries Teichert et al. (2017), while microtransaction

research highlights additional enforcement ambiguity in game-like environments Gibson et al. (2022). The public-health relevance of gambling harms is further supported by prevalence-oriented review evidence Tran et al. (2024). These works provide domain context, but they do not synthesize information-security mechanisms such as domain rotation, hidden-link abuse, covert app distribution, defacement injection, blockchain role concealment, or adversarial robustness testing.

Appendix A and Table E1 compare the present review with existing review-oriented studies. The comparison shows that prior reviews explain legal, behavioral, clinical, or public-health aspects of online gambling, but they do not connect IOG evasion techniques to feature representations, model architectures, evaluation protocols, robustness evidence, and operational countermeasures. This is the gap addressed by the present SLR.

2.2 AI-based and computational detection studies for IOG-related artifacts

A second stream consists of AI-based and computational studies that detect gambling-related artifacts on specific platforms. Textual and social-media studies detect semi-supervised YouTube gambling promotions Santika et al. (2025), IndoBERT-based YouTube comment spam Kamdan et al. (2025), disguised promotional comments with Indonesian and multilingual BERT Prakoso et al. (2025), Twitter/X promotion patterns Perdana et al. (2024), and Indonesian spam or promotion evidence Caraka et al. (2025). These studies typically use text preprocessing, TF-IDF, classical machine learning, CNNs, IndoBERT, RoBERTa, multilingual BERT, or hybrid transformer architectures. They show strong within-platform performance, but many remain limited to one language, platform, or dataset.

Website and multimodal studies detect OCR-derived gambling cues from screenshots Wang et al. (2022a), pornographic/gambling websites using visual–textual decision fusion Chen et al. (2020), gambling websites with hybrid multimodal fusion Wang et al. (2022b), multimodal advertisements through text/OCR/ASR extraction Muzakir et al. (2025b), and adjacent visual-advertising signals Jilcha and Kwak (2022). These studies are especially relevant to information security because they address hidden or relocated evidence: gambling indicators may appear in images, web screenshots, HTML resources, OCR output, or visual advertisements rather than ordinary text. Traffic and infrastructure studies extend this view to newly registered domains Guo et al. (2026), POST-message templates and traffic features Feng et al. (2025); Zhang et al. (2025), and illicit website communities built from third-party service identifiers Wang et al. (2023). Mobile-app studies measure illegal gambling app ecosystems Gao et al. (2021), detect underground economy apps through UI transition graph similarity Chen et al. (2023), analyze hybrid dynamic–static UTG detection for illicit apps Lu et al. (2025), and connect underground distribution to Telegram-based app promotion Zhao et al. (2024). Blockchain and financial-flow studies detect crypto-gambling contracts and addresses Huang et al. (2022), hidden delegatees in crypto casinos Wang et al. (2025b), betting anomalies linked to illegal gambling prevention Kim et al. (2024), malicious blockchain-domain behavior Sachan et al. (2023), and broader fraud or behavioral patterns Sun et al. (2022).

These computational studies form the primary technical evidence base of the review. However, most of them are written as individual detection or measurement studies rather than as an integrated information-security synthesis. Their limitations are therefore cross-cutting: heterogeneous datasets, inconsistent reporting, limited dataset openness, incomplete labeling protocols, uneven baseline comparison, and limited adversarial or longitudinal validation. The present review uses RQ1–RQ4 to synthesize these studies across modalities rather than summarizing them one by one.

2.3 Adjacent information-security abuse and countermeasure studies

The third stream consists of adjacent information-security studies that are not always gambling-specific but are important for understanding the security mechanisms used by IOG ecosystems. Adjacent information-security studies show several infrastructure-abuse mechanisms: pop-up scams on typosquatting URLs Dam et al. (2019), typosquatting-oriented malicious URL evidence Dam et al. (2020), DNS and malicious-domain behavior Luo et al. (2018); Li et al. (2020), serverless or cloud-abuse surfaces Hynek et al. (2022); Schulmann and Zhao (2024), web-delivery abuse Wen et al. (2018), and hosting-layer abuse Liu et al. (2025). These studies are relevant because IOG services often depend on rapidly changing domains, redirect chains, hidden infrastructure, and abuse-tolerant service layers.

Other adjacent studies focus on web and platform abuse. Web and platform-abuse studies clarify crawler and indexing manipulation through proactive web defacement Zhao (2023), adversarial hidden links Xiong et al. (2025), hidden gambling-page injection Xiong et al. (2024), repository-based SEO manipulation Wu et al. (2025), pirated-video or third-party abuse contexts Li et al. (2022), and pirated-video website detection evidence Zhang et al. (2022). Mobile and application-security studies further show concealed functionality through underground app distribution Zhao et al. (2024), devious or predatory in-app advertising Liu et al. (2020); Pervez et al. (2026), postMessage-based abuse surfaces Jeong et al. (2023), and iOS mask-app behavior Guo et al. (2024). Blockchain and graph-security studies provide context for Bitcoin address identification Liang et al. (2019), blockchain game mechanics Scholten et al. (2019), blockchain social networks Wu et al. (2021), graph anomaly rules Dong et al. (2025), and fine-grained graph fraud behavior Wang et al. (2025c).

These adjacent studies matter for an AI and information-security audience because they situate IOG detection within broader security problems: abuse-infrastructure detection, adversarial web measurement, platform misuse, fraud analytics, graph-based anomaly detection, software supply-chain abuse, and digital forensics. They also justify the conservative inclusion of several partially aligned primary studies when those studies provide direct technical evidence for evasion or countermeasure design.

2.4 Scope of the systematic literature review

This review focuses on AI-based and computational evidence related to IOG detection, adversarial evasion, and countermeasure methods. The scope includes studies on gambling websites, promotional comments, social media advertisements, mobile

gambling applications, underground app distribution, POST-message traffic, domain and infrastructure abuse, hidden pages, web defacement, SMS-based reporting, filtering systems, behavioral fraud, financial-flow indicators, betting anomalies, blockchain addresses, gambling contracts, and crypto-casino roles.

The review excludes purely legal, clinical, psychological, or public-health gambling studies from the primary-study synthesis unless they provide direct technical or AI-relevant evidence for detection, evasion, robustness, or countermeasure design. Contextual studies on regulation, responsible gambling, prevalence, or gaming–gambling convergence are used only to support framing. In this sense, the review is broader than single-platform gambling detection but narrower than gambling studies in general: it focuses specifically on technical evidence for detection targets, evasion mechanisms, countermeasure methods, robustness evidence, and deployment-related limitations.

2.5 Conceptual background and domain scope

This section introduces the conceptual foundations needed to interpret IOG detection as a technical SLR domain. Following the structure of review papers that first clarify the domain and its basic constructs before moving into technical synthesis, the discussion proceeds from domain scope, definition, core elements, detection tasks, and illustrative examples to the technical layers of detection, evasion, and countermeasure design.

2.5.1 Domain of illegal online gambling detection

The domain of this review is illegal online gambling detection within cybersecurity, content moderation, fraud detection, digital forensics, and infrastructure monitoring. Online gambling itself varies by jurisdiction: some markets regulate responsible-gaming duties and risk prediction Buttigieg et al. (2022), player tracking and early-risk identification Auer and Griffiths (2024), and account-based behavioral monitoring Haefeli et al. (2011), while other settings involve unlicensed, prohibited, or weakly supervised platforms. This review does not redefine legality across jurisdictions. Instead, it treats *illegal online gambling* (IOG) as a technical unit of analysis for gambling services that are unlicensed, prohibited, hidden, covertly distributed, or operationally difficult to supervise.

Within this domain, IOG is not represented by a single object type. Website and multimodal evidence includes screenshot-based co-training Wang et al. (2022a) and visual–textual website detection Chen et al. (2020). Social-media evidence includes semi-supervised YouTube promotion detection Santika et al. (2025), IndoBERT-based YouTube comment classification Kamdan et al. (2025), and Indonesian promotion detection with multiple ML/DL models Liem et al. (2025). Traffic and app evidence includes UI-transition graph app detection Chen et al. (2023) and POST-message traffic analysis Feng et al. (2025). Behavioral, financial-flow, or blockchain evidence includes crypto-gambling contract/address detection Huang et al. (2022), betting-anomaly detection Kim et al. (2024), and gambling fraud detection with semi-supervised GANs Charitou et al. (2020). Contextual evidence from problem-gambling language models Smith et al. (2024) and lottery segmentation Chagas et al.

(2022) delimits adjacent gambling-data research, while gaming–gambling convergence Teichert et al. (2017) and financial-compliance studies Mawardi et al. (2026) clarify neighboring legal and conceptual boundaries.

2.5.2 Definition, elements, tasks, and examples

In this review, an IOG detection system is defined as an AI-enabled or computational pipeline that transforms gambling-related digital artifacts into evidence for classification, filtering, moderation, monitoring, forensic investigation, or enforcement support. The input artifacts may include HTML text, screenshots, advertisements, comments, mobile application packages, URLs, cookies, POST-message bodies, hidden pages, filesystem events, payment traces, graph relations, or blockchain transactions. The output may be a class label, suspiciousness score, cluster, alert, reconstructed URL, graph anomaly, role inference, or forensic indicator.

The core elements of an IOG detection problem are: (1) the *detection target*, such as a website, comment, app, URL, domain, transaction, or blockchain address; (2) the *evidence representation*, such as text embeddings, OCR outputs, visual features, POST-message vectors, UI transition graphs, or transaction graphs; (3) the *evasion mechanism*, such as textual obfuscation, visual embedding, domain churn, hidden-link manipulation, app concealment, payment obfuscation, or blockchain role hiding; (4) the *countermeasure*, such as machine learning, deep learning, transformer models, multimodal fusion, graph analysis, anomaly detection, honeypot reporting, filtering, or forensic monitoring; and (5) the *evaluation claim*, such as effectiveness, robustness, scalability, or deployment readiness.

The tasks in this domain can be grouped into six recurring forms. First, content classification detects gambling-related websites, comments, advertisements, or social media posts. Second, multimodal detection combines screenshots, OCR output, HTML content, and visual features. Third, infrastructure analysis detects suspicious URLs, domains, POST-message traffic, hidden links, and malicious communities. Fourth, mobile and app analysis detects illegal gambling apps, WebView behavior, UI transition patterns, underground distribution channels, and payment-account linkage. Fifth, operational response supports filtering, reporting, alerting, hidden-page detection, intrusion monitoring, and forensic reconstruction. Sixth, behavioral, financial-flow, and blockchain analytics identify fraud records, betting anomalies, payment traces, blockchain contracts, address relationships, and crypto-casino roles.

For example, a gambling promotion classifier may analyze Indonesian YouTube comments that use font manipulation and promotional templates Santika et al. (2025); a multimodal detector may combine website screenshots and HTML text Wang et al. (2022a); a POST-message detector may classify request-line, cookie, and request-body patterns generated by gambling website templates Feng et al. (2025); a mobile detector may infer illegal app behavior from UI transition graphs Lu et al. (2025); and a blockchain detector may infer hidden casino-related roles from transaction or address relationships Wang et al. (2025b). Table 2 summarizes these conceptual foundations before the technical detection layers are discussed.

Table 2 Conceptual foundations of IOG detection.

Conceptual component	Meaning in this review	Typical evidence	Illustrative task
Domain	Cybersecurity-oriented detection of unlicensed, prohibited, hidden, or operationally evasive online gambling services.	Websites, apps, comments, advertisements, traffic, domains, hidden pages, payment traces, and blockchain entities.	Identify whether a digital artifact belongs to an IOG ecosystem.
Detection target	The object or entity being classified, monitored, reconstructed, clustered, or investigated.	Website pages, YouTube comments, APKs, URLs, POST messages, filesystem events, transactions, addresses, and graph nodes.	Detect a gambling website, app, comment, hidden page, suspicious transaction, or blockchain role.
Evidence representation	The model-ready form extracted from raw IOG artifacts.	HTML text, OCR text, screenshots, TF-IDF, embeddings, visual keypoints, URL parameters, UI transition graphs, and transaction graphs.	Convert obfuscated or multimodal evidence into features for detection.
Evasion mechanism	The adversarial strategy used to reduce visibility or bypass detection.	Text mutation, visual embedding, domain rotation, hidden links, POST-template changes, WebView concealment, payment obfuscation, and blockchain anonymity.	Test whether a detector remains reliable under adaptive evasion.
Countermeasure task	The computational or operational response used to identify or mitigate IOG activity.	Classifiers, transformers, multimodal fusion, clustering, graph models, filtering, alerting, honeypots, and forensic monitoring.	Link evasion surfaces to appropriate detection and response workflows.
Evaluation claim	The type of evidence used to support system performance or maturity.	Accuracy, precision, recall, F1-score, AUC, scale of collection, baseline comparison, robustness test, and deployment evidence.	Distinguish task-specific effectiveness from adversarial resilience.

2.5.3 Technical layers of detection, evasion, and countermeasures

Technically, IOG detection involves four recurring stages. Data acquisition collects social-media evidence from YouTube settings Santika et al. (2025) and Twitter/X settings Perdana et al. (2024), mobile ecosystem evidence from illegal gambling apps Gao et al. (2021), network evidence from POST-message traffic Feng et al. (2025), operational evidence from SMS reporting workflows Min et al. (2021), blockchain evidence from smart contracts and addresses Huang et al. (2022), and betting-anomaly evidence from sports-betting records Kim et al. (2024). Feature extraction converts website artifacts into visual–textual and screenshot-based features Chen et al. (2020); Wang et al. (2022b), multimodal advertisements into text/OCR/ASR representations

Muzakir et al. (2025b), app artifacts into UI transition graphs Chen et al. (2023), traffic artifacts into URL, cookie, and request-body features Zhang et al. (2025), and blockchain artifacts into address relationships and graph features Wang et al. (2025b). Related database-driven feature extraction supports gambling-behavior feature construction Stechschulte et al. (2024); shapelet analysis Suzuki et al. (2019) and temporal-stability evidence Murch et al. (2024) further clarify time-sensitive feature interpretation.

Model construction applies co-training and visual learning for website classification Wang et al. (2022a), hybrid heuristic–visual learning for gambling imagery Sánchez et al. (2026), semi-supervised text learning for social-media evidence Santika et al. (2025), transformer-based text learning Anderson et al. (2026), clustering and traffic analysis for POST-message evidence Feng et al. (2025), graph anomaly detection Dong et al. (2025), and heterogeneous graph detection for malicious website infrastructure Fang et al. (2024). Operational response connects detection outputs to honeypot reporting and URL reconstruction Min et al. (2021), hidden-page monitoring workflows Erawan et al. (2024), IDS-style alerting Harahap and Ridho (2024), and forensic or administrator-notification response Teppap et al. (2024). Adjacent non-primary work clarifies neighboring abuse surfaces through piracy-site advertisements and third-party abuse Li et al. (2022), pirated-video website detection Zhang et al. (2022), deceptive mobile-app advertising Liu et al. (2020), predatory in-app advertising Pervez et al. (2026), and push-service abuse through `postMessage` Jeong et al. (2023). Related work on Bitcoin address identification Liang et al. (2019), blockchain games Scholten et al. (2019), blockchain social networks Wu et al. (2021), and Indonesian spam detection Cholistyo and Lhaksmana (2025) provides contextual boundaries for blockchain and spam-like abuse evidence without being treated as primary IOG evidence.

The key technical implication is that conventional detection performance and adversarial robustness are different claims. Website detection reports strong visual–textual performance in PG-VTDM Chen et al. (2020), multimodal fusion reports high gambling-website classification performance Wang et al. (2022b), and POST-message traffic classification reports high accuracy and F1-oriented evidence Feng et al. (2025). However, explicit stress testing against adaptive evasion, cross-platform validation, and deployment-scale resilience remain limited or partially documented, including in platform-specific text detection Santika et al. (2025). Therefore, high performance under conventional train-test splits should not be interpreted as sufficient evidence of resilience against adaptive IOG evasion.

2.6 Analytical taxonomy of illegal online gambling detection systems

To strengthen the technical depth of the review, the synthesis is organized using a taxonomy that separates evidence surfaces, evasion mechanisms, countermeasure architectures, and evaluation evidence. Figure 2 summarizes this taxonomy. The taxonomy is used as an analytical device to avoid a purely descriptive list of studies and to support comparison across heterogeneous studies on text, visual content, mobile apps, URLs, network traffic, financial traces, blockchain entities, and operational monitoring.

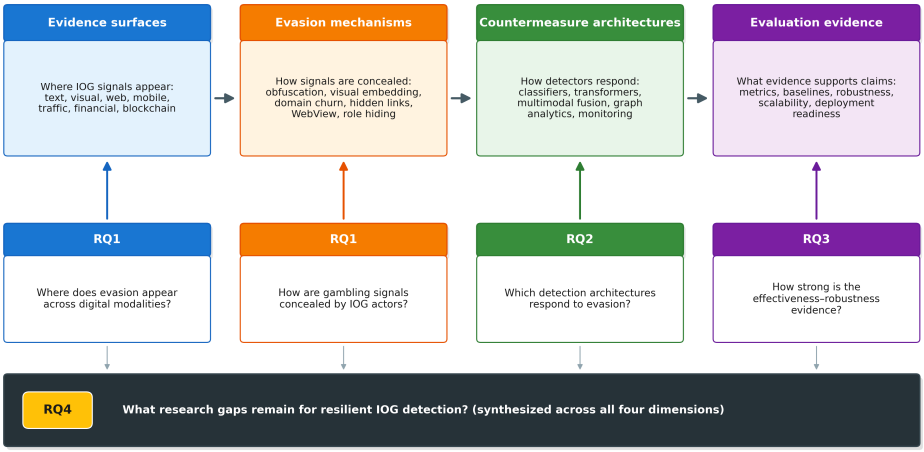


Fig. 2 Taxonomy-guided view of illegal online gambling detection systems, connecting evidence surfaces, evasion mechanisms, countermeasure architectures, and evaluation evidence.

The taxonomy has four dimensions. The first dimension is the *evidence surface*, which identifies where gambling-related evidence is observed. The second dimension is the *evasion mechanism*, which captures how IOG actors conceal or move signals across platforms, content formats, applications, infrastructure, payments, and blockchain relations. The third dimension is the *AI and countermeasure architecture*, which includes rule-based systems, machine learning, deep learning, transformer models, multimodal fusion, graph analytics, anomaly detection, and operational monitoring. The fourth dimension is the *evaluation setup*, which separates ordinary AI performance metrics from stronger evidence of robustness, scalability, external validity, and deployment readiness.

2.7 Positioning of the present review

The present review differs from prior work in two ways. First, it treats IOG detection as an AI-driven adversarial information-security problem, not merely as gambling-content classification. Second, it evaluates whether reported AI and hybrid systems provide enough evidence for resilience against adaptive evasion. The technical streams covered by the 53 primary studies are synthesized in Section 4. The Related Work section therefore establishes the AI-centered gap and scope, while the Results section provides the detailed evidence synthesis for RQ1–RQ4.

3 Methods

This study was conducted as a systematic literature review (SLR) to synthesize artificial intelligence, adversarial evasion techniques, and countermeasure methods in illegal online gambling (IOG) detection systems. The method follows a transparent review protocol covering research-question formulation, data acquisition, study selection, quality assessment, data extraction, and data analysis. This organization is used

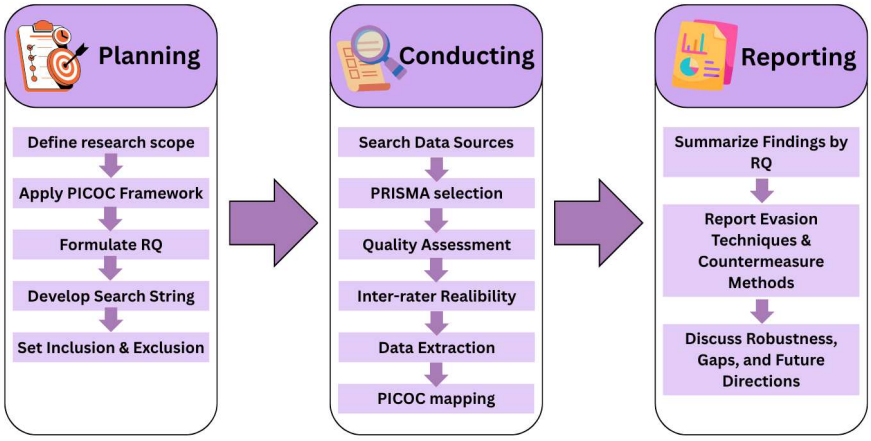


Fig. 3 PICOC-guided SLR workflow from scope definition and study identification to data extraction, synthesis, and reporting.

to make the review reproducible and to align the manuscript with a structured review format in which the search process, selection logic, extraction variables, and synthesis procedure are reported separately.

3.1 Review protocol

The review protocol consisted of five steps: (1) defining the scope and research questions, (2) identifying relevant studies through database searching, (3) selecting eligible studies using predefined criteria, (4) extracting evidence into a structured data form, and (5) collating and synthesizing the extracted evidence according to the research questions. The protocol followed general systematic-review guidance Boland et al. (2017), software-engineering SLR recommendations Kitchenham and Charters (2007), and recent computer-science SLR process guidance Carrera-Rivera et al. (2022). Figure 3 summarizes the workflow.

The PICOC framework was used to keep the review scope consistent with the search string, eligibility criteria, and extraction form. Table 3 defines the PICOC elements used in this review.

3.2 Research questions

The research questions were designed to move from evasion evidence to countermeasure evidence, effectiveness and robustness assessment, and future framework requirements. Table 4 lists the research questions and their rationale.

3.3 Analytical taxonomy

The review used the taxonomy introduced in Section 2.6 as an AI-oriented analytical guide for data extraction and synthesis. The taxonomy separates four dimensions: evidence surface, evasion mechanism, AI and countermeasure architecture, and evaluation

Table 3 PICOC framework used to define the scope of the SLR.

Component	Definition in this SLR
Population	Illegal online gambling platforms, gambling websites, gambling applications, gambling-related digital content, and online gambling ecosystems.
Intervention	AI-based detection methods, adversarial evasion techniques, countermeasure strategies, and automated detection systems.
Comparison	Rule-based approaches, signature-based detection, machine learning methods, deep learning methods, transformer-based systems, graph-based AI, multimodal AI, and hybrid detection architectures.
Outcome	AI model performance, detection metrics, robustness, evasion resistance, scalability, and operational effectiveness.
Context	Artificial intelligence, cybersecurity, fraud detection, digital forensics, web security, network security, content moderation, and adversarial machine learning.

Table 4 Research questions and rationale.

RQ	Research question	Rationale
RQ1	What adversarial evasion techniques challenge AI-based illegal online gambling detection across different digital modalities?	To identify the evasion surfaces that affect AI models across textual, visual, web, mobile, operational, financial, and blockchain contexts.
RQ2	What AI-based countermeasure methods and detection architectures have been proposed to address these evasion techniques?	To synthesize how AI and hybrid methods transform raw IOG evidence into classifications, alerts, clusters, graph relations, anomaly scores, or forensic indicators.
RQ3	How effective, robust, and scalable are existing AI-based illegal online gambling detection systems against adversarial evasion strategies?	To distinguish task-specific AI performance from robustness, scalability, external validity, and deployment readiness.
RQ4	What research gaps, open challenges, and future directions exist for developing resilient AI-based illegal online gambling detection frameworks?	To derive technical requirements for adaptive, scalable, and operationally deployable AI-based IOG detection systems.

setup. This separation is important because an AI detector may report strong performance on one surface, such as text classification, while providing limited evidence for robustness against other evasion surfaces, such as image embedding, domain rotation, POST-message camouflage, WebView concealment, or blockchain role hiding.

Table 5 Database sources and initial records identified on 6 May 2026.

Database	Initial records	Search fields used where available
Scopus	137	Title, abstract, and author keywords.
ScienceDirect	718	Title, abstract, and keyword-related database fields.
IEEE Xplore	63	Metadata fields supported by IEEE Xplore advanced search.
ACM Digital Library	140	Title, abstract, and keyword-related metadata fields.
Total	**1,058**	Four database sources.

3.4 Data acquisition

3.4.1 Search criteria

The search string was derived from the PICOC framework and research questions. No publication-year restriction was applied because, during review formulation, no prior SLR specifically focusing on adversarial evasion and countermeasure methods in IOG detection systems was identified. The final search string was:

```
("online gambling" OR "gambling platform" OR "gambling website" OR
"gambling app" OR "illegal gambling") AND ("detection" OR "evasion" OR
"countermeasure" OR "classification")
```

The search string was intentionally broad and was not restricted to the term "artificial intelligence" because many relevant IOG countermeasure studies use mixed terminology such as machine learning, deep learning, transformer models, multimodal fusion, graph analysis, anomaly detection, filtering, or forensic monitoring. AI relevance was therefore assessed during eligibility screening, quality assessment, and structured data extraction rather than by a narrow keyword filter.

The search string was kept unchanged during revision to preserve methodological consistency with the original protocol. Searches were applied to article title, abstract, and keywords where the database interface supported this field combination.

3.4.2 Data sources and metadata management

The search was conducted on 6 May 2026 using four academic databases: Scopus, ScienceDirect, IEEE Xplore, and ACM Digital Library. The initial search identified 1,058 records. As shown in Table 5, ScienceDirect contributed the largest number of initial records, followed by ACM Digital Library, Scopus, and IEEE Xplore.

The retrieved records were managed using a metadata acquisition sheet. The acquisition sheet contained bibliographic and indexing fields such as record key, item type, publication year, author, title, publication venue, ISBN/ISSN, DOI, URL, abstract note, date added, pages, issue, volume, publisher, language, tags, notes, and file-attachment information. This metadata layer was used to support duplicate checking, accessibility checking, screening, and later linkage between bibliographic records and

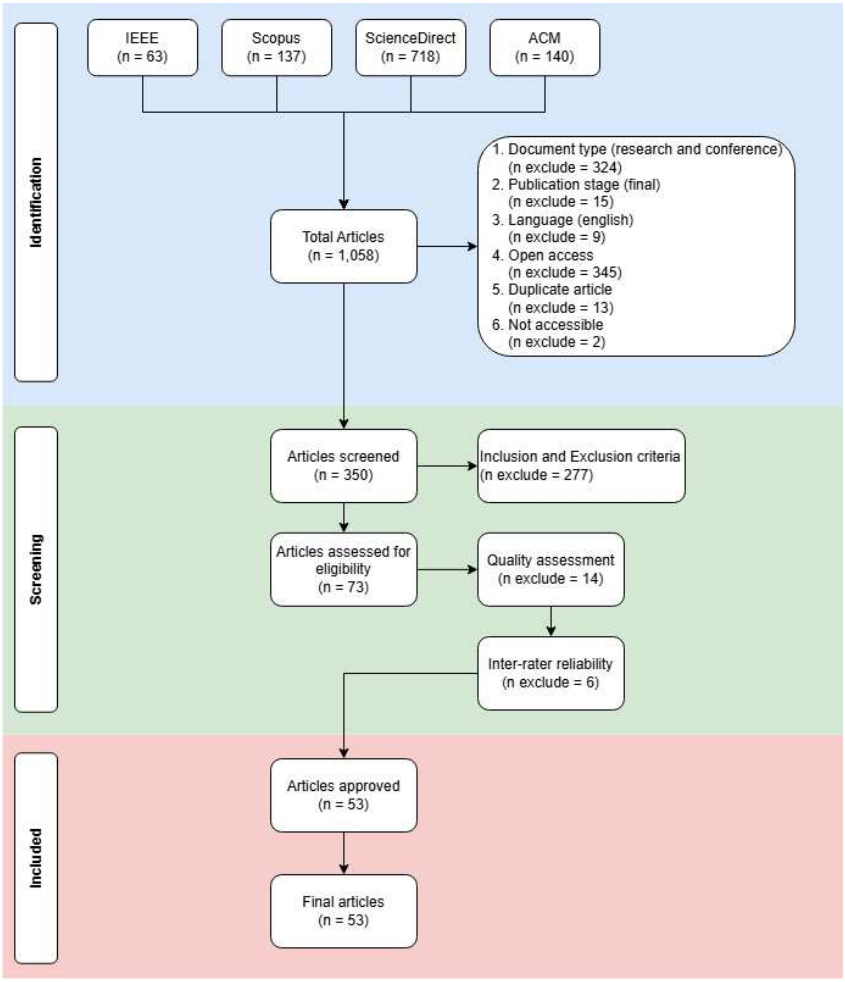


Fig. 4 PRISMA-style flow of study identification, screening, eligibility assessment, and final inclusion.

full-text evidence. The acquisition sheet contained 352 post-filter metadata records, consisting of 217 journal-article records and 135 conference-paper records.

3.4.3 Selection of studies

The selection process followed a PRISMA-style flow. Database-specific filtering excluded 324 records by document type, 15 by publication stage, 9 by language, 345 by open-access or accessibility status, 13 duplicates, and 2 inaccessible full-text records. After these steps, 350 accessible records remained for title, abstract, and keyword screening. The screening stage excluded 277 records that did not satisfy the eligibility criteria, leaving 73 articles for quality assessment, reliability checking, and adjudication. Figure 4 summarizes the study-selection flow.

Table 6 Eligibility criteria used for study selection.

Component	Inclusion criteria	Exclusion criteria
Publication type	Journal and conference papers indexed in reputable academic sources.	News articles, blogs, opinion papers, review papers, theses, or non-scientific publications.
Topical relevance	The study focuses on online gambling detection, adversarial evasion, countermeasures, or closely related technical detection problems.	The study is unrelated to IOG detection or uses gambling-related terms only incidentally.
Technical evidence	The study reports a method, model, framework, measurement, dataset, evaluation, or countermeasure relevant to detection or evasion.	The study provides no technical evidence relevant to detection, evasion, robustness, or countermeasure design.
Accessibility	Full text is accessible for assessment and extraction.	Full text is not accessible.

The eligibility criteria were defined before the final screening stage. Table 6 summarizes the inclusion and exclusion criteria.

3.4.4 Quality assessment

The 73 assessed articles were evaluated using ten binary quality-assessment questions. The checklist was designed to assess whether each study provided sufficient evidence for technical synthesis, not merely whether it mentioned online gambling. Each item was scored as Yes = 1 or No = 0, producing a total score from 0 to 10. Scores of 9–10 were classified as high quality, 6–8 as moderate quality, and 0–5 as low quality. Table 7 presents the checklist.

3.4.5 Inter-rater reliability and adjudication

Inter-rater reliability was assessed for all 73 articles that reached the quality-assessment and adjudication stage. No article was excluded from the Kappa calculation. The reviewer-decision sheet records the study identifier, QA score, quality category, relevance category, Reviewer A decision, Reviewer B decision, agreement status, final decision, adjudication note, related RQ mapping, and inclusion or exclusion rationale. Three decision categories were used: *Include*, *Exclude*, and *Needs adjudication*. Cohen's Kappa Cohen (1960) was used because it measures categorical agreement while accounting for chance agreement. The interpretation followed the agreement scale proposed by Landis and Koch Landis and Koch (1977). Table 8 summarizes the reliability results, and Table 9 presents the reviewer-decision matrix.

Using the matrix in Table 9, the observed agreement was $P_o = 66/73 = 0.904110$, the expected agreement was $P_e = 0.368174$, and Cohen's Kappa was calculated using Eq. (1).

$$\kappa = \frac{P_o - P_e}{1 - P_e} = \frac{0.904110 - 0.368174}{1 - 0.368174} = 0.848233. \tag{1}$$

Table 7 Quality assessment checklist used in this review.

Code	Assessment question	Score
QA1	Does the study clearly state its objective related to illegal online gambling detection, adversarial evasion, or countermeasure methods?	1/0
QA2	Does the study implement or analyze technical detection, evasion, or countermeasure techniques?	1/0
QA3	Does the study provide information about the dataset, data source, or collected gambling-related data?	1/0
QA4	Does the study describe preprocessing, feature engineering, or extraction methods used in the detection process?	1/0
QA5	Does the study explain the detection system architecture, framework, or workflow?	1/0
QA6	Does the study report evaluation metrics or experimental performance results?	1/0
QA7	Does the study compare its approach with baseline, existing, or state-of-the-art methods?	1/0
QA8	Does the study discuss adversarial evasion strategies, detection challenges, or system vulnerabilities?	1/0
QA9	Does the study address robustness, reliability, scalability, bias mitigation, or validation strategies?	1/0
QA10	Does the study provide recommendations, future research directions, or improvement opportunities for gambling detection systems?	1/0

Table 8 Summary of inter-rater reliability and final study-selection outcomes.

Component	Result
Total articles assessed	73
Papers used for Cohen's Kappa	73
Papers excluded from Cohen's Kappa	0
Direct Include	31
Include with adjudication note	22
Total final primary studies	53
Background only	6
Excluded	14
Observed agreement (P_o)	0.904110
Expected agreement (P_e)	0.368174
Cohen's Kappa (κ)	0.848233
Interpretation	Almost perfect agreement

The value in Eq. (1) indicates almost perfect agreement. After adjudication, 53 studies were retained as primary studies, consisting of 31 direct includes and 22 studies included with adjudication notes. Six studies were retained as background-only sources, and 14 studies were excluded from the synthesis. Table 10 summarizes the final decision categories.

The adjudication notes were retained in the extraction record to avoid overstating the contribution of partially related studies. In particular, studies marked as *Include*

Table 9 Reviewer-decision matrix for Cohen's Kappa calculation based on 73 assessed articles.

Reviewer A / Reviewer B	**Include**	**Exclude**	**Needs adjudication**	**Row total**
Include	31	0	7	38
Exclude	0	14	0	14
Needs adjudication	0	0	21	21
Column total	**31**	**14**	**28**	**73**

Table 10 Final decision categories after quality assessment and adjudication.

Final decision category	**Number of studies**	**Description**
Direct Include	31	Counted as primary studies without adjudication note.
Include with adjudication note	22	Counted as primary studies after adjudication or stricter relevance review.
Background only	6	Used for contextual support only and not counted in the primary-study synthesis.
Excluded	14	Removed from the primary-study synthesis.
Final primary studies	**53**	Main evidence base for the SLR.

with adjudication note were used when they provided technical evidence relevant to at least one research question, such as blockchain crime detection, hidden-link evasion, parental-control filtering, underground mobile distribution, general fraud detection, or adversarial infrastructure abuse. These studies were interpreted conservatively in the synthesis and were not treated as direct evidence for every IOG detection subproblem.

Snowballing was considered as a supplementary identification strategy. However, the available acquisition and extraction records do not document a separate backward- or forward-snowballing stage, citation database, snowballing count, or additional snowballed inclusion. Therefore, snowballing was not added as a separate numerical stage in the PRISMA-style flow.

3.5 Data extraction

Data extraction was conducted using a structured extraction sheet containing 53 primary studies and 53 extraction fields. The extraction form was designed to capture bibliographic details, study context, evasion evidence, detection architecture, dataset information, evaluation evidence, robustness, scalability, limitations, research gaps, and RQ-level contributions. Table 11 summarizes the extraction dimensions and their analytical use.

To reduce over-interpretation, missing information was coded as *Not Reported*, and incomplete information was coded as *Partially Reported*. This conservative coding was applied to dataset availability, labeling strategy, training/testing split, baseline comparison, robustness evidence, scalability evidence, and real-world applicability.

Table 11 Data extraction dimensions and analytical use.

Extraction dimension	Extracted fields	Analytical use
Bibliographic profile	Study ID, authors, year, title, source or venue, publication type, rank, and study type.	Describes the evidence base and publication profile.
Research context	Research objective, IOG context, digital modality, and population target.	Defines the detection target and review scope.
RQ1: evasion evidence	Evasion technique, evasion description, evasion category, and adversarial or evasion evidence level.	Synthesizes adversarial evasion surfaces and tactics.
RQ2: countermeasure evidence	Detection method, AI model family, method category, architecture or workflow, algorithm or model name, model type, model process, feature representation, and features used.	Compares AI-based, hybrid, and technical detection and countermeasure architectures.
RQ3: evaluation evidence	Dataset name, source, URL or repository, size, format, labeling strategy, availability, split, evaluation method, metrics, main results, baseline, robustness evidence, scalability evidence, and real-world applicability.	Separates task-specific effectiveness from robustness, scalability, and deployment readiness.
RQ4: synthesis evidence	Reported limitations, research gaps, future work, RQ contribution, PICOC mapping, novelty category, novelty description, evidence strength, and extraction notes.	Derives research gaps and future framework requirements.

The final primary-study set consisted of 35 conference papers and 18 journal articles. Rank information was available for part of the journal set and was retained only when explicitly documented in the extraction sheet.

3.6 Data analysis

The extracted data were analyzed using structured narrative synthesis and thematic evidence mapping. RQ1 was analyzed by grouping evasion techniques according to digital modality and evasion category. RQ2 was analyzed by grouping methods according to AI model family, countermeasure family, feature representation, and architecture or workflow. RQ3 was analyzed by comparing reported metrics, baselines, dataset evidence, robustness status, scalability evidence, and real-world applicability. RQ4 was analyzed by aggregating reported limitations, research gaps, future-work suggestions, and extraction notes into framework-level research directions.

Because several studies report more than one modality, method, or evasion mechanism, multi-label coding was used where appropriate. As a result, some modality and evasion-category counts exceed the number of unique primary studies. This is expected because a single study may address, for example, both textual and visual evasion, or

both URL/domain and financial-flow evidence. The synthesis tables in the Results section therefore report category assignments rather than mutually exclusive classes.

The included studies vary substantially in platform, dataset type, model architecture, evaluation protocol, and metric definition. For this reason, a pooled-effect performance meta-analysis of accuracy or F1-score was not conducted. Treating heterogeneous tasks, such as YouTube comment classification, website screenshot classification, POST-message classification, mobile app detection, filtering evaluation, betting anomaly detection, and blockchain forensics, as directly comparable prediction problems would be statistically misleading. Instead, the review combines narrative synthesis, thematic grouping, structured comparison, evidence mapping, and a descriptive meta-analysis of binary evidence indicators such as metric reporting, baseline reporting, robustness evidence, scalability evidence, split reporting, labeling transparency, and dataset availability.

3.7 Technical evidence maturity assessment

To address the need for deeper technical analysis, the review added an evidence maturity score (EMS). Each primary study was coded according to whether it reported evaluation metrics, baseline or comparative methods, explicit or partial robustness evaluation, scalability evidence, and real-world applicability. These dimensions were converted into the EMS in Eq. (2).

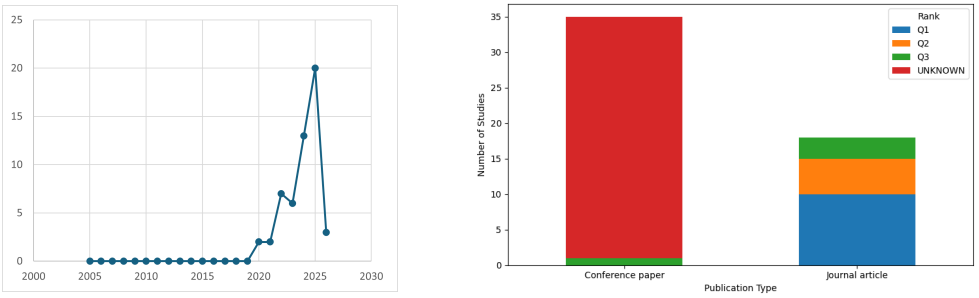
$$EMS_i = M_i + B_i + R_i + S_i + D_i, \tag{2}$$

where M_i indicates that study i reports evaluation metrics, B_i indicates baseline or comparative evaluation, R_i indicates explicit or partial robustness evaluation, S_i indicates moderate or strong scalability evidence, and D_i indicates moderate or strong real-world applicability. Eq. (2) is not used to rank overall scientific quality. It is used only to evaluate how complete the AI and technical evidence is for each study. This assessment strengthens RQ3 by distinguishing reported task-specific effectiveness from adversarial resilience and deployment readiness.

4 Results

According to the four research questions, the results are presented in four main subsections: (1) adversarial evasion techniques in IOG ecosystems, (2) AI-based and hybrid countermeasure architectures, (3) effectiveness, robustness, scalability, and reproducibility evidence, and (4) research gaps and future directions. A final subsection integrates these findings into an evidence-based framework for adversarially resilient AI-based IOG detection.

The RQ-aligned presentation follows recent SLR reporting practice in which evidence is organized by analytical dimensions before being integrated into a framework Awadh et al. (2025). The synthesis is based on 53 primary studies retained after quality assessment, reliability checking, and adjudication. Because several studies address more than one modality, method, or evasion mechanism, category counts are interpreted as multi-label assignments rather than mutually exclusive study counts. The evidence base shows that IOG detection is no longer limited to standalone



(a) Annual publication trend

(b) Publication type and source rank

Fig. 5 Overview of the 53 primary studies by publication year and publication-source profile.

website or keyword classification: it spans textual promotion detection, multimodal website analysis, mobile-app ecosystems, traffic and URL analysis, hidden-page detection, financial-flow indicators, and blockchain analytics. Figure 5 summarizes the publication-year distribution and source profile of the primary studies.

4.1 RQ1: Adversarial Evasion Techniques

RQ1 asks what adversarial evasion techniques are employed by IOG actors across digital modalities. The review shows six recurring evasion surfaces: textual and social media manipulation, visual and multimodal concealment, web/URL/infrastructure evasion, mobile-app concealment, behavioral or financial-flow obfuscation, and operational bypass.

Textual evasion appears through symbol insertion, Unicode variation, modified domains, slang, fonts, emojis, and special-character manipulation in promotional comments Prakoso et al. (2025); Anderson et al. (2026); Santosa et al. (2025). Visual and multimodal evasion shifts gambling cues into screenshots, banners, images, or audiovisual content that require OCR, visual encoders, or multimodal extraction Wang et al. (2022a,b); Muzakir et al. (2025b). Infrastructure-level evasion is reflected in domain churn, POST-message templates, hidden links, crawler resistance, and traffic patterns Guo et al. (2026); Feng et al. (2025); Zhang et al. (2025). Mobile and app-ecosystem evasion includes covert distribution, WebView concealment, URL modification, and payment-service abuse Gao et al. (2021); Min et al. (2022); Lu et al. (2025). Financial and blockchain evasion appears through betting anomalies, fraud patterns, transaction relationships, and hidden crypto-casino roles Kim et al. (2024); Charitou et al. (2020); Wang et al. (2025b). Operational bypass appears through filtering gaps, hidden pages, compromised websites, and incomplete reporting or blocking workflows Min et al. (2021); Liberato et al. (2025); Rojas et al. (2025).

Figure 6 indicates that web, URL, and infrastructure evasion is the most persistent surface across the reviewed years, while textual/social-media and operational-bypass evidence becomes more visible in recent studies. This pattern supports the main RQ1 finding: IOG evasion is not a single-content problem but a layered adversarial behavior that moves across content, infrastructure, applications, payments, and graphs.

Table 12 summarizes the modality-level mapping. The implication is that single-feature detectors remain fragile even when they perform well in controlled datasets.

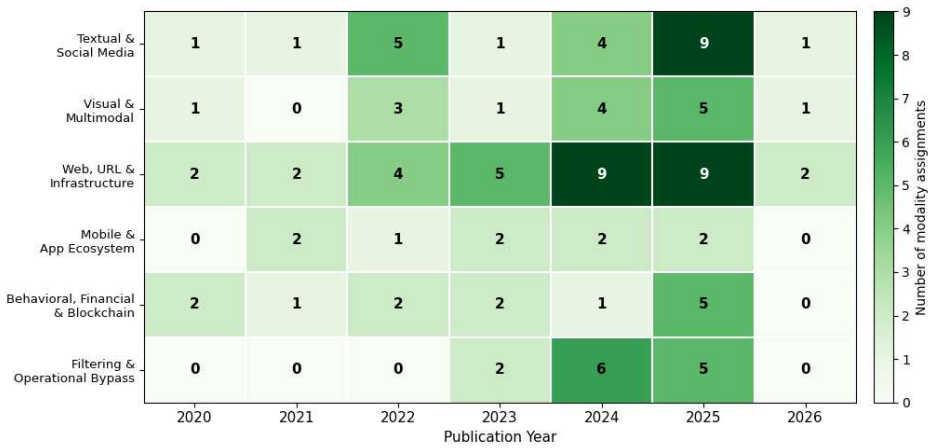


Fig. 6 Year-by-year distribution of evasion-modality assignments across the 53 primary studies.

Future AI-based IOG detection should combine textual, visual, URL, mobile, behavioral, financial, operational, and graph-based evidence and evaluate whether these representations remain reliable under adaptive evasion.

4.2 RQ2: AI-Based Detection and Countermeasure Architectures

RQ2 asks what AI-based, hybrid, and computational countermeasures have been proposed to address the evasion surfaces identified in RQ1. The synthesis maps each method to its evidence surface, feature representation, model family, and remaining limitation. Six countermeasure families recur across the primary studies.

Text-oriented methods use preprocessing, TF-IDF, CNNs, AutoML, IndoBERT, RoBERTa, multilingual BERT, and hybrid transformers for promotional comments and social-media advertisements Santika et al. (2025); Kamdan et al. (2025); Liem et al. (2025); Prakoso et al. (2025); Anderson et al. (2026); Muzakir et al. (2025a). Visual and multimodal methods combine OCR, screenshots, HTML, visual descriptors, Doc2Vec/Word2Vec, ResNet-based backbones, Bi-LSTM, self-attention, and fusion strategies Wang et al. (2022a); Chen et al. (2020); Wang et al. (2022b); Wen et al. (2024); Zhang et al. (2023). Traffic, domain, URL, and infrastructure methods analyze POST messages, URL/cookie/body parameters, newly registered domains, third-party identifiers, heterogeneous graph relations, and malicious-domain indicators Feng et al. (2025); Zhang et al. (2025); Guo et al. (2026); Wang et al. (2023); Fang et al. (2024). Mobile-app approaches use app-domain association, UI transition graphs, WebView behavior, Telegram distribution evidence, and joint-promotion relationships Gao et al. (2021); Chen et al. (2023); Lu et al. (2025); Guo et al. (2024); Han et al. (2023). Operational methods use honeypots, URL reconstruction, filesystem monitoring, Suricata alerts, filtering evaluation, and administrator notification Min et al. (2021); Erawan et al. (2024); Harahap and Ridho (2024); Teppap et al. (2024). Behavioral, financial,

Table 12 Adversarial evasion techniques by modality. Counts are multi-label assignments over the 53 primary studies.

Evasion modality	Mapped primary studies	Main evasion signals	Representative primary studies
Textual and social media	22 (41.5%)	Templated comments, font manipulation, emojis, slang, prohibited-word avoidance, multilingual characters, media files, poll options, and illicit jargon.	Prakoso et al. (2025); Anderson et al. (2026); Santosa et al. (2025)
Visual and multimodal	15 (28.3%)	Gambling cues embedded in screenshots, banners, images, videos, OCR noise, and gaming-like visual content.	Wang et al. (2022a,b); Muzakir et al. (2025b)
Web, URL, and infrastructure	33 (62.3%)	Disguising, misleading content, blocking, URL churn, newly registered domains, hidden links, POST templates, traffic camouflage, SEO manipulation, and crawler-resistant structures.	Chen et al. (2020); Feng et al. (2025); Zhang et al. (2025)
Mobile and app ecosystem	9 (17.0%)	Covert app distribution, underground channels, Telegram-based distribution, operator identity hiding, WebView behavior, mask apps, missing malicious payloads, and payment-service abuse.	Gao et al. (2021); Min et al. (2022); Lu et al. (2025)
Behavioral, financial-flow, and blockchain	13 (24.5%)	Fraud imbalance, betting anomalies, transaction layering, payment obfuscation, hidden roles, address anonymity, smart-contract relationships, and transaction-network opacity.	Kim et al. (2024); Zhang et al. (2025); Charitou et al. (2020)
Filtering and operational bypass	13 (24.5%)	Filtering gaps, parental-control limitations, hidden-page injection, compromised websites, defacement persistence, repository abuse, and delayed response.	Min et al. (2021); Liberato et al. (2025); Rojas et al. (2025)

and blockchain methods use betting-anomaly classifiers, sparse autoencoders, semi-supervised GANs, illicit financial-flow analysis, multimodal retrieval, graph learning, and smart-contract/address classification Kim et al. (2024); Charitou et al. (2020); Chen et al. (2024); Huang et al. (2022); Wang et al. (2025b); Dong et al. (2025).

Table 13: Mapping between evasion surfaces, countermeasure families, and remaining evidence limitations.

Evasion surface	Countermeasure family	Typical methods and representations	Evidence limitation	Representative studies
Textual and social media obfuscation	Text-oriented AI detection	Text preprocessing, font normalization, TF-IDF, contextual embeddings, CNN, XGBoost, Random Forest, IndoBERT, RoBERTa, and hybrid classifiers.	Strong platform-specific performance is reported, but adaptive text-mutation, slang variation, multilingual obfuscation, and cross-platform robustness testing remain limited.	Santika et al. (2025); Kamdan et al. (2025); Liem et al. (2025)
Visual embedding and multimodal concealment	Visual and multimodal AI detection	OCR extraction, screenshots, HTML text, Doc2Vec, Word2Vec, ResNet34, visual–textual fusion, decision-level fusion, and self-attention-based multimodal architectures.	Visual–textual fusion addresses hidden cues, but robustness against website redesign, OCR noise, image manipulation, adversarial visual changes, and cross-site variation is rarely stress-tested.	Wang et al. (2022a); Chen et al. (2020); Wang et al. (2022b)
Traffic, domain, URL, and hidden-link manipulation	Traffic, domain, and infrastructure analysis	POST-message features, URL/cookie/body parameters, clustering, graph analysis, heterogeneous neural networks, domain lifecycle indicators, third-party service identifiers, and crawler-based inspection.	Stronger scale-oriented evidence exists, but adaptive domain rotation, crawler resistance, hidden-link mutation, SEO manipulation, and infrastructure migration remain only partially evaluated.	Guo et al. (2026); Feng et al. (2025); Zhang et al. (2025)
Mobile app concealment and underground distribution	Mobile and app-ecosystem detection	APK analysis, app-domain association, UI transition graphs, similarity analysis, Telegram ecosystem measurement, WebView behavior analysis, mask-app detection, and payment-account linkage.	Large app-ecosystem evidence is available, but app repackaging, WebView concealment, distribution-channel migration, and adversarial app-interface changes are not consistently tested as adversarial scenarios.	Gao et al. (2021); Min et al. (2022); Lu et al. (2025)

Continued on next page

Table 13 continued from previous page

Evasion surface	Countermeasure family	Typical methods and representations	Evidence limitation	Representative studies
Operational bypass, hidden pages, and compromised websites	Operational and forensic response	Honeypot reporting, URL reconstruction, Wazuh alerts, Velociraptor artifacts, filesystem monitoring, filtering evaluation, defacement detection, administrator notification, and incident-response workflows.	Practical deployment value is visible, but integration with AI robustness evaluation, longitudinal adversarial testing, and end-to-end operational validation remains limited.	Min et al. (2021); Liberato et al. (2025); Rojas et al. (2025)
Behavioral, financial-flow, and blockchain evasion	Behavioral, financial, and graph-based AI detection	Betting-anomaly classifiers, sparse autoencoders, semi-supervised GANs, clustering, transaction analysis, address classification, graph attention models, smart-contract analysis, and crypto-role inference.	Financial-flow and blockchain evidence broadens detection beyond content, but adversarial validation for payment obfuscation, transaction layering, blockchain role hiding, and graph perturbation remains sparse.	Kim et al. (2024); Zhang et al. (2025); Charitou et al. (2020)

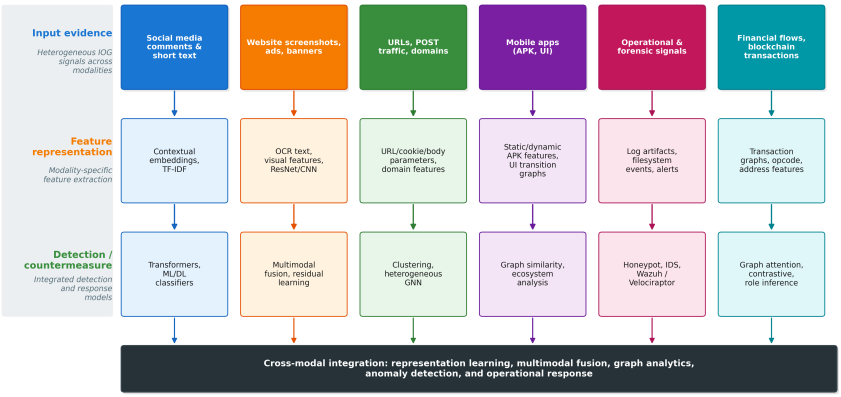


Fig. 7 Architecture synthesis of detection and countermeasure methods in RQ2.

Table 13 shows that no single method family covers all evasion surfaces. Text models are suitable for promotional language, multimodal fusion is needed for visual concealment, graph and traffic methods are needed for infrastructure and blockchain signals, and operational tools are needed for reporting and response. The main RQ2 finding is therefore that resilient IOG detection should be treated as an evidence-integration problem rather than as a search for one universal classifier.

4.3 RQ3: AI Effectiveness, Robustness, and Scalability

RQ3 asks how effective and robust existing IOG detection systems are against adversarial evasion. The analysis separates four evidence dimensions: reported effectiveness, baseline comparison, robustness evidence, and scalability or deployment-oriented evidence. This separation is necessary because high accuracy or F1-score in one dataset does not automatically demonstrate adversarial resilience.

Figure 7 summarizes the common architecture pattern: heterogeneous IOG evidence is transformed into modality-specific representations and then passed to detection, monitoring, inference, or response components. The figure is conceptual rather than an implemented system; it visualizes how the reviewed studies distribute detection work across input evidence, feature construction, AI modeling, and operational response.

Table 14 shows that metric reporting is common, but explicit adversarial validation is limited. Most studies provide task-specific effectiveness evidence, many report baseline comparisons or scalability-oriented data collection, but far fewer evaluate adaptive evasion directly. This creates a gap between AI detectability and adversarial resilience.

4.3.1 Descriptive meta-analysis of evidence proportions

Because the 53 studies differ in task, modality, dataset, label definition, and metric family, a pooled-effect performance meta-analysis would be misleading. Instead, this review uses descriptive evidence indicators and Wilson 95% confidence intervals to summarize how often key forms of evidence are reported.

Table 14 Evidence-depth summary for effectiveness, robustness, scalability, and reproducibility across the primary studies.

Evidence dimension	Mapped studies	Interpretation for technical depth	RQ link
Evaluation metrics reported	51/53 (96.2%)	Most studies report task-specific effectiveness using accuracy, precision, recall, F1-score, AUC, counts, or operational measures. These metrics are not directly comparable across heterogeneous tasks.	RQ3
Baseline or comparison reported	36/53 (67.9%)	Comparative evidence is present in many studies, but baseline choice varies across text, multimodal, traffic, app, filtering, and blockchain contexts.	RQ3
Explicit robustness evaluation	9/53 (17.0%)	Controlled or explicit evasion-oriented evaluation remains limited and is the weakest evidence dimension.	RQ3/RQ4
Partial robustness evaluation	11/53 (20.8%)	Some studies provide indirect robustness evidence through real-world, large-scale, noisy, imbalanced, or deployment-like evaluation, but not full adversarial stress testing.	RQ3/RQ4
Evasion discussed but not tested	31/53 (58.5%)	Evasion is frequently recognized as a threat, but usually handled as motivation or limitation rather than as a controlled test condition.	RQ1/RQ3
Moderate or strong scalability evidence	43/53 (81.1%)	Many studies use real-world crawls, platform data, domain-scale datasets, app collections, traffic logs, or blockchain traces, but scalability is uneven across modalities.	RQ3
Training/testing split reported	24/53 (45.3%)	Reproducibility is constrained because fewer than half of the studies clearly report train–test or validation split information.	RQ3/RQ4
Dataset labeling strategy reported	32/53 (60.4%)	Labeling transparency remains incomplete, limiting confidence in cross-study performance interpretation.	RQ3/RQ4
EMS ≥ 4	38/53 (71.7%)	Many studies provide several technical-evidence components, but high EMS should not be interpreted as adversarial robustness because robustness is only one EMS component.	RQ3

Table 15 confirms that the literature is technically active but unevenly mature. Evaluation metrics are reported in 51/53 studies, whereas explicit robustness evaluation appears in only 9/53 studies. Dataset openness and repository-level reproducibility are also limited, which reduces the possibility of independent replication and adversarial stress testing.

Table 15 Descriptive meta-analysis of evidence indicators across the primary studies.

Evidence indicator	x/n	Proportion	95% CI	Interpretation
Evaluation metrics reported	51/53	96.2%	87.2–99.0%	Most studies report task-specific metrics, but the metrics are heterogeneous.
Baseline or comparison reported	36/53	67.9%	54.5–78.9%	Comparative evidence is common but baseline choice varies across modalities.
Any robustness evidence	20/53	37.7%	25.9–51.2%	Combines explicit and partial robustness evidence.
Explicit robustness evaluation	9/53	17.0%	9.2–29.2%	Controlled evasion-oriented testing remains limited.
Moderate/strong scalability evidence	43/53	81.1%	68.6–89.4%	Many studies use real-world, platform, domain, app, traffic, or blockchain evidence.
Training/testing split reported	24/53	45.3%	32.7–58.5%	Split reporting is incomplete and affects reproducibility.
Dataset labeling strategy reported	32/53	60.4%	46.9–72.4%	Labeling transparency remains uneven.
Repository/project/DOI link reported	19/53	35.8%	24.3–49.3%	Dataset, model, project, DOI, or repository link is reported.
Explicit public/released/open or DOI-published dataset	6/53	11.3%	5.3–22.6%	Only studies clearly coded as public, released, open, public-source, or DOI-published are counted here.

Figure 8 visualizes the same imbalance: ordinary metric reporting is far more common than explicit robustness testing, public dataset release, or complete reproducibility evidence.

4.3.2 Descriptive meta-analysis of reported performance results

Thirty studies contained extractable standard point estimates from the *Main Performance Results* column. The median reported value was 97.00%, with values ranging from 85.01% to 99.85%. These values show strong within-task performance across several modalities, including transformer-based text detection, multimodal website classification, POST-message analysis, mobile-app detection, defacement monitoring, and blockchain-oriented classification. However, the values are descriptive only: a YouTube-comment F1-score, a gambling-website accuracy score, a POST-message F1-score, and a smart-contract classification F1-score measure different prediction problems and should not be ranked as if they were the same task.

Table 16 and Figure 9 should therefore be interpreted as a performance-profile summary, not as a pooled benchmark. The main RQ3 message is that high reported performance is strongest as evidence of within-dataset detectability. It becomes evidence of adversarial resilience only when accompanied by temporal validation, cross-platform validation, adversarial mutation tests, out-of-domain testing, false-positive analysis, and deployment evidence.

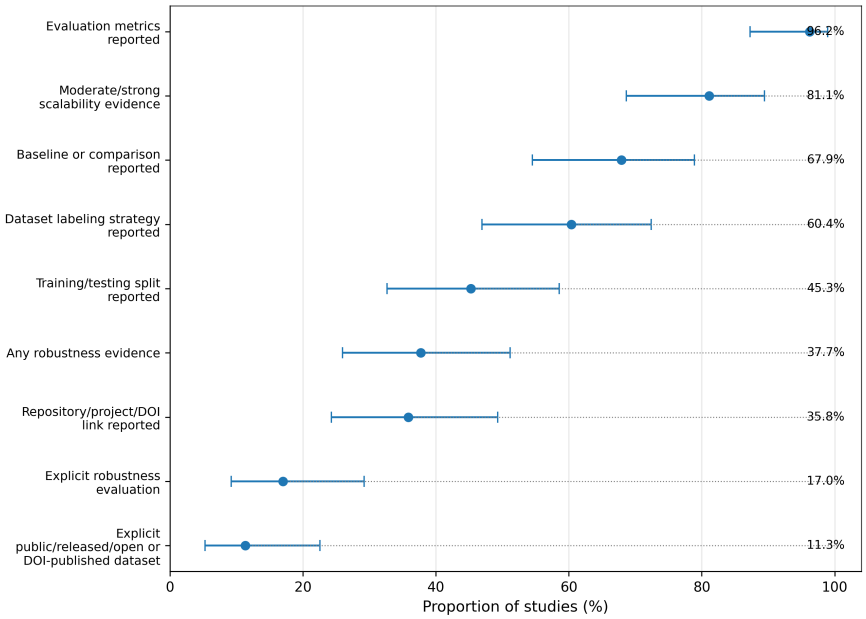


Fig. 8 Descriptive meta-analysis of evidence indicators using pooled proportions and Wilson 95% confidence intervals across the 53 primary studies.

Table 16 Descriptive meta-analysis of extractable point estimates from the *Main Performance Results* column.

Metric family	Studies	Mean	Median	Range
Accuracy-family	10	93.54%	92.90%	85.01–99.18%
F1-family	18	96.09%	97.73%	89.00–99.85%
Other metric	2	97.86%	97.86%	96.60–99.12%

4.3.3 Published and repository-linked datasets

Dataset availability was analyzed separately because reproducibility depends on access to data, label definitions, split protocols, and preprocessing details. Across the 53 primary studies, 19 reported a dataset, model, project, DOI, or repository link, while 6 were explicitly coded as public, released, open, public-source, or DOI-published in the extraction sheet. Appendix D lists these studies and distinguishes confirmed public datasets from partially confirmed links.

This finding strengthens the reproducibility interpretation of RQ3. Many studies report high performance, but fewer provide enough public evidence to support independent replication, external validation, or adversarial testing. Future studies should publish clear label definitions, annotation protocols, train–validation–test splits, preprocessing scripts, and benchmark baselines whenever legal and ethical constraints allow.

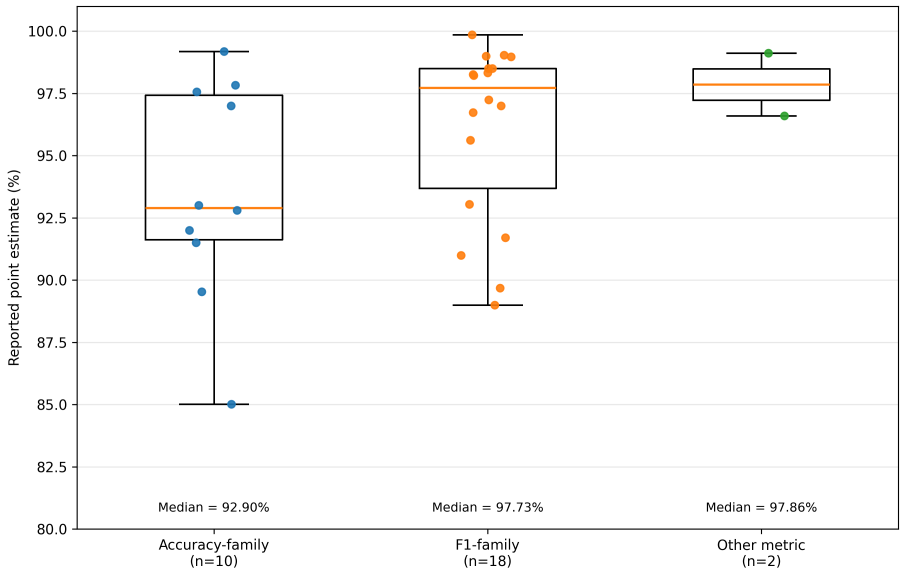


Fig. 9 Descriptive meta-analysis of reported point estimates extracted from the *Main Performance Results* column.

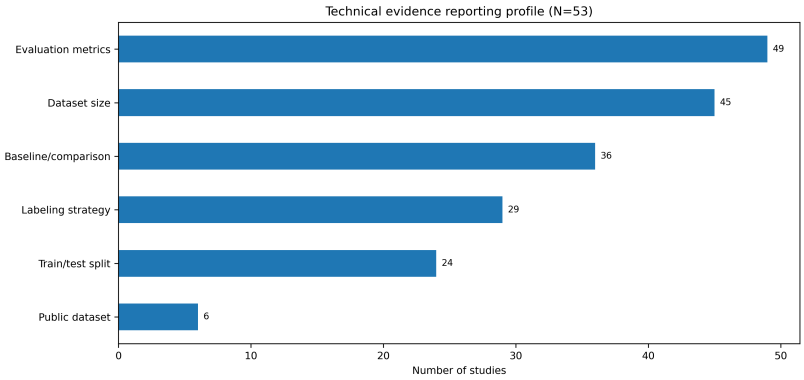


Fig. 10 Technical evidence reporting profile across the 53 primary studies.

4.3.4 AI Evidence Profile and Evidence Maturity

Figure 10 shows that evaluation metrics and dataset size are reported more consistently than public availability, labeling details, explicit splits, and baseline comparisons. This reporting imbalance limits reproducibility and cross-study comparison.

Figure 11 shows that only 9 of 53 studies explicitly tested robustness against evasion, while most studies discussed evasion without controlled testing. This is the central RQ3 limitation: the literature demonstrates detectability more strongly than adversarial resilience.

Table 17 RQ3 evidence summary of effectiveness, robustness, and scalability.

Evidence group	Effectiveness evidence	Scalability evidence	Robustness limitation	Representative studies
Textual and social media detection	Strong task-specific F1 or accuracy is reported in several comment and social media classifiers.	Mostly platform-specific datasets, especially YouTube, Twitter/X, and Indonesian-language content.	Most studies handle obfuscation through preprocessing or model choice, but explicit adaptive text-mutation testing is limited.	Santika et al. (2025); Kamdan et al. (2025); Liem et al. (2025)
Visual and multimodal detection	High website or advertisement detection performance is reported using visual-textual fusion and OCR-based features.	Moderate scalability; evidence often depends on collected websites, screenshots, or advertisement datasets.	Robustness against image manipulation, OCR noise, and visual redesign is usually discussed more than directly stress-tested.	Wang et al. (2022a); Chen et al. (2020); Wang et al. (2022b)
Traffic, domain, and infrastructure analysis	POST-message, domain, hidden-link, and graph-based studies report useful detection or clustering performance.	Stronger scalability appears in domain, traffic, and infrastructure studies using large-scale crawled or network data.	Adaptive domain rotation, crawler resistance, and infrastructure migration remain only partially evaluated.	Guo et al. (2026); Feng et al. (2025); Zhang et al. (2025)
Mobile and app ecosystem detection	Strong performance is reported in app, UI-transition, and underground-economy app detection studies.	Stronger scalability appears in large app-ecosystem studies with thousands of apps or domains.	App repackaging, WebView concealment, and distribution-channel migration are not consistently evaluated as adversarial tests.	Gao et al. (2021); Min et al. (2022); Lu et al. (2025)
Operational and forensic response	Operational studies show practical value for filtering, hidden-page detection, alerting, and incident response.	Scalability is context-specific and depends on deployment environment, monitoring coverage, and response workflow.	Operational bypass is visible, but controlled robustness testing against adaptive attackers remains limited.	Min et al. (2021); Erawan et al. (2024); Harahap and Ridho (2024)
Behavioral, financial-flow, and blockchain analytics	Fraud, betting anomaly, graph anomaly, and blockchain studies report promising detection or inference results.	Moderate to strong depending on graph size, transaction coverage, or platform-scale data availability.	Financial and blockchain role obfuscation is addressed through graph or behavioral modeling, but adversarial validation remains sparse.	Kim et al. (2024); Charitou et al. (2020); Chen et al. (2024)

Figure 12 shows that many studies have moderate or high evidence maturity scores because they report metrics, real-world data, or scalability-oriented evidence. Nevertheless, EMS should not be read as proof of robustness because robustness is only one

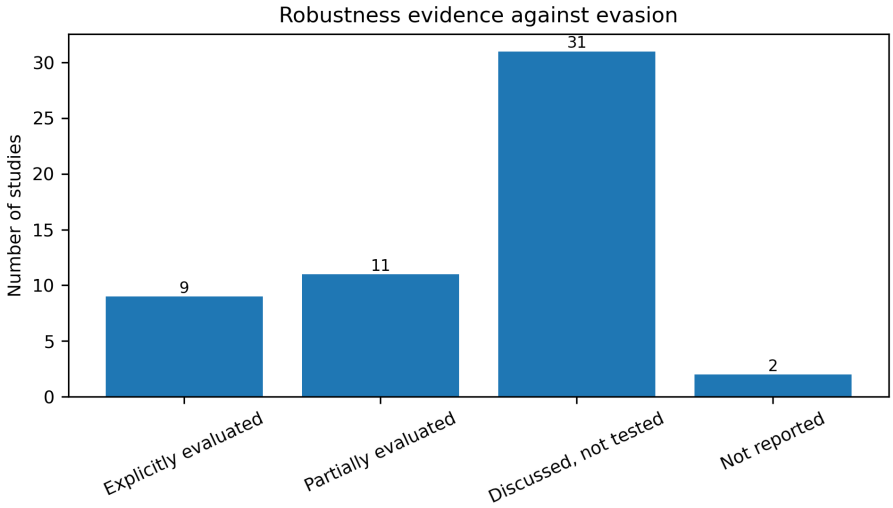


Fig. 11 Robustness evidence against evasion across the 53 primary studies. Explicit evaluation remains limited, while most studies discuss evasion without controlled testing.

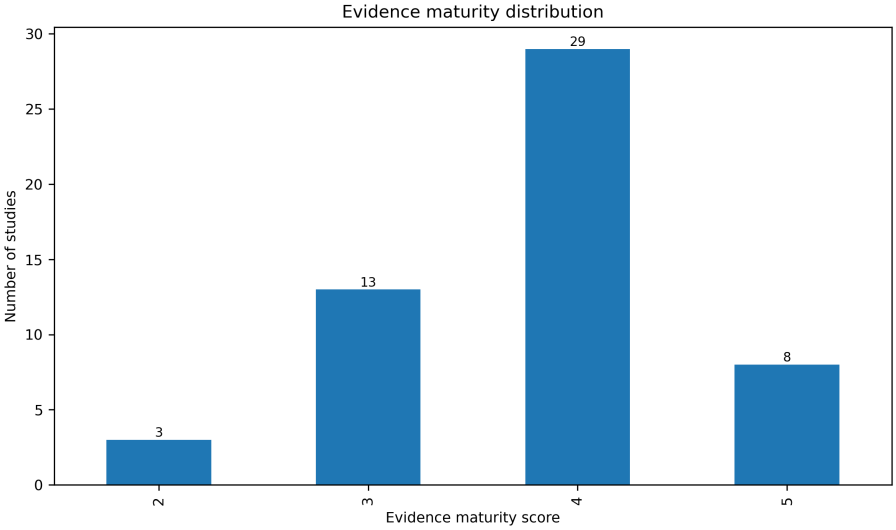


Fig. 12 Evidence maturity score (EMS) distribution across the 53 primary studies.

component of the score. The main RQ3 conclusion remains that adversarial validation is the weakest evidence component.

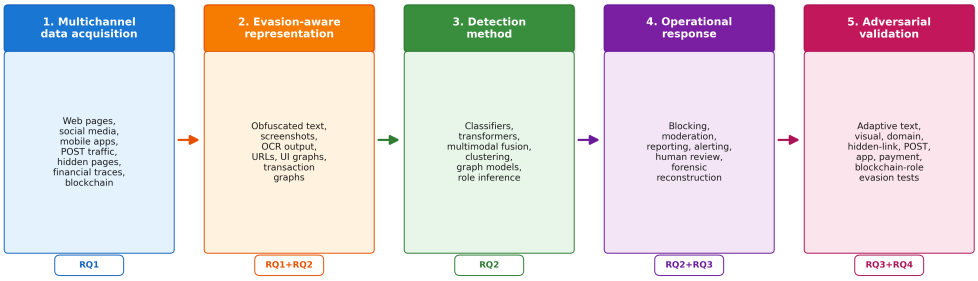


Fig. 13 Proposed adversarially resilient IOG detection framework derived from the RQ1–RQ4 synthesis.

4.4 RQ4: Research Gaps, Open Challenges, and Future Directions

RQ4 synthesizes the unresolved gaps that remain after answering RQ1–RQ3. Five gaps recur across modalities and methods. First, adversarial robustness testing is limited: many studies are evasion-aware by design, but few test adaptive text, visual, URL, app, payment, or graph perturbations. Second, cross-platform, cross-language, and cross-domain generalization remains weak, especially for social-media and language-specific models. Third, detection evidence is fragmented across content, infrastructure, app, payment, operational, and blockchain layers. Fourth, dataset reporting and reproducibility are incomplete. Fifth, operational and longitudinal evaluation is still underdeveloped, even when studies provide practical monitoring or response workflows.

Table 18 maps each gap to resolution paths. The central RQ4 finding is that the field does not mainly lack algorithms; it lacks integrated evaluation practice. Future IOG detection research should prioritize adversarial benchmarks, shared and well-documented datasets, cross-domain testing, multimodal and graph-based integration, longitudinal monitoring, and feedback loops between AI detection, investigation, and enforcement.

4.5 Evidence-based adversarially resilient IOG detection framework

The proposed framework is the integrative output of this review. It translates the RQ1–RQ4 evidence into a layered design logic for future AI-based IOG detection systems. The framework is conceptual rather than an implemented detector; its role is to connect what current studies usually treat separately: evasion surfaces, AI countermeasures, evidence quality, robustness testing, and operational response.

Figure 13 and Table 19 show how future systems can combine multichannel evidence acquisition, modality-aware representation learning, AI/hybrid detection, adversarial validation, operational response, and continuous feedback. The framework contributes a synthesis logic for moving from isolated high-performing classifiers toward resilient, scalable, and deployable AI-based IOG detection ecosystems.

Table 18 Major research gaps and resolution paths for resilient IOG detection.

Rank	Gap	Closest evidence and limitation	Path to resolution
1	Limited adversarial robustness testing across modalities.	Evasion-aware designs appear in website, traffic, social media, hidden-link, and domain studies, but adaptive manipulation is rarely evaluated directly Chen et al. (2020); Wang et al. (2022b); Feng et al. (2025).	Build adversarial benchmarks for text, image, URL, hidden-link, POST-message, mobile, filtering, and graph-based evasion.
2	Limited cross-platform, cross-language, and cross-domain generalization.	Social media and multimodal studies often report strong platform-specific performance but limited external validation Santika et al. (2025); Kamdan et al. (2025); Perdana et al. (2024).	Construct shared multilingual and multimodal datasets across websites, social media, apps, SMS, advertisements, and underground platforms.
3	Fragmented integration of content, infrastructure, app, payment, operational, and blockchain evidence.	Mobile, multimodal, traffic, graph, and blockchain studies cover important layers but do not provide a unified ecosystem detector Gao et al. (2021); Muzakir et al. (2025b); Zhang et al. (2025).	Develop cross-modal architectures linking content, apps, domains, payment services, POST traffic, incident response, and blockchain graphs.
4	Incomplete dataset reporting and limited reproducibility.	Dataset source, labeling, class balance, availability, or split protocol is partially documented in several studies Chen et al. (2020); Zhang et al. (2025); Charitou et al. (2020).	Standardize reporting for dataset source, labels, split, availability, class balance, evaluation protocol, and limitations.
5	Limited operational and longitudinal evaluation.	Honeypot, filtering, defacement, and monitoring studies provide operational evidence, but are weakly integrated with ML robustness evaluation Min et al. (2021); Erawan et al. (2024); Harahap and Ridho (2024).	Combine controlled experiments with longitudinal deployment, incident-response workflows, cloud-specific evaluation, and human-in-the-loop review.

5 Discussion

This section interprets the RQ-based results without repeating the detailed evidence reported in Section 4. The discussion focuses on what the synthesis means for AI research, cybersecurity practice, and future IOG detection design.

Table 19 Condensed evidence-based framework for adversarially resilient AI-based IOG detection.

Framework layer	Core role	Evidence base
AI-ready multichannel data acquisition	Collects evidence from websites, social media, mobile apps, POST traffic, hidden pages, operational artifacts, financial traces, and blockchain networks.	Gao et al. (2021); Feng et al. (2025); Muzakir et al. (2025b)
Evasion-aware representation	Transforms obfuscated text, screenshots, OCR outputs, URLs, app interfaces, graph relations, and transaction traces into detection-ready features.	Wang et al. (2022a,b); Anderson et al. (2026)
AI detection and inference	Applies classifiers, deep learning, transformers, multimodal fusion, clustering, graph models, anomaly rules, and blockchain role inference.	Chen et al. (2020); Liem et al. (2025); Fang et al. (2024)
Operational responses	Links detection outputs to blocking, moderation, reporting, alerting, human review, forensic reconstruction, and incident response.	Min et al. (2021); Erawan et al. (2024); Harahap and Ridho (2024)
Adversarial validation	Evaluates resilience against adaptive text, visual, domain, hidden-link, POST-template, app, payment, and blockchain-role evasion.	Gap derived from RQ3 and RQ4.

5.1 Summary of the review

The review shows that AI-based IOG detection has moved beyond simple keyword or website classification. The 53 primary studies cover text and social media, multimodal website and advertisement evidence, mobile apps, POST-message traffic, domain and infrastructure signals, hidden pages, operational reporting, financial-flow indicators, betting anomalies, and blockchain relationships. This diversity makes IOG detection a multichannel AI problem rather than a single-modality classification task.

Across the four RQs, the main result is consistent: current studies demonstrate that many IOG signals are detectable, but they do not yet provide equally strong evidence of adversarial resilience. The strongest evidence is ordinary task-specific effectiveness; the weakest evidence is controlled robustness testing against adaptive evasion.

5.2 Critical appraisal of the evidence

The key weakness of the literature is the gap between model performance and adversarial validation. Text preprocessing, OCR, multimodal fusion, UI transition graphs, POST-message analysis, and graph-based analytics all respond to meaningful evasion surfaces. However, many studies evaluate these designs under observed or static conditions rather than under controlled adversarial variants such as text mutation, visual redesign, app repackaging, domain rotation, graph perturbation, or temporal drift.

Another concern is comparability. Reported accuracy, precision, recall, F1-score, AUC, blocking rate, and operational success are produced from different tasks and

datasets. Therefore, high values should be read as within-task evidence unless the study also reports external validation, temporal validation, adversarial testing, and reproducibility details. This is why the descriptive meta-analysis in this review focuses on evidence indicators and extractable point estimates rather than a pooled universal performance score.

5.3 Implications of the review

For AI researchers, the findings suggest that future IOG detection should be evaluated as an adaptive, adversarial, and multimodal learning problem. Robustness tests, cross-platform validation, transparent splits, public benchmarks, and threat-model assumptions should become standard reporting elements.

For system designers, the findings imply that resilient detection requires evidence integration. A practical IOG ecosystem may combine social-media promotion, image-based advertisements, newly registered domains, WebView-based apps, hidden defacement pages, payment intermediaries, and blockchain transactions. A modality-specific classifier may detect one layer but miss the broader operational chain.

For cybersecurity and regulatory practice, the review indicates that keyword lists and URL blacklists remain useful but insufficient. More resilient monitoring should combine AI-based content analysis with infrastructure monitoring, mobile-app vetting, SOC alert triage, forensic investigation, and blockchain analytics.

5.4 Limitations of the review

This review is limited by the selected databases, search string, accessibility criteria, and extraction records. Searches were conducted in Scopus, ScienceDirect, IEEE Xplore, and ACM Digital Library on 6 May 2026. Relevant studies indexed elsewhere or using adjacent terms such as *betting*, *wagering*, *fraud*, *abuse infrastructure*, or *underground economy* may have been missed.

The included studies are heterogeneous in modality, task definition, dataset source, model architecture, metric choice, and evaluation protocol. For this reason, a pooled-effect performance meta-analysis was not conducted. The review instead uses structured narrative synthesis, thematic mapping, descriptive evidence indicators, and extractable point-estimate summaries.

Several included studies are partially aligned rather than directly focused on IOG. They were retained only when they provided relevant technical evidence for adjacent evasion or countermeasure mechanisms, such as illicit app detection, hidden-link threats, parental-control filtering, malicious domain verification, graph anomaly detection, or blockchain crime detection. Finally, the review synthesizes reported evidence and does not independently reproduce experiments, relabel datasets, or run new adversarial tests.

6 Conclusion

This SLR synthesized 53 primary studies on AI-based illegal online gambling detection, adversarial evasion, and countermeasure methods. The RQ-based synthesis shows

that IOG evasion is multimodal and infrastructure-dependent; current countermeasures span machine learning, deep learning, transformer models, multimodal fusion, graph analytics, mobile UI transition graphs, anomaly detection, honeypots, forensic monitoring, and blockchain analytics; reported effectiveness is often strong within specific tasks; but explicit robustness, reproducibility, and operational validation remain limited.

The main novelty of this review is an AI-oriented adversarial synthesis that links evasion surfaces, model families, countermeasure workflows, evidence maturity, and remaining research gaps. The review shows that the field should move beyond isolated high-performing classifiers toward integrated, adversarially tested, and operationally deployable AI systems. Future progress requires transparent datasets, standardized reporting, adversarial benchmark construction, cross-domain validation, longitudinal monitoring, threat-model reporting, responsible human–AI decision workflows, and explicit links between model outputs and operational response.

Author contributions. A.A.A. conceived the study, designed the review protocol, formulated the research questions, conducted the literature search, performed study screening, data extraction, quality assessment, evidence synthesis, and wrote the main manuscript text. R.F.A. supervised the study design, refined the review methodology, validated the research scope, and critically reviewed the manuscript. A.J.M. assisted in study screening, data extraction, quality assessment, and verification of the primary-study evidence. M.P. contributed to methodological refinement, interpretation of the technical findings, and critical revision of the manuscript. W.J. provided overall technical supervision, contributed to the interpretation of the information-security implications, and critically reviewed the manuscript. All authors reviewed and approved the final manuscript.

Acknowledgements. The authors would like to thank the Intelligent Robot and Systems (IRoS) Lab, Faculty of Computer Science, Universitas Indonesia, and all parties who supported this research.

Funding. This research is funded by the Indonesian Endowment Fund for Education (LPDP) on behalf of the Indonesian Ministry of Higher Education, Science and Technology and managed under the EQUITY Program (Contract No. 4302/B3/DT.03.08/2025 and 573/PKS/R/UI/2025) and The Indonesia Endowment Fund for Education, Republic of Indonesia (LPDP RI), with grant no LOG-3599/LPDP.3/2025.

Data availability. Not applicable.

Declarations

Ethics approval and consent to participate. Not applicable. This systematic literature review did not involve human or animal subjects and therefore did not require ethics approval.

Consent to publish. Not applicable. This study used publicly available data and did not require approval for publication.

Competing interest. The authors declare that they have no conflict of interest.

Appendix A Comparison with related work

See Table E1.

Appendix B Technical evidence matrix

See Table E2. The table is generated from the structured extraction sheet and uses the same citation keys as the reference list.

Appendix C List of primary studies

See Table E3. This table mirrors the appendix style of the reference article by listing the reviewed studies with automatic continued headers across pages.

Appendix D Published and repository-linked datasets

Table E4 lists studies that reported a dataset, project, model, DOI, or repository link, or that were explicitly coded as public, released, open, or public-source in the extraction sheet. The table distinguishes confirmed public or released datasets from partially confirmed repository links to avoid overstating dataset availability.

Appendix E Descriptive meta-analysis performance table

Table E5 lists the study-level point estimates extracted from the *Main Performance Results* column. These values are descriptive and are not statistically pooled because the underlying tasks and metrics differ substantially.

Table E1: Comparison of this review with closely related review-oriented studies on online gambling and detection contexts.

Study	Review / study type	Main objective and evidence base	Coverage of gambling context	Difference from the present SLR
Cokronegoro and Prasetyo (2025)	Literature review, case study, comparative legal analysis, and critical regulatory analysis.	Analyzes online gambling promoters, endorsements, legal enforcement, and regulatory strategies in Indonesia.	Directly related to online gambling enforcement and platform accountability.	Provides legal and regulatory context, but does not systematically synthesize computational detection, adversarial evasion, robustness testing, or countermeasure architectures.
Gibson et al. (2022)	Systematic review.	Reviews videogame microtransactions, problem gaming, and problem gambling using 19 cross-sectional studies.	Indirectly related through gaming–gambling convergence and gambling-like game mechanics.	Supports the blurred boundary between gaming and gambling, but does not address illegal gambling platforms, detection systems, adversarial evasion, or technical countermeasures.
Tran et al. (2024)	Systematic review and meta-analysis.	Estimates global prevalence of gambling, any-risk gambling, and problematic gambling using 380 representative samples from 68 countries and territories.	Broadly related to gambling prevalence and public-health significance.	Provides prevalence motivation, but does not examine detection models, evasion strategies, robustness, scalability, or computational mitigation workflows.
Present SLR	Systematic literature review using PICOC, PRISMA-style selection, quality assessment, inter-rater reliability, adjudication, and structured data extraction.	Synthesizes adversarial evasion and countermeasure methods using 53 primary studies from Scopus, ScienceDirect, IEEE Xplore, and ACM Digital Library.	Directly focused on illegal gambling websites, social media promotion, mobile apps, traffic, domains, hidden pages, financial traces, and blockchain entities.	Provides the missing technical synthesis by connecting evasion taxonomy, detection architectures, effectiveness evidence, scalability, robustness gaps, and resilient framework design.

Table E2: Architecture-level technical evidence matrix for the 53 primary studies.

Study	Evidence surface and evasion	Countermeasure architecture	Evaluation evidence	Robustness and scalability evidence	Strength
S01 Gao et al. (2021)	*Surface:* Mobile app / UI-based; URL/domain-based; Network-based; Financial transaction-based *Evasion:* Mobile app / UI-level evasion; Platform-level evasion; URL/domain-level evasion; Financial-flow obfuscation (Explicitly Discussed)	*Method:* Web crawling-based detection; Mobile app static/ecosystem analysis; Hybrid detection *Model:* Machine-learning based website identification; code-level similarity clustering; developer-signature clustering; guilt-by-association expansion	*Metrics:* Counts, prevalence, infrastructure counts, payment account counts *Result:* Collected 3,366 unique apps with 5,344 versions and identified over 140K suspicious gambling domains and over 57K app candidates.	*Robustness:* Discussed but Not Tested *Scale:* Strong: five-month collection, thousands of apps, more than 40 million domains, and large expansion results.	Strong
S02 Anderson et al. (2026)	*Surface:* Textual *Evasion:* Textual evasion; Platform-level evasion (Explicitly Discussed)	*Method:* Transformer-based detection; Machine learning-based detection *Model:* IndoBERT; IndoBERTweet; RoBERTa; XLM-RoBERTa	*Metrics:* F1-score; preprocessing comparison *Result:* IndoBERT achieved F1-score 0.9850 when font preprocessing was applied.	*Robustness:* Discussed but Not Tested *Scale:* Moderate: dataset is domain-specific but limited to YouTube comments.	Strong
S03 Charitou et al. (2020)	*Surface:* Financial transaction-based; Behavioral *Evasion:* Behavioral evasion; Financial-flow obfuscation (Explicitly Discussed)	*Method:* Deep learning-based detection; Generative model; Fraud detection *Model:* Sparse Autoencoder; Semi-supervised GAN; Logistic Regression; Random Forest; Multilayer Perceptron; SMOTE; ADASYN	*Metrics:* Not Reported in extracted evidence; comparative performance described qualitatively *Result:* Proposed framework outperformed logistic regression, random forest, and multilayer perceptron, including variants with SMOTE and ADASYN, according to extracted evidence.	*Robustness:* Partially Evaluated *Scale:* Moderate: validated on real-world gambling fraud data and benchmark datasets, but online gambling dataset size is not reported in extracted evidence.	Moderate

Continued on next page

Table E2: Architecture-level technical evidence matrix for the 53 primary studies (continued).

Study	Evidence surface and evasion	Countermeasure architecture	Evaluation evidence	Robustness and scalability evidence	Strength
S04 Chen et al. (2020)	*Surface:* Multimodal; Visual; Textual; Web/content-based *Evasion:* Visual evasion; Textual evasion; URL/domain-level evasion; Platform-level evasion (Explicitly Discussed)	*Method:* Multimodal detection; Hybrid detection; Machine learning-based detection *Model:* Doc2Vec; BoVW; Spa-BoVW; Logistic Regression; SVM; Random Forest	*Metrics:* Accuracy; precision; F-measure *Result:* Accuracy, precision, and F-measure were reported as over 99%; extracted evidence reports gambling website accuracy 99.18% and F-measure 99.13%.	*Robustness:* Discussed but Not Tested *Scale:* Moderate: system is evaluated on crawled websites but dataset size and deployment-scale runtime are not fully reported in extracted evidence.	Strong
S05 Chen et al. (2023)	*Surface:* Mobile app / UI-based *Evasion:* Mobile app / UI-level evasion; Platform-level evasion (Explicitly Discussed)	*Method:* Mobile app static analysis; Graph-based detection; Machine learning-based detection *Model:* DeUEDroid; UI Transition Graph; UTG similarity	*Metrics:* F1-score; classification accuracy; prevalence counts *Result:* Detection F1-score 98.22%; classification accuracy 98.97%; real-world evaluation found 54% of apps in the wild and 11% of app-store apps were UEware.	*Robustness:* Partially Evaluated *Scale:* Strong: 25,717 apps and real-world evaluation on 24,017 apps.	Strong
S06 Chen et al. (2024)	*Surface:* Web/content-based; Financial transaction-based; URL/domain-based *Evasion:* URL/domain-level evasion; Platform-level evasion; Financial-flow obfuscation (Explicitly Discussed)	*Method:* Web crawling-based detection; Financial-flow analysis; Machine learning-based clustering *Model:* DBSCAN; TF-IDF; Selenium crawler; regular expressions; third-party bank-card API	*Metrics:* Counts and clustering output; no accuracy/F1 reported *Result:* 23,142 gambling websites clustered into 386 categories; 2,617 records collected with 535 unique records.	*Robustness:* Discussed but Not Tested *Scale:* Strong for crawling scale and financial-flow data collection; weak for predictive model evaluation.	Moderate

Continued on next page

Table E2: Architecture-level technical evidence matrix for the 53 primary studies (continued).

Study	Evidence surface and evasion	Countermeasure architecture	Evaluation evidence	Robustness and scalability evidence	Strength
S07 Dong et al. (2025)	*Surface:* Behavioral; Financial transaction-based; Graph-based *Evasion:* Behavioral evasion; Financial-flow obfuscation (Explicitly Discussed)	*Method:* Graph-based detection; Rule-based detection; Hybrid detection *Model:* Anomaly Graph Rules; rule-matching algorithm; ML literals	*Metrics:* Accuracy; efficiency *Result:* Not Reported in extracted evidence	*Robustness:* Discussed but Not Tested *Scale:* Moderate: algorithm evaluated on real and synthetic datasets, but gambling-specific dataset/evaluation not reported.	Moderate
S08 Erawan et al. (2024)	*Surface:* Web/content-based; Network-based *Evasion:* Platform-level evasion; URL/domain-level evasion; Web/content-level evasion (Explicitly Discussed)	*Method:* Rule-based detection; Signature/monitoring-based detection; Web crawling-based detection *Model:* Bash script; Linux inotify; Telegram bot notification	*Metrics:* CPU usage; memory use; detection/notification behavior; exact values Partially Reported *Result:* Not Reported in extracted evidence	*Robustness:* Partially Evaluated *Scale:* Moderate: lightweight design targets resource-constrained government infrastructure; scale not fully quantified in extracted evidence.	Moderate
S09 Fang et al. (2024)	*Surface:* URL/domain-based; Network-based; Graph-based *Evasion:* URL/domain-level evasion; Network-level evasion; Platform-level evasion (Explicitly Discussed)	*Method:* Graph-based detection; Deep learning-based detection; Hybrid detection *Model:* HMPN; GraphSAGE/GAT baselines Partially Reported; heterogeneous graph neural network	*Metrics:* Macro-Average F1 *Result:* Macro-Avg-F1 around 91.70%.	*Robustness:* Partially Evaluated *Scale:* Moderate: evaluated on real-world network traffic, but dataset scale is not extracted.	Strong
S10 Feng et al. (2025)	*Surface:* Network-based; Behavioral; Web/content-based *Evasion:* Network-level evasion; Behavioral evasion; URL/domain-level evasion (Explicitly Discussed)	*Method:* Machine learning-based detection; Network traffic classification; Hybrid detection *Model:* Word2Vec; TF-IDF; Hierarchical clustering; Mutual Information; SVM; PSO	*Metrics:* Accuracy; precision; recall; F1-score; macro average; weighted average *Result:* Accuracy 0.9985; macro average precision 0.9982, recall 0.9989, F1 0.9985; weighted average F1 0.9985.	*Robustness:* Partially Evaluated *Scale:* Moderate to Strong: 6,525 POST entries and high test performance, but generalization to other regions/platforms not reported in extracted evidence.	Strong

Continued on next page

Table E2: Architecture-level technical evidence matrix for the 53 primary studies (continued).

Study	Evidence surface and evasion	Countermeasure architecture	Evaluation evidence	Robustness and scalability evidence	Strength
S11 Kim et al. (2024)	*Surface:* Behavioral; Financial transaction-based *Evasion:* Behavioral evasion; Financial-flow obfuscation (Explicitly Discussed)	*Method:* Machine learning-based detection; Fraud detection *Model:* Logistic Regression; Random Forest; Support Vector Machine; K-Nearest Neighbor; Ensemble model	*Metrics:* Accuracy; precision; recall; F1-score; confusion matrix *Result:* Random Forest, KNN, and ensemble models recorded accuracy over 92%; LR and SVM were approximately 80% accurate.	*Robustness:* Discussed but Not Tested *Scale:* Moderate: uses data from 12 betting companies and real-time match application, but exact dataset size is not reported.	Moderate
S12 Ghani et al. (2025)	*Surface:* Network-based; Web/content-based *Evasion:* Network-level evasion; Platform-level evasion; Web/content-level abuse (Explicitly Discussed)	*Method:* Machine learning-based detection; Hybrid detection; Network-layer anomaly detection *Model:* Isolation Forest; Random Forest	*Metrics:* Accuracy; precision; recall; F1-score; confusion matrix *Result:* Random Forest achieved 93% accuracy; reported precision was 88% for DDoS, 97% for Malware or Backdoor, and 100% for Normal traffic; Isolation Forest was reported as 100% accurate for DDoS and malware/backdoor mitigation with minimal false positives.	*Robustness:* Partially Evaluated *Scale:* Moderate: evaluated in a university network case context, but dataset size is not reported.	Moderate
S13 Guo et al. (2024)	*Surface:* Mobile app / UI-based; Platform-level; Web/content-based *Evasion:* Platform-level evasion; Mobile app / UI-level evasion; URL/domain-level evasion (Explicitly Discussed)	*Method:* Web crawling-based detection; Mobile app ecosystem analysis; Hybrid detection *Model:* Telegram bots/search; snowballing; semi-automated crawler; app behavior analysis	*Metrics:* Prevalence counts; category proportions; behavioral statistics; not primarily classification metrics *Result:* Significant prevalence of underground apps on Telegram; apps primarily cater to gambling and pornography services; total subscription counts across channels are described as equivalent to 1% of Telegram's user base.	*Robustness:* Discussed but Not Tested *Scale:* Strong: uses a large-scale Telegram message/channel/app dataset and ecosystem measurement.	Moderate

Continued on next page

Table E2: Architecture-level technical evidence matrix for the 53 primary studies (continued).

Study	Evidence surface and evasion	Countermeasure architecture	Evaluation evidence	Robustness and scalability evidence	Strength
S14 Guo et al. (2026)	*Surface:* URL/domain-based; Web/content-based; Multimodal text/web monitoring *Evasion:* URL/domain-level evasion; Platform-level evasion; Network-level evasion (Explicitly Discussed)	*Method:* Web crawling-based detection; Machine learning-based detection; URL/domain-based detection *Model:* NRDMiner; multilingual BERT text-based classification; machine learning methods	*Metrics:* Accuracy; counts; percentages; temporal activation statistics *Result:* Machine-learning identification achieved 97.57% accuracy; 3.3% of NRDs were flagged as illicit websites; 60% of abusive domains were activated on the day of registration; over 90% initiated illicit operations within one week.	*Robustness:* Partially Evaluated *Scale:* Strong: 27.6 million NRDs across 481 TLDs over 133 days.	Strong
S15 Han et al. (2023)	*Surface:* Mobile app / UI-based; Web/content-based; URL/domain-based; Financial transaction-based *Evasion:* Platform-level evasion; URL/domain-level evasion; Mobile app / UI-level evasion; Financial-flow obfuscation (Explicitly Discussed)	*Method:* Graph-based detection; Web crawling-based detection; Mobile app ecosystem analysis *Model:* Knowledge graph; headless browser crawler; page classification rules; cluster analysis	*Metrics:* Counts, clustering/-community evidence, overlap statistics; standard classifier metrics not reported in extracted evidence *Result:* Constructed a knowledge graph of online gambling apps centered on promotion relationships; detailed numerical performance metrics are not reported in extracted evidence.	*Robustness:* Discussed but Not Tested *Scale:* Strong: starts from 277,816 suspected illegal-industry domains and constructs a knowledge graph.	Moderate
S16 Harahap and Ridho (2024)	*Surface:* Web/content-based; Network-based; Textual *Evasion:* Textual evasion; Web/content-level abuse; Platform-level evasion (Explicitly Discussed)	*Method:* Signature-based detection; Rule-based detection; Keyword-based detection *Model:* Suricata; bigram TF-IDF; Chi-square test	*Metrics:* TPR; TNR; PPV; NPV; accuracy; F1-score *Result:* TPR 0.95, TNR 0.99, PPV 0.99, NPV 0.96, accuracy 0.97, and F1-score 0.97.	*Robustness:* Discussed but Not Tested *Scale:* Moderate: practical IDS deployment context, but dataset size is not reported in extracted evidence.	Strong

Continued on next page

Table E2: Architecture-level technical evidence matrix for the 53 primary studies (continued).

Study	Evidence surface and evasion	Countermeasure architecture	Evaluation evidence	Robustness and scalability evidence	Strength
S17 Huang et al. (2022)	*Surface:* Financial transaction-based; Behavioral; Multimodal *Evasion:* Financial-flow obfuscation; Behavioral evasion; Platform-level evasion (Explicitly Discussed)	*Method:* Machine learning-based detection; Graph-based detection; Financial-flow analysis *Model:* ETHGamDet; LightGBM with memory component; transaction graph; opcode feature extractor	*Metrics:* F1-score; classification metrics *Result:* ETHGamDet achieved F1-score 0.72 for address classification and 0.89 for contract classification.	*Robustness:* Discussed but Not Tested *Scale:* Strong: large-scale released dataset and blockchain-wide detection formulation.	Strong
S18 Izzaturrahim et al. (2025)	*Surface:* Textual *Evasion:* Not Reported (Not Reported)	*Method:* Transformer-based detection; Machine learning-based detection; Deep learning-based detection; Hybrid detection *Model:* IndoBERT; CNN; XGBoost; Wordformer; TF-IDF	*Metrics:* Macro F1; accuracy/-precision/recall likely reported but not fully summarized in extracted evidence *Result:* CNN combined with IndoBERT achieved F1 testing score of 0.9897; performance improved from 0.5267 in the reported comparison.	*Robustness:* Not Reported *Scale:* Moderate: practical multi-source comment detection, but dataset size and availability are not clearly reported in extracted evidence.	Moderate
S19 Kamdan et al. (2025)	*Surface:* Textual *Evasion:* Textual evasion; Platform-level evasion (Explicitly Discussed)	*Method:* Transformer-based detection; Deep learning-based detection; Machine learning-based detection *Model:* IndoBERT	*Metrics:* Accuracy; precision; recall; F1-score; confusion matrix *Result:* IndoBERT achieved 98.26% accuracy; extracted evidence states precision 98.28%, recall 98.26%, and F1 98.26%.	*Robustness:* Discussed but Not Tested *Scale:* Moderate: platform-specific data collection from YouTube, but dataset size is not reported in extracted evidence.	Strong
S20 Ke et al. (2022)	*Surface:* Textual *Evasion:* Textual evasion; Platform-level evasion (Explicitly Discussed)	*Method:* Transformer-based detection; Machine learning-based detection; Keyword-evasion detection *Model:* DC-BERT; cross-corpus similarity analysis	*Metrics:* Detection accuracy *Result:* The proposed framework achieved detection accuracy of 91.5%.	*Robustness:* Discussed but Not Tested *Scale:* Moderate: darknet corpus from six Tor websites, but exact scale is not reported in extracted evidence.	Moderate

Continued on next page

Table E2: Architecture-level technical evidence matrix for the 53 primary studies (continued).

Study	Evidence surface and evasion	Countermeasure architecture	Evaluation evidence	Robustness and scalability evidence	Strength
S21 Prakoso et al. (2025)	*Surface:* Textual *Evasion:* Textual evasion; Platform-level evasion (Explicitly Discussed)	*Method:* Transformer-based detection; Machine learning-based detection *Model:* IndoBERT Base; IndoBERT Lite; IndoBERT Large; mBERT; Indonesian RoBERTa 522M; Indonesian RoBERTa 1.5G	*Metrics:* Accuracy; Precision; Recall; F1-score *Result:* All models achieved F1 above 0.97; IndoBERT Base performed best with F1 = 0.9904.	*Robustness:* Discussed but Not Tested *Scale:* Moderate: platform-specific real comments are used, but dataset size is not visible in extracted evidence.	Strong
S22 Kurniawan and Triayudi (2024)	*Surface:* Web/content-based; Network-based *Evasion:* Platform-level evasion; Network-level evasion; Web/content-level evasion (Explicitly Discussed)	*Method:* Signature-based detection; Rule-based detection; Hybrid detection; Digital forensic countermeasure *Model:* Wazuh; Velociraptor; MITRE ATT&CK; Telegram notification	*Metrics:* Average response time; alert count; isolation success *Result:* Wazuh detected attacks in real time with average response time of 1.2 seconds per alert; Velociraptor isolation succeeded on AWS but not GCP.	*Robustness:* Partially Evaluated *Scale:* Moderate: tested on AWS and GCP reconstruction scenarios.	Moderate
S23 Liberato et al. (2025)	*Surface:* URL/domain-based; Network-based *Evasion:* Platform-level evasion; Network-level evasion; URL/domain-level weakness (Explicitly Discussed)	*Method:* Rule-based detection; DNS filtering; URL/domain-based filtering; Countermeasure evaluation *Model:* TP-Link parental control; Netgear parental control; ASUS parental control; OpenDNS; DNS0.eu; Norton Family	*Metrics:* Blocking rate; underblocking; overblocking; inconsistency *Result:* Some systems fail to block up to 96% of sensitive domains; ASUS allows access to over 95% of gambling websites.	*Robustness:* Partially Evaluated *Scale:* Moderate: broad system comparison, but gambling is one category among multiple sensitive categories.	Moderate

Continued on next page

Table E2: Architecture-level technical evidence matrix for the 53 primary studies (continued).

Study	Evidence surface and evasion	Countermeasure architecture	Evaluation evidence	Robustness and scalability evidence	Strength
S24 Liem et al. (2025)	*Surface:* Textual *Evasion:* Textual evasion; Platform-level evasion (Explicitly Discussed)	*Method:* Machine learning-based detection; Deep learning-based detection; Transformer-based detection *Model:* BERT; IndoBERT; RoBERTa; Random Forest; XGBoost; CNN; LSTM; FastText; TF-IDF; Word2Vec	*Metrics:* Accuracy; Precision; Recall; F1-score; F-beta score *Result:* IndoBERT achieved 98.53% accuracy, 97.24% F1-score, 95.65% recall, 98.88% precision, and 98.21% F-beta score.	*Robustness:* Discussed but Not Tested *Scale:* Moderate	Strong
S25 Lu et al. (2025)	*Surface:* Mobile app / UI-based; Multimodal *Evasion:* Mobile app / UI-level evasion; Platform-level evasion (Explicitly Discussed)	*Method:* Graph-based detection; Mobile app static analysis; Hybrid detection; Deep learning-based detection *Model:* TACDROID; Graph Autoencoder; User Interface Transition Graph; link prediction	*Metrics:* F1-score; classification accuracy *Result:* TACDROID achieved F1-score 96.73% for illicit app detection and 92.09% for classification; UTG construction F1-score 83.48%.	*Robustness:* Partially Evaluated *Scale:* Strong: large-scale illicit app evaluation and comparison with baselines.	Strong
S26 Min et al. (2021)	*Surface:* Textual; URL/domain-based; Mobile environment *Evasion:* Textual evasion; URL/domain-level evasion; Platform-level evasion (Explicitly Discussed)	*Method:* Web crawling-based detection; Keyword-based detection; Machine learning-based detection; Countermeasure strategy *Model:* Readable Transformation Technique; n-gram features; machine learning algorithms	*Metrics:* URL discovery counts; accessible URL count; blocking result *Result:* 5,253 IOG-related URLs were discovered, and 3,745 accessible URLs were blocked by authorities.	*Robustness:* Discussed but Not Tested *Scale:* Strong: applied to nine months of KRA data and thousands of URLs.	Strong

Continued on next page

Table E2: Architecture-level technical evidence matrix for the 53 primary studies (continued).

Study	Evidence surface and evasion	Countermeasure architecture	Evaluation evidence	Robustness and scalability evidence	Strength
S27 Min et al. (2022)	*Surface:* Textual; URL/domain-based; Mobile environment *Evasion:* Textual evasion; URL/domain-level evasion (Explicitly Discussed)	*Method:* Machine learning-based detection; Text mining; URL/domain-based detection; Countermeasure strategy *Model:* Artificial intelligence-based classifier; text mining; binary classification	*Metrics:* Accuracy *Result:* Detection accuracy rate of 97%.	*Robustness:* Discussed but Not Tested *Scale:* Strong: grounded in national-scale spam reporting context.	Strong
S28 Muzakir et al. (2025b)	*Surface:* Multimodal; Textual; Visual; Audio/video-derived text *Evasion:* Textual evasion; Visual evasion; Platform-level evasion (Explicitly Discussed)	*Method:* Multimodal detection; Deep learning-based detection; Transformer-based detection; Hybrid detection *Model:* RNN; CNN; BERT; AutoGluon AutoML; OCR	*Metrics:* Accuracy; Precision; Recall; F1-score *Result:* RNN achieved accuracy 93.07%, precision 93.22%, recall 92.87%, and F1-score 93.04%; CNN achieved accuracy 92.80%; AutoML achieved accuracy 90.19%; BERT achieved accuracy 68.12%.	*Robustness:* Discussed but Not Tested *Scale:* Strong: multiple platforms, multimodal extraction, and over 15K entries.	Strong
S29 Muzakir et al. (2025a)	*Surface:* Textual; Multimodal-derived text *Evasion:* Textual evasion; Platform-level evasion (Explicitly Discussed)	*Method:* Machine learning-based detection; Deep learning-based detection *Model:* AutoML; CNN; TF-IDF; Word2Vec	*Metrics:* Accuracy; Precision; Recall *Result:* CNN achieved 92.80% accuracy, outperforming AutoML at 90.19%, with higher precision and recall.	*Robustness:* Discussed but Not Tested *Scale:* Strong: multi-platform dataset with 15,286 annotated sentences.	Strong
S30 Perdana et al. (2024)	*Surface:* Textual *Evasion:* Textual evasion; Platform-level evasion (Explicitly Discussed)	*Method:* Machine learning-based detection; Deep learning-based detection; Text mining *Model:* Random Forest; Logistic Regression; CNN; TF-IDF	*Metrics:* Recall; Precision; Accuracy/F1 Not Reported in extracted evidence *Result:* TF-IDF and Random Forest achieved recall 0.958 and precision 0.966.	*Robustness:* Discussed but Not Tested *Scale:* Moderate: thousands of tweets from a real platform.	Strong

Continued on next page

Table E2: Architecture-level technical evidence matrix for the 53 primary studies (continued).

Study	Evidence surface and evasion	Countermeasure architecture	Evaluation evidence	Robustness and scalability evidence	Strength
S31 Wang et al. (2022a)	*Surface:* Visual; Textual; Web/content-based; Multimodal *Evasion:* Visual evasion; Textual evasion; Web/content-level evasion (Explicitly Discussed)	*Method:* Multimodal detection; Deep learning-based detection; Semi-supervised learning; OCR-based detection *Model:* CNN-5; TextRNN; ResNet34; OCR; co-training	*Metrics:* Precision; Recall; F1-score *Result:* Co-TextRNN achieved overall Precision 0.9835, Recall 0.9833, and F1 0.9833 after co-training; Co-CNN-5 achieved F1 0.9000.	*Robustness:* Discussed but Not Tested *Scale:* Moderate: uses a self-collected dataset and unlabeled data, but scale is moderate.	Strong
S32 Rojas et al. (2025)	*Surface:* URL/domain-based; Network-based; Mobile app / UI-based *Evasion:* Platform-level evasion; Network-level evasion; URL/domain-level evasion (Explicitly Evaluated)	*Method:* Rule-based detection; Signature-based/filtering countermeasure; URL/domain-based filtering; Network-based countermeasure *Model:* Protective DNS; Android parental control apps; Functionality Index	*Metrics:* Filtering accuracy; false positives/false negatives; Functionality Index; evasion resistance *Result:* Results show inconsistent effectiveness; some tools provide strong performance while others fail to provide basic protections; PDNS services show inconsistent accuracy with false positives and low adult filtering effectiveness.	*Robustness:* Explicitly Evaluated *Scale:* Moderate: 1,427 categorized websites and multiple deployed tools/services.	Moderate
S33 Sachan et al. (2023)	*Surface:* URL/domain-based; Network-based; Financial transaction-based *Evasion:* URL/domain-level evasion; Financial-flow obfuscation; Platform-level evasion (Partially Reported)	*Method:* Machine learning-based detection; URL/domain-based detection; Fraud detection *Model:* TPOT AutoML; supervised ML; unsupervised clustering; cosine similarity	*Metrics:* Balanced accuracy; counts of malicious DNs; persistent behavior counts *Result:* Supervised approach achieved 89.53% balanced accuracy; 73,769 blockchain DNs showed malicious behavior at least once; 34,171 showed persistent malicious behavior; 2,479 malicious blockchain DNs over time.	*Robustness:* Partially Evaluated *Scale:* Strong for DNS-scale data and blockchain DNs; partial for gambling-specific scope.	Moderate

Continued on next page

Table E2: Architecture-level technical evidence matrix for the 53 primary studies (continued).

Study	Evidence surface and evasion	Countermeasure architecture	Evaluation evidence	Robustness and scalability evidence	Strength
S34 Sánchez et al. (2026)	*Surface:* Visual; Web/content-based *Evasion:* Visual evasion; Textual evasion; Platform-level evasion (Explicitly Discussed)	*Method:* Deep learning-based detection; Hybrid detection; Visual detection *Model:* GADIA; ResNet50; SVM; Random Forest; EfficientNet (related comparison context)	*Metrics:* Accuracy; Precision; Recall; F1-score; AUC *Result:* ResNet50 achieved 85.01% accuracy and 90.42% recall.	*Robustness:* Discussed but Not Tested *Scale:* Moderate: real-time pipeline motivation and balanced dataset; deployment not fully proven in extracted evidence.	Strong
S35 Santika et al. (2025)	*Surface:* Textual *Evasion:* Not Reported (Not Reported)	*Method:* Transformer-based detection; Semi-supervised learning; Machine learning-based detection *Model:* IndoBERT; SVM	*Metrics:* Accuracy; Precision; Recall; F1-score *Result:* Initial IndoBERT achieved accuracy 0.953, precision 0.945, recall 0.977, and F1 0.961; semi-supervised IndoBERT achieved accuracy 0.986, precision 0.991, recall 0.980, and F1 0.985.	*Robustness:* Not Reported *Scale:* Moderate: uses over 10K comments and semi-supervised learning but no explicit deployment test.	Moderate
S36 Santosa et al. (2025)	*Surface:* Textual *Evasion:* Textual evasion; Platform-level evasion; URL/domain-level evasion (Explicitly Discussed)	*Method:* Machine learning-based detection; Transformer-based detection; Hybrid annotation workflow *Model:* Bernoulli Naive Bayes; Random Forest; IndoBERT; RoBERTa; Mistral-7B-Instruct	*Metrics:* Accuracy; weighted F1-score *Result:* IndoBERT achieved accuracy 95.63% and weighted F1-score 95.62% $\pm$ 0.0084.	*Robustness:* Discussed but Not Tested *Scale:* Moderate: platform-specific dataset collected from trending Shorts; scale not explicitly reported in extracted evidence.	Strong
S37 Sun et al. (2022)	*Surface:* Financial transaction-based; Graph-based; Network-based *Evasion:* Financial-flow obfuscation; Behavioral evasion; Network-level evasion (Explicitly Discussed)	*Method:* Graph-based detection; Deep learning-based detection; Fraud detection; Financial-flow analysis *Model:* GAME-BC; Graph Attention Network	*Metrics:* Classification accuracy *Result:* Experimental results are reported as better classification accuracy, but exact value is not explicitly stated in extracted evidence.	*Robustness:* Discussed but Not Tested *Scale:* Partially Reported: blockchain data are large-scale by nature, but exact dataset size is not in extracted evidence.	Moderate

Continued on next page

Table E2: Architecture-level technical evidence matrix for the 53 primary studies (continued).

Study	Evidence surface and evasion	Countermeasure architecture	Evaluation evidence	Robustness and scalability evidence	Strength
S38 Teppap et al. (2024)	*Surface:* Web/content-based; URL/domain-based; Textual *Evasion:* URL/domain-level evasion; Web/content-level evasion; Platform-level evasion (Explicitly Discussed)	*Method:* Web crawling-based detection; Rule-based detection; Automated reporting countermeasure *Model:* BeautifulSoup; Google Search extraction; LINE Notify	*Metrics:* Accuracy; Precision; Recall; F1-score *Result:* Proposed model achieved accuracy 0.89, precision 0.83, recall 1.00, and F1-score 0.91; baseline accuracy 0.53 and F1-score 0.69.	*Robustness:* Discussed but Not Tested *Scale:* Moderate: evaluated thousands of URLs and supports automatic notification.	Strong
S39 Wang and Shang (2024)	*Surface:* URL/domain-based; Visual; Network-based *Evasion:* URL/domain-level evasion; Platform-level evasion; Network-level evasion (Explicitly Discussed)	*Method:* URL/domain-based detection; Graph-based detection; Hybrid detection *Model:* Graph computing; image similarity; clustering analysis; domain name pool	*Metrics:* Not Reported *Result:* Not Reported	*Robustness:* Discussed but Not Tested *Scale:* Moderate: aims at network-wide domain scanning and dynamic tracking, but exact scale not reported in extracted evidence.	Moderate
S40 Wang et al. (2022b)	*Surface:* Visual; Textual; Web/content-based; Multimodal *Evasion:* Visual evasion; Textual evasion; Web/content-level evasion (Explicitly Discussed)	*Method:* Multimodal detection; Deep learning-based detection; Hybrid detection; OCR-based detection *Model:* ResNet34; OCR; Word2Vec; Bi-LSTM; self-attention; late fusion; CNN-5; CNN-7	*Metrics:* Accuracy; Precision; Recall; F1-score *Result:* Final method reported accuracy, precision, recall, and F1-score all over 99%; ResNet34 image classifier alone achieved accuracy 0.9200 and macro F1 0.9199.	*Robustness:* Discussed but Not Tested *Scale:* Moderate: self-collected balanced dataset and computational model comparison.	Strong

Continued on next page

Table E2: Architecture-level technical evidence matrix for the 53 primary studies (continued).

Study	Evidence surface and evasion	Countermeasure architecture	Evaluation evidence	Robustness and scalability evidence	Strength
S41 Xiong et al. (2025)	*Surface:* Web / crawler-based; HTML/link structure; visual inspection *Evasion:* Hidden-link evasion; crawler-level evasion; black-hat SEO evasion (Explicitly Evaluated)	*Method:* Security measurement; adversarial detection; crawler/LLM robustness evaluation *Model:* Crawler mechanisms; Large Language Models; visual inspection mitigation framework	*Metrics:* Detection effectiveness of hidden-link threats; qualitative/empirical evaluation metrics *Result:* Partially Reported: evaluation demonstrates framework efficacy, but exact numeric metrics are not available in the extracted evidence.	*Robustness:* Discussed but Not Tested for gambling-specific evasion; Explicitly Evaluated for hidden-link evasion *Scale:* Partially Reported: real-world web pages are used, but scale is not fully specified in extracted evidence.	Moderate
S42 Zhang et al. (2023)	*Surface:* Website / domain; HTML/resource; textual semantic *Evasion:* SEO/promotion-channel evasion; domain/resource/semantic evasion (Explicitly Discussed)	*Method:* Hybrid ML/DL detection; multi-level feature fusion *Model:* DRSDetector; Transformer Encoder; LightGBM; HAN	*Metrics:* Accuracy *Result:* DRSDetector achieved 97.83% accuracy.	*Robustness:* Discussed but Not Tested *Scale:* Partially Reported: automatic website detection is evaluated, but operational deployment scale is not fully specified.	Strong
S43 Zhang et al. (2025)	*Surface:* Network traffic / POST messages; behavioral *Evasion:* Network/traffic-level evasion; behavioral camouflage (Explicitly Discussed)	*Method:* Traffic-based detection; clustering; user behavior analysis *Model:* Word2Vec; TF-IDF; SDAE; Agglomerative Clustering; cosine similarity consensus voting	*Metrics:* Detection and clustering performance; user behavior classification outcomes *Result:* Partially Reported: extracted evidence reports improved detection and classification but does not provide all numeric metrics.	*Robustness:* Discussed but Not Tested *Scale:* Partially Reported: real-world network environment is used, but full scale is unavailable in extracted evidence.	Strong

Continued on next page

Table E2: Architecture-level technical evidence matrix for the 53 primary studies (continued).

Study	Evidence surface and evasion	Countermeasure architecture	Evaluation evidence	Robustness and scalability evidence	Strength
S44 Zhao (2023)	*Surface:* Web / JavaScript / rendered HTML / visual comparison *Evasion:* JavaScript/browser-level evasion; content cloaking; SEO evasion (Explicitly Evaluated)	*Method:* Browser-based detection; visual comparison; web security measurement *Model:* PACTOR	*Metrics:* Detected incidents, control scripts, illicit content providers, URL-safety evasion evidence *Result:* 2,454 incidents studied; proactive defacements can evade existing URL safety-checking tools.	*Robustness:* Explicitly Evaluated *Scale:* Moderate: two-month empirical study across thousands of incidents.	Strong
S45 Zhao et al. (2024)	*Surface:* Mobile app / iOS; UI/review/recommendation/code similarity *Evasion:* App-store review evasion; UI/-functionality masking; trigger-based evasion (Explicitly Evaluated)	*Method:* Mobile app detection; app-store security analysis; similarity-based detection *Model:* Mask-Catcher	*Metrics:* Recall, precision *Result:* Mask-Catcher achieved high recall and precision; prior QA evidence reported recall 99.12% and precision 100%.	*Robustness:* Explicitly Evaluated *Scale:* Partially Reported: real-world App Store data used; operational scale not fully available in extracted evidence.	Strong
S46 Wang et al. (2023)	*Surface:* Website / domain / third-party service / infrastructure *Evasion:* Cloaking; third-party service abuse; infrastructure-level evasion (Explicitly Evaluated)	*Method:* Community-level detection; graph/infrastructure-based detection; ML classification *Model:* IDTracker	*Metrics:* F1-score; number of illicit communities/websites discovered *Result:* F1-score 0.8968; deployment identified 6,830 illicit communities containing 165,378 illicit websites.	*Robustness:* Explicitly Evaluated *Scale:* Strong: three-month ISP deployment with large-scale communities and websites.	Strong
S47 Wang et al. (2025a)	*Surface:* Social media / textual / image / campaign metadata *Evasion:* Textual and visual social-media evasion; campaign-level evasion (Explicitly Evaluated)	*Method:* Social-media content moderation; ML-based detection; campaign analysis *Model:* PIP hunter; PIP analyzer	*Metrics:* Counts, category distribution, account/contact counts, survival rate *Result:* In six months, 12M distinct PIPs were found across 10 illicit categories; about 90% survived the first two months after publication.	*Robustness:* Explicitly Evaluated *Scale:* Strong: six-month large-scale Twitter study with millions of PIPs.	Strong

Continued on next page

Table E2: Architecture-level technical evidence matrix for the 53 primary studies (continued).

Study	Evidence surface and evasion	Countermeasure architecture	Evaluation evidence	Robustness and scalability evidence	Strength
S48 Wang et al. (2025b)	*Surface:* Blockchain / transaction graph / behavioral role *Evasion:* Blockchain anonymity; role-concealment evasion; collusion (Explicitly Evaluated)	*Method:* Graph neural network; unsupervised role identification; blockchain forensic detection *Model:* Voting influence metric; multi-view GNN; structural entropy; self-supervised contrastive learning	*Metrics:* Match with judicially confirmed identities; role partitioning effectiveness *Result:* The method achieved higher match with judicially confirmed identities than existing methods according to extracted evidence.	*Robustness:* Explicitly Evaluated *Scale:* Partially Reported: multi-chain experiments demonstrate generalizability, but scale is not fully visible.	Strong
S49 Wang et al. (2025c)	*Surface:* Graph / behavioral / financial fraud *Evasion:* Behavioral evasion / fine-grained fraud heterogeneity (Conceptual / Indirect)	*Method:* Graph neural network; prototype learning; fraud detection *Model:* APGNN	*Metrics:* Fraud detection performance metrics; interpretability results *Result:* Extensive experiments show better fraud detection and more understandable results, but extracted evidence does not show numeric metrics.	*Robustness:* Discussed but Not Tested for gambling-specific evasion *Scale:* Partially Reported: tested on four real-world datasets.	Moderate
S50 Wen et al. (2024)	*Surface:* Website / image / text / HTML / OCR *Evasion:* Visual/-textual multimodal evasion; camouflage; single-modal deception (Explicitly Discussed)	*Method:* Multimodal detection; semi-supervised learning; data augmentation *Model:* MEDAL; tri-training; multimodal mutual assistance module	*Metrics:* Accuracy, precision, recall, F1 *Result:* Experimental results show MEDAL performs well on accuracy, precision, recall, and F1; exact values are not in extracted evidence.	*Robustness:* Discussed but Not Tested *Scale:* Partially Reported: uses unlabeled data for data augmentation; scale not shown in extracted evidence.	Strong
S51 Wu et al. (2025)	*Surface:* Software repository / SEO / URL redirection / package metadata *Evasion:* Blackhat SEO evasion; repository abuse; redirection-based promotion (Explicitly Evaluated)	*Method:* Large-scale abuse detection; supply-chain/security measurement *Model:* RepSEO detection tool	*Metrics:* Counts of abusive packages; ecosystem/supply-chain measurements *Result:* 3,801,682 abusive packages were discovered; vulnerabilities and packages were reported to repository operators.	*Robustness:* Explicitly Evaluated for RepSEO abuse; Discussed but Not Tested for gambling-specific evasion *Scale:* Strong: ten-year, multi-repository dataset and millions of abusive packages.	Strong

Continued on next page

Table E2: Architecture-level technical evidence matrix for the 53 primary studies (continued).

Study	Evidence surface and evasion	Countermeasure architecture	Evaluation evidence	Robustness and scalability evidence	Strength
S52 Xiong (2022)	*Surface:* Website / text / access path / platform attributes *Evasion:* Domain rotation; blacklist evasion (Explicitly Discussed)	*Method:* Similarity-based detection; case-based recognition *Model:* Sensitive feature similarity method	*Metrics:* Recognition performance; qualitative accuracy statement *Result:* Experiment shows better results when historical cases are fewer and case data quality is better; exact metrics not visible in extracted evidence.	*Robustness:* Discussed but Not Tested *Scale:* Partially Reported: intended for cases where typical illegal website cases are difficult to obtain.	Moderate
S53 Xiong et al. (2024)	*Surface:* Web / crawler / hidden link / visual inspection *Evasion:* Hidden-link evasion; crawler-level evasion; black-hat SEO evasion (Explicitly Evaluated)	*Method:* Adversarial detection; crawler security; visual inspection countermeasure *Model:* Element visual inspection mitigation framework	*Metrics:* Detection efficacy of mitigation framework; attack impact analysis *Result:* Evaluation demonstrates framework efficacy in detecting and addressing hidden-link threats; numeric values not visible in extracted evidence.	*Robustness:* Explicitly Evaluated *Scale:* Partially Reported: real-world web pages assessed, but scale not visible in extracted evidence.	Moderate

Table E3: Selected primary studies included in the SLR.

ID	Primary study	Year	Publication type	Source / venue
S01	Demystifying Illegal Mobile Gambling Apps Gao et al. (2021)	2021	Conference paper	Proceedings of the Web Conference 2021 (WWW '21)
S02	Detecting Online Gambling Promotion in Youtube Comments Using Transformer-Based Models Anderson et al. (2026)	2026	Conference paper	2026 International Seminar on Intelligent Business and Edge-Computing Research (ISIBER)
S03	Semi-supervised GANs for Fraud Detection Charitou et al. (2020)	2020	Conference paper	2020 International Joint Conference on Neural Networks (IJCNN)
S04	Automatic Detection of Pornographic and Gambling Websites Based on Visual and Textual Content Using a Decision Mechanism Chen et al. (2020)	2020	Journal article	Sensors
S05	DeUEDroid: Detecting Underground Economy Apps Based on UTG Similarity Chen et al. (2023)	2023	Conference paper	Proceedings of the 32nd ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA)
S06	Data mining based analysis of online gambling sites and illicit financial flows Chen et al. (2024)	2024	Conference paper	International Conference on Cloud Computing and Big Data (ICCBD)
S07	Anomaly Detection Using Graph Anomaly Rules Dong et al. (2025)	2025	Journal article	Big Data Mining and Analytics
S08	A Lightweight Design Approach to Detect Malicious Injections in Government Websites: Combating Hidden Online Gambling Pages Erawan et al. (2024)	2024	Conference paper	2024 Beyond Technology Summit on Informatics International Conference (BTS-I2C)
S09	Detecting and Exploring Malicious Websites through Multi-Message Passing Heterogeneous Neural Network Fang et al. (2024)	2024	Conference paper	IEEE Symposium on Computers and Communications (ISCC)
S10	Anomalous Traffic Identification Method for POST Messages Based on Gambling Website Templates Feng et al. (2025)	2025	Journal article	Computer Science and Information Systems
S11	AI-based betting anomaly detection system to ensure fairness in sports and prevent illegal gambling Kim et al. (2024)	2024	Journal article	Scientific Reports

Continued on next page

Table E3: Selected primary studies included in the SLR (continued).

ID	Primary study	Year	Publication type	Source / venue
S12	AI-Driven Network Security: Detecting and Mitigating DDoS, Malware, and Backdoor Attacks with Isolation and Random Forest Algorithm Ghani et al. (2025)	2025	Conference paper	2025 International Conference on Smart Computing, IoT and Machine Learning (SIML)
S13	Beyond App Markets: Demystifying Underground Mobile App Distribution Via Telegram Guo et al. (2024)	2024	Journal article	Proceedings of the ACM on Measurement and Analysis of Computing Systems
S14	Darkness at dawn: understanding illicit websites in newly registered domain names Guo et al. (2026)	2026	Journal article	Cybersecurity
S15	Measurement of Illegal Android Gambling App Ecosystem From Joint Promotion Perspective Han et al. (2023)	2023	Conference paper	2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA)
S16	Detection of Online Gambling Web Defacement in University Domains Using Attack Signatures Harahap and Ridho (2024)	2024	Conference paper	2024 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD)
S17	Who is Gambling? Finding Cryptocurrency Gamblers Using Multi-modal Retrieval Methods Huang et al. (2022)	2022	Journal article	International Journal of Multimedia Information Retrieval
S18	Transformer Based Hybridization Text Classifiers In Online Gambling Spam Identification Case Study Izzaturrahim et al. (2025)	2025	Conference paper	2025 International Conference on Computer, Control, Informatics and its Applications (IC3INA)
S19	Performance Analysis of IndoBERT for Detection of Online Gambling Promotion in YouTube Comments Kamdan et al. (2025)	2025	Conference paper	Engineering Proceedings
S20	An Unsupervised Detection Framework for Chinese Jargons in the Darknet Ke et al. (2022)	2022	Conference paper	Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining (WSDM '22)

Continued on next page

Table E3: Selected primary studies included in the SLR (continued).

ID	Primary study	Year	Publication type	Source / venue
S21	Detecting Online Gambling Promotional Comments on YouTube Using Indonesian BERT and Multilingual BERT Prakoso et al. (2025)	2025	Conference paper	2025 International Conference on Artificial Intelligence and Technological Solutions (ICAITech)
S22	Reconstruction and Detection of Gambling Web Defacement Attack Using Wazuh and Velociraptor Kurniawan and Triayudi (2024)	2024	Conference paper	2024 International Conference on Information Technology Research and Innovation (ICITRI)
S23	To Block Or Not To Block? Evaluating Parental Controls Across Routers, DNS Services, and Software Liberato et al. (2025)	2025	Conference paper	TMA 2025 - Proceedings of the 9th Network Traffic Measurement and Analysis Conference
S24	Deep Learning And Machine Learning For Indonesian Online Gambling Website Promotion Detection Liem et al. (2025)	2025	Conference paper	Procedia Computer Science
S25	TACDROID: Detection of Illicit Apps Through Hybrid Analysis of UI-Based Transition Graphs Lu et al. (2025)	2025	Conference paper	2025 IEEE/ACM 47th International Conference on Software Engineering (ICSE)
S26	Honeypot System for Automatic Reporting of Illegal Online Gambling Sites Utilizing SMS Spam Min et al. (2021)	2021	Conference paper	2021 World Automation Congress (WAC)
S27	Detecting Illegal Online Gambling (IOG) Services in the Mobile Environment Min et al. (2022)	2022	Journal article	Security and Communication Networks
S28	A Deep Learning and AutoML-Based Multimodal Text Extraction Framework for Detecting Online Gambling Advertisements in Indonesian Social Media Muzakir et al. (2025b)	2025	Journal article	International Journal of Safety and Security Engineering
S29	Text-Based Detection of Online Gambling Advertisements on Social Media Using AutoML and CNN Muzakir et al. (2025a)	2025	Conference paper	2025 2nd International Conference on Information System and Information Technology (ICISIT)

Continued on next page

Table E3: Selected primary studies included in the SLR (continued).

ID	Primary study	Year	Publication type	Source / venue
S30	Detecting Online Gambling Promotions on Indonesian Twitter Using Text Mining Algorithm Perdana et al. (2024)	2024	Journal article	International Journal of Advanced Computer Science and Applications
S31	Identifying Gambling Websites with Co-training Wang et al. (2022a)	2022	Conference paper	The 34th International Conference on Software Engineering and Knowledge Engineering
S32	Technical Evaluation of Android Parental Control Tools and PDNS Filtering Services: The Chilean Context Rojas et al. (2025)	2025	Journal article	IEEE Access
S33	Identifying malicious accounts in blockchains using domain names and associated temporal properties Sachan et al. (2023)	2023	Journal article	Blockchain: Research and Applications
S34	Comparative Evaluation of Deep Learning Architectures and Hybrid Heuristics for Automated Gambling Content Detection Sánchez et al. (2026)	2026	Journal article	International Journal of Advanced Computer Science and Applications
S35	Online Gambling Promotion Detection in Indonesian YouTube Comments Using Semi-Supervised IndoBERT Classification Santika et al. (2025)	2025	Conference paper	Procedia Computer Science
S36	Comparative Analysis of AI Models for Detecting Online Gambling Promotion From Trending YouTube Shorts' Comments in Indonesia Santosa et al. (2025)	2025	Conference paper	2025 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD)
S37	GAME-BC: A Graph Attention Model for Exploring Bitcoin Crime Sun et al. (2022)	2022	Conference paper	2022 6th International Symposium on Computer Science and Intelligent Control (ISCSIC)
S38	Automating Hidden Gambling Detection in Web Sites: A BeautifulSoup Implementation Teppap et al. (2024)	2024	Conference paper	2024 21st International Joint Conference on Computer Science and Software Engineering (JCSSE)
S39	The research on a rapid verification method for malicious domain names Wang and Shang (2024)	2024	Conference paper	BDEIM 2023 / ACM

Continued on next page

Table E3: Selected primary studies included in the SLR (continued).

ID	Primary study	Year	Publication type	Source / venue
S40	A Hybrid Multimodal Data Fusion-Based Method for Identifying Gambling Websites Wang et al. (2022b)	2022	Journal article	Electronics
S41	Assessing the effectiveness of crawlers and large language models in detecting adversarial hidden link threats in meta computing Xiong et al. (2025)	2025	Journal article	High-Confidence Computing
S42	DRSDetector: Detecting Gambling Websites by Multi-level Feature Fusion Zhang et al. (2023)	2023	Conference paper	2023 IEEE Symposium on Computers and Communications (ISCC)
S43	Identification and Detection of Illegal Gambling Websites and Analysis of User Behavior Zhang et al. (2025)	2025	Journal article	Computer Science and Information Systems
S44	The Chameleon on the Web: an Empirical Study of the Insidious Proactive Web Defacements Zhao (2023)	2023	Conference paper	Proceedings of the ACM Web Conference 2023 (WWW '23)
S45	No Source Code? No Problem! Demystifying and Detecting Mask Apps in iOS Zhao et al. (2024)	2024	Conference paper	2024 IEEE/ACM 32nd International Conference on Program Comprehension (ICPC)
S46	IDTracker: Discovering Illicit Website Communities via Third-party Service IDs Wang et al. (2023)	2023	Conference paper	2023 53rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)
S47	Detecting and Understanding the Promotion of Illicit Goods and Services on Twitter Wang et al. (2025a)	2025	Conference paper	Proceedings of the ACM Web Conference 2025 (WWW '25)
S48	Gamblers or Delegatees: Identifying Hidden Participant Roles in Crypto Casinos Wang et al. (2025b)	2025	Conference paper	Proceedings of the ACM Web Conference 2025 (WWW '25)
S49	Towards Fraud Detection via Fine-Grained Classification of User Behavior Wang et al. (2025c)	2025	Journal article	IEEE Transactions on Big Data
S50	MEDAL: A Multimodality-Based Effective Data Augmentation Framework for Illegal Website Identification Wen et al. (2024)	2024	Journal article	Electronics

Continued on next page

Table E3: Selected primary studies included in the SLR (continued).

ID	Primary study	Year	Publication type	Source / venue
S51	Exposing the Hidden Layer: Software Repositories in the Service of SEO Manipulation Wu et al. (2025)	2025	Conference paper	2025 IEEE/ACM 47th International Conference on Software Engineering (ICSE)
S52	Recognition of Illegal Websites Based on Similarity of Sensitive Features of Mixed Elements Xiong (2022)	2022	Conference paper	International Conference on Computation, Big-Data and Engineering 2022
S53	Adversarial Hidden Link Threats in Meta Computing Xiong et al. (2024)	2024	Conference paper	2024 International Conference on Meta Computing (ICMC)

Table E4: Published and repository-linked datasets identified from the 53 primary studies.

Study	Dataset name	Source/platform	Dataset size	Availability status	Repository / project link
S01 Gao et al. (2021)	Self-collected illegal mobile gambling app dataset	Illegal gambling websites, app download links, ISP-assisted domain...	3,366 unique gambling apps; 5,344 app versions; over 40 million domains analyzed; over 11K suspicious server	Repository/project link reported; Self-collected	https://mobile-app-research.github.io/
S02 Anderson et al. (2026)	IndoGambling-Comments Dataset	YouTube comments	13,114 Indonesian-language YouTube comments; 250 unique gambling websites mentioned	Repository/project link reported; Self-collected / Not publicly confirmed	https://github.com/valinaevelyn/Detecting-Online-Gambling-Promotion-in-Youtube-Co
S05 Chen et al. (2023)	Large-scale ground-truth UEware dataset	Apps collected from app stores and the wild	25,717 apps collected; 1,700 ground-truth apps; 24,017 apps used in real-world evaluation	Explicit public/released/open-coded; Publicly released by authors	https://github.com/HypoopyH/DeUEDroid
S13 Guo et al. (2024)	Self-collected Telegram underground app dataset	Telegram channels and promotional websites	Over 200 million messages, more than 10,000 downloaded apps, and about 1,000 unique apps were reported	Repository/project link reported; Self-collected	https://github.com/security-pride/TUApps
S15 Han et al. (2023)	Gambling app knowledge graph	Suspected illegal-industry domains, porn/gambling domains, promotio...	277,816 related suspected illegal-industry domains; exact final app/entity counts are partially reported	Explicit public/released/open-coded; Publicly released by authors	https://github.com/yadispace/gambling_app_KG
S17 Huang et al. (2022)	Bitcoin-Gambling-Dataset / Ethereum gambling dataset	Ethereum blockchain smart contracts and transaction records	260 gambling contracts; 1,040 non-gambling contracts; 10,423 gambling addresses; 51,004 non-gambling addresses.	Explicit public/released/open-coded; Publicly released by authors	https://github.com/AwesomeHuang/Bitcoin-Gambling-Dataset

Table E4: Published and repository-linked datasets identified from the 53 primary studies (continued).

Study	Dataset name	Source/platform	Dataset size	Availability status	Repository / project link
S18 Izzatur-rahim et al. (2025)	Not Reported	Multi-source online platforms	Following the pre-processing pipeline, the number of data instances was reduced from 31,240 to 26,786.	Repository/project link reported; Self-collected / Not Reported	https://github.com/mctosima/online_bet_spam_detection https://www.kaggle.com/datasets/ferdiansakti/gambling-comments-from-youtube-platform?resource=download
S19 Kamdan et al. (2025)	Not Reported	YouTube Data API	The dataset consists of 1170 spam comments promoting online gambling and 1145 non-spam comments.	Repository/project link reported; Self-collected / Not Reported	https://huggingface.co/jelialmutaali/online-gambling-spam-detector
S20 Ke et al. (2022)	Dark corpus dataset	Six Chinese Tor websites	Data from six Chinese Tor websites; exact corpus size not reported in extracted evidence.	Repository/project link reported; Self-collected	https://github.com/KL4MVP/Chinese-Jargon-Detection
S24 Liem et al. (2025)	Not Reported	Not Reported	The dataset contains 3241 data points	Repository/project link reported; Not Reported	https://github.com/tyogoutomo/INDONESIAN_ONLINE_GAMBLING_WEBSITE_PROMOTION_DATASET
S25 Lu et al. (2025)	Not Reported	Android apps; illicit app datasets	2,927 potential illicit applications from 2,239 websites.	Repository/project link reported; Not Reported	https://github.com/initfunction/TacDroid
S33 Sachan et al. (2023)	Cisco Umbrella top 1 million January 2020.	Cisco Umbrella; DGA dataset; PhishTank; OpenPhish; Cryptoscam black	335 million DNS queries; 1.77 million distinct queries; 87,557 malicious-DN DNS queries; 41,991 malicious DNS	Explicit public/released/open-coded; Public / mixed external sources	Not reported
S35 Santika et al. (2025)	Self-collected Indonesian YouTube comment dataset	YouTube	10,101 comments; 10% manually labeled	DOI-published repository; Self-collected	https://doi.org/10.5281/zenodo.15424844
S36 Santosa et al. (2025)	Self-collected Indonesian YouTube Shorts comments dataset	YouTube Data API v3	training, validation, and test sets using a 64:16:20 ratio	Repository/project link reported; Self-collected	https://github.com/OhMyBreee/
S45 Zhao et al. (2024)	Real-world Apple App Store dataset	Apple App Store	1500 benign apps and the manually collected 180 Mask Apps	Repository/project link reported; Not Reported	https://github.com/Junzy71/Code_of_Mask_App

Table E4: Published and repository-linked datasets identified from the 53 primary studies (continued).

Study	Dataset name	Source/platform	Dataset size	Availability status	Repository / project link
S46 Wang et al. (2023)	ISP deployment dataset	ISP environment and labelled websites	6,830 illicit communities and 165,378 illicit websites identified in deployment.	Repository/project link reported; Not Reported	https://github.com/IDTrackerSystem/dataset
S47 Wang et al. (2025a)	Twitter PIP dataset	Twitter; also observed visibility on YouTube, Facebook, and TikTok	12,401,082 distinct PIPs; 580,530 Twitter accounts; 37,000 instant messaging accounts reported	Repository/project link reported; Not Reported	https://illicit-promotion.github.io/
S48 Wang et al. (2025b)	Real-world crypto casino cases	Ethereum, TRON, and Arbitrum blockchains	TRON represents the largest dataset, significantly leading with over 1.2 million nodes and 4.3 million edges.	Repository/project link reported; Not Reported	https://github.com/njublockchain/crypto-casino-delegatee-mining
S50 Wen et al. (2024)	Self-developed illegal website dataset	Collected illegal/normal websites	1200 gambling, pornography, and attraction website data points each and 4800 normal websites	Repository/project link reported; Self-developed	https://datacon.qianxin.com/opendata/openpage?resourcesId=8
S51 Wu et al. (2025)	Ten-year software repository dataset	npm, Docker Hub, NuGet	3,801,682 abusive packages discovered	Explicit public/released/open-coded; Reported as open-sourced, but repository URL.	Not reported

Table E5: Study-level performance values extracted from the Main Performance.

Study	Primary metric	Value	Approx	Extracted performance statement
S02 Anderson et al. (2026)	F1-score	98.50%	No	IndoBERT achieved F1-score 0.9850 when font preprocessing was applied.
S04 Chen et al. (2020)	Accuracy	99.18%	No	Accuracy, precision, and F-measure were reported as over 99%; extracted evidence reports gambling website accuracy 99.18% and F...
S05 Chen et al. (2023)	F1-score	98.22%	No	Detection F1-score 98.22%; classification accuracy 98.97%; real-world evaluation found 54% of apps in the wild and 11% of app-s...
S09 Fang et al. (2024)	Macro-F1	91.70%	Yes	Macro-Avg-F1 around 91.70%.
S10 Feng et al. (2025)	F1-score	99.85%	No	Accuracy 0.9985; macro average precision 0.9982, recall 0.9989, F1 0.9985; weighted average F1 0.9985.
S11 Kim et al. (2024)	Accuracy	92.00%	Yes	Random Forest, KNN, and ensemble models recorded accuracy over 92%; LR and SVM were approximately 80% accurate.
S12 Ghani et al. (2025)	Accuracy	93.00%	No	Random Forest achieved 93% accuracy; reported precision was 88% for DDoS, 97% for Malware or Backdoor, and 100% for Normal traf...
S14 Guo et al. (2026)	Accuracy	97.57%	No	Machine-learning identification achieved 97.57% accuracy; 3.3% of NRDs were flagged as illicit websites; 60% of abusive domains...
S16 Harahap and Ridho (2024)	F1-score	97.00%	No	TPR 0.95, TNR 0.99, PPV 0.99, NPV 0.96, accuracy 0.97, and F1-score 0.97.
S17 Huang et al. (2022)	F1-score	89.00%	No	ETHGamDet achieved F1-score 0.72 for address classification and 0.89 for contract classification.

Table E5: Study-level performance values extracted from the Main Performance Results column (continued).

Study	Primary metric	Value	Approx	Extracted performance statement
S18 Izzaturrahim et al. (2025)	F1-score	98.97%	No	CNN combined with IndoBERT achieved F1 testing score of 0.9897; performance improved from 0.5267 in the reported comparison.
S19 Kamdan et al. (2025)	F1-score	98.26%	No	IndoBERT achieved 98.26% accuracy; extracted evidence states precision 98.28%, recall 98.26%, and F1 98.26%.
S20 Ke et al. (2022)	Accuracy	91.50%	No	The proposed framework achieved detection accuracy of 91.5%.
S21 Prakoso et al. (2025)	F1-score	99.04%	No	All models achieved F1 above 0.97; IndoBERT Base performed best with F1 = 0.9904.
S24 Liem et al. (2025)	F1-score	97.24%	No	IndoBERT achieved 98.53% accuracy, 97.24% F1-score, 95.65% recall, 98.88% precision, and 98.21% F-beta score.
S25 Lu et al. (2025)	F1-score	96.73%	No	TACDROID achieved F1-score 96.73% for illicit app detection and 92.09% for classification; UTG construction F1-score 83.48%.
S27 Min et al. (2022)	Accuracy	97.00%	No	Detection accuracy rate of 97%.
S28 Muzakir et al. (2025b)	F1-score	93.04%	No	RNN achieved accuracy 93.07%, precision 93.22%, recall 92.87%, and F1-score 93.04%; CNN achieved accuracy 92.80%; AutoML achiev...
S29 Muzakir et al. (2025a)	Accuracy	92.80%	No	CNN achieved 92.80% accuracy, outperforming AutoML at 90.19%, with higher precision and recall.
S30 Perdana et al. (2024)	Precision	96.60%	No	TF-IDF and Random Forest achieved recall 0.958 and precision 0.966.
S31 Wang et al. (2022a)	F1-score	98.33%	No	Co-TextRNN achieved overall Precision 0.9835, Recall 0.9833, and F1 0.9833 after co-training; Co-CNN-5 achieved F1 0.9000.

Table E5: Study-level performance values extracted from the Main Performance Results column (continued).

Study	Primary metric	Value	Approx	Extracted performance statement
S33 Sachan et al. (2023)	Balanced accuracy	89.53%	No	Supervised approach achieved 89.53% balanced accuracy; 73,769 blockchain DNs showed malicious behavior at least once; 34,171 sh...
S34 Sánchez et al. (2026)	Accuracy	85.01%	No	ResNet50 achieved 85.01% accuracy and 90.42% recall.
S35 Santika et al. (2025)	F1-score	98.50%	No	Initial IndoBERT achieved accuracy 0.953, precision 0.945, recall 0.977, and F1 0.961; semi-supervised IndoBERT achieved accura...
S36 Santosa et al. (2025)	Weighted F1-score	95.62%	No	IndoBERT achieved accuracy 95.63% and weighted F1-score 95.62% ± 0.0084.
S38 Teppap et al. (2024)	F1-score	91.00%	No	Proposed model achieved accuracy 0.89, precision 0.83, recall 1.00, and F1-score 0.91; baseline accuracy 0.53 and F1-score 0.69.
S40 Wang et al. (2022b)	F1-score	99.00%	Yes	Final method reported accuracy, precision, recall, and F1-score all over 99%; ResNet34 image classifier alone achieved accuracy...
S42 Zhang et al. (2023)	Accuracy	97.83%	No	DRSDetector achieved 97.83% accuracy.
S45 Zhao et al. (2024)	Recall	99.12%	No	Mask-Catcher achieved high recall and precision; prior QA evidence reported recall 99.12% and precision 100%.
S46 Wang et al. (2023)	F1-score	89.68%	No	F1-score 0.8968; deployment identified 6,830 illicit communities containing 165,378 illicit websites.

References

Anderson H, Pranoto V, Winata V, et al (2026) Detecting online gambling promotion in youtube comments using transformer-based models. In: 2026 International Seminar on Intelligent Business and Edge-Computing Research (ISIBER), pp 692–697, https://doi.org/10.1109/ISIBER68248.2026.11469786

Andersson S, Carlbring P, Lyon K, et al (2025) Insights into the temporal dynamics of identifying problem gambling on an online casino: A machine learning study on routinely collected individual account data. Journal of Behavioral Addictions 14(1):490–500. https://doi.org/10.1556/2006.2025.00013

Auer M, Griffiths M (2024) Predicting high-risk gambling based on the first seven days of gambling activity after registration using account-based tracking data. International Journal of Mental Health and Addiction 22(6):3397–3413. https://doi.org/10.1007/s11469-023-01056-4

Awadh W, Sulaiman R, Mahmoud M (2025) Aspect-based sentiment analysis in moocs: a systematic literature review introducing the masc-mef framework. Journal of King Saud University - Computer and Information Sciences 37(2). https://doi.org/10.1007/s44443-025-00018-1

Boland A, Cherry M, Dickson R (2017) Doing a Systematic Review: A Student's Guide, 2nd edn. SAGE Publications, London, UK

Buttigieg K, Caruana M, Suda D (2022) Identifying problematic gamblers using multiclass and two-stage binary neural network approaches. In: International Conference on Agents and Artificial Intelligence, pp 336–342, https://doi.org/10.5220/0010821100003116

Caraka R, Nurhaliza S, Arato A, et al (2025) Exploring the digital pulse of sentiment analysis on online gambling in indonesia using χ^2 feature selection. Procedia Computer Science 269:1406–1413. https://doi.org/10.1016/j.procs.2025.09.082

Carrera-Rivera A, Ochoa W, Larrinaga F, et al (2022) How-to conduct a systematic literature review: A quick guide for computer science research. MethodsX 9:101895. https://doi.org/10.1016/j.mex.2022.101895

Chagas B, Gomes J, Griffiths M (2022) Consumer profile segmentation in online lottery gambling utilizing behavioral tracking data from the portuguese national lottery. Journal of Gambling Studies 38(3):917–939. https://doi.org/10.1007/s10899-021-10072-9

Challet-Bouju G, Hardouin J, Thiabaud E, et al (2020) Modeling early gambling behavior using indicators from online lottery gambling tracking data: Longitudinal analysis. Journal of Medical Internet Research https://doi.org/10.2196/17675

Charitou C, Garcez A, Dragicevic S (2020) Semi-supervised gans for fraud detection. In: 2020 International Joint Conference on Neural Networks (IJCNN), https://doi.org/10.1109/IJCNN48605.2020.9206844

Chen J, Zheng S, Cheng Y, et al (2024) Data mining based analysis of online gambling sites and illicit financial flows. In: Proceedings of the 2024 International Conference on Cloud Computing and Big Data, pp 205–211, https://doi.org/10.1145/3695080.

3695116

Chen Y, Zheng R, Zhou A, et al (2020) Automatic detection of pornographic and gambling websites based on visual and textual content using a decision mechanism. Sensors 20(14):3989. https://doi.org/10.3390/s20143989

Chen Z, Liu J, Hu Y, et al (2023) DeUEDroid: Detecting underground economy apps based on UTG similarity. In: Proceedings of the 32nd ACM SIGSOFT International Symposium on Software Testing and Analysis, https://doi.org/10.1145/3597926.3598051

Cholistyo R, Lhaksmana K (2025) Comparative evaluation of indobert-base and indobert-lite for detecting indonesian-language spam comments on instagram. In: 2025 2nd Beyond Technology Summit on Informatics International Conference (BTS-I2C), https://doi.org/10.1109/BTS-I2C67944.2025.11399297

Cohen J (1960) A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20(1):37–46. https://doi.org/10.1177/001316446002000104

Cokronegoro H, Prasetyo H (2025) Optimizing law enforcement in the case of online gambling promoters/endorsements. Media Iuris 8(1):73–94. https://doi.org/10.20473/mi.v8i1.59348

Dam T, Klausner L, Buhov D, et al (2019) Large-scale analysis of pop-up scam on typosquatting urls. In: ACM International Conference Proceeding Series, https://doi.org/10.1145/3339252.3340332

Dam T, Klausner L, Schrittwieser S (2020) Typosquatting for fun and profit: Cross-country analysis of pop-up scam. Journal of Cyber Security and Mobility 9(2):265–300. https://doi.org/10.13052/JCSM2245-1439.924

Dong B, Wang W, Sun L, et al (2025) Anomaly detection using graph anomaly rules. Big Data Mining and Analytics 8(5):1075–1091. https://doi.org/10.26599/BDMA.2025.9020012

Erawan E, Satria R, Nastiti A, et al (2024) A lightweight design approach to detect malicious injections in government websites: Combating hidden online gambling pages. In: 2024 Beyond Technology Summit on Informatics International Conference (BTS-I2C), pp 365–369, https://doi.org/10.1109/BTS-I2C63534.2024.10942063

Fang X, Ma X, Wang C, et al (2024) Detecting and exploring malicious websites through multi-message passing heterogeneous neural network. In: 2024 IEEE Symposium on Computers and Communications (ISCC), https://doi.org/10.1109/ISCC61673.2024.10733612

Feng Z, Han D, Wu S, et al (2025) Anomalous traffic identification method for POST messages based on gambling website templates. Computer Science and Information

Systems 22(1):79–103. https://doi.org/10.2298/CSIS240728069F

Fiedler I (2018) Regulation of online gambling. Economics and Business Letters 7(4):162–168. https://doi.org/10.17811/ebl.7.4.2018.162-168

Gao Y, Wang H, Li L, et al (2021) Demystifying illegal mobile gambling apps. In: Proceedings of the Web Conference 2021, pp 1447–1458, https://doi.org/10.1145/3442381.3449932

Ghani E, Sofwan A, Somantri M (2025) Ai-driven network security: Detecting and mitigating ddos, malware, and backdoor attacks with isolation and random forest algorithm. In: 2025 International Conference on Smart Computing, IoT and Machine Learning (SIML), https://doi.org/10.1109/SIML65326.2025.11080951

Gibson E, Griffiths M, Calado F, et al (2022) The relationship between videogame micro-transactions and problem gaming and gambling: A systematic review. Computers in Human Behavior 131:107219. https://doi.org/10.1016/j.chb.2022.107219

Guo B, Zhang Y, Shi F, et al (2026) Darkness at dawn: understanding illicit websites in newly registered domain names. Cybersecurity 9(1). https://doi.org/10.1186/s42400-025-00523-w

Guo Y, Wang D, Wang L, et al (2024) Beyond app markets: Demystifying underground mobile app distribution via telegram. Proceedings of the ACM on Measurement and Analysis of Computing Systems 8(3). https://doi.org/10.1145/3700432

Haefeli J, Lischer S, Schwarz J (2011) Early detection items and responsible gambling features for online gambling. International Gambling Studies 11(3):273–288. https://doi.org/10.1080/14459795.2011.604643

Han Y, Wang S, Li Y, et al (2023) Measurement of illegal android gambling app ecosystem from joint promotion perspective. In: 2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA), https://doi.org/10.1109/DSAA60987.2023.10302499

Harahap H, Ridho F (2024) Detection of online gambling web defacement in university domains using attack signatures. In: 2024 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD), pp 73–78, https://doi.org/10.1109/ICoABCD63526.2024.10704413

Huang Z, Liu Z, Chen J, et al (2022) Who is gambling? finding cryptocurrency gamblers using multi-modal retrieval methods. International Journal of Multimedia Information Retrieval 11(4):539–551. https://doi.org/10.1007/s13735-022-00264-3

Hynek K, Vekshin D, Luxemburk J, et al (2022) Summary of dns over https abuse. IEEE Access https://doi.org/10.1109/ACCESS.2022.3175497

Izzaturrahim M, Husnul Khotimah P, Darmayanti N (2025) Transformer based hybridization text classifiers in online gambling spam identification case study. In: 2025 International Conference on Computer, Control, Informatics and its Applications (IC3INA), pp 126–131, https://doi.org/10.1109/IC3INA68387.2025.11325712

Jeong Y, Lee W, Hur J (2023) A honey postmessage, but a heart of gall: Exploiting push service in service workers via postmessage. In: Proceedings of the 2023 ACM Asia Conference on Computer and Communications Security, https://doi.org/10.1145/3579856.3590342

Jilcha L, Kwak J (2022) Machine learning-based advertisement banner identification technique for effective piracy website detection process. Computers, Materials and Continua 71(2):2883–2899. https://doi.org/10.32604/cmc.2022.023167

Kamdan K, Anugrah M, Almutaali M, et al (2025) Performance analysis of indobert for detection of online gambling promotion in youtube comments. Engineering Proceedings 107(1). https://doi.org/10.3390/engproc2025107066

Ke L, Chen X, Wang H (2022) An unsupervised detection framework for chinese jargons in the darknet. In: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, pp 458–466, https://doi.org/10.1145/3488560.3498469

Kim C, Park J, Lee J (2024) Ai-based betting anomaly detection system to ensure fairness in sports and prevent illegal gambling. Scientific Reports 14(1). https://doi.org/10.1038/s41598-024-57195-8

Kitchenham B, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering. Tech. rep., Keele University and University of Durham, EBSE 2007-001

Kurniawan C, Triayudi A (2024) Reconstruction and detection of gambling web defacement attack using wazuh and velociraptor. In: 2024 International Conference on Information Technology Research and Innovation (ICITRI), pp 257–262, https://doi.org/10.1109/ICITRI62858.2024.10699215

Landis J, Koch G (1977) The measurement of observer agreement for categorical data. Biometrics 33(1):159–174

Li T, Kou G, Peng Y (2020) Improving malicious urls detection via feature engineering: Linear and nonlinear space transformation methods. Information Systems https://doi.org/10.1016/j.is.2020.101494

Li Z, Zhang S, Yin J, et al (2022) Fighting against piracy:an approach to detect pirated video websites enhanced by third-party services. In: 2022 IEEE Symposium on Computers and Communications (ISCC), https://doi.org/10.1109/ISCC55528.

2022.9912777

Liang J, Li L, Chen W, et al (2019) Targeted addresses identification for bitcoin with network representation learning. In: 2019 IEEE International Conference on Intelligence and Security Informatics (ISI), https://doi.org/10.1109/ISI.2019.8823249

Liberato M, Affinito A, Meijerink B, et al (2025) To block or not to block? evaluating parental controls across routers, dns services, and software. In: TMA 2025 - Proceedings of the 9th Network Traffic Measurement and Analysis Conference, https://doi.org/10.23919/TMA66427.2025.11096991

Liem J, Utomo D, Ignasio B, et al (2025) Deep learning and machine learning for indonesian online gambling website promotion detection. Procedia Computer Science 269:979–992. https://doi.org/10.1016/j.procs.2025.09.040

Liu T, Wang H, Li L, et al (2020) Maddroid: Characterizing and detecting devious ad contents for android apps. In: Proceedings of The Web Conference 2020, https://doi.org/10.1145/3366423.3380242

Liu Y, Liu M, Zhang Y, et al (2025) Dive into the cloud: Unveiling the (ab)usage of serverless cloud function in the wild. In: Proceedings of the 2025 ACM Internet Measurement Conference, https://doi.org/10.1145/3730567.3732915

Lu Y, Lin H, He Z, et al (2025) TacDroid: Detection of illicit apps through hybrid analysis of ui-based transition graphs. In: Proceedings of the IEEE/ACM 47th International Conference on Software Engineering, pp 2790–2802, https://doi.org/10.1109/ICSE55347.2025.00111

Luo X, Wang L, Xu Z, et al (2018) A large scale analysis of dns water torture attack. In: Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence, https://doi.org/10.1145/3297156.3297272

Mawardi R, Nuraeni V, Jasman J, et al (2026) Digital financial compliance challenges: Applying routine activity theory to online gambling networks analysis. Emerging Science Journal 10(1):315–332. https://doi.org/10.28991/ESJ-2026-010-01-015

Min M, Lee J, Park H, et al (2021) Honeypot system for automatic reporting of illegal online gambling sites utilizing sms spam. In: 2021 World Automation Congress (WAC), pp 180–185, https://doi.org/10.23919/WAC50355.2021.9559478

Min M, Lee J, Lee K (2022) Detecting illegal online gambling (IOG) services in the mobile environment. Security and Communication Networks 2022:3286623. https://doi.org/10.1155/2022/3286623

Murch W, Kairouz S, French M (2024) Establishing the temporal stability of machine learning models that detect online gambling-related harms. Computers in Human

Behavior Reports 14. https://doi.org/10.1016/j.chbr.2024.100427

Muzakir A, Ependi U, Suyanto P, et al (2025a) Text-based detection of online gambling advertisements on social media using automl and cnn. In: 2025 2nd International Conference on Information System and Information Technology (ICISIT), https://doi.org/10.1109/ICISIT66233.2025.11402999

Muzakir A, Ependi U, Suyanto S (2025b) A deep learning and automl-based multimodal text extraction framework for detecting online gambling advertisements in indonesian social media. International Journal of Safety and Security Engineering 15(9):1837–1848. https://doi.org/10.18280/ijsse.150908

Perdana R, Ardin B, I. S, et al (2024) Detecting online gambling promotions on indonesian twitter using text mining algorithm. International Journal of Advanced Computer Science and Applications 15(8):942–949. https://doi.org/10.14569/IJACSA.2024.0150893

Pervez M, Atiq Ullah M, Ahmed Khan I, et al (2026) (mis-)informed consent: Predatory apps and the exploitation of populations with limited literacy. In: Proceedings of the ACM Web Conference 2026, https://doi.org/10.1145/3774904.3792969

Prakoso G, Doewes A, Anggrainingsih R (2025) Detecting online gambling promotional comments on youtube using indonesian bert and multilingual bert. In: 2025 International Conference on Artificial Intelligence and Technological Solutions (ICAITech), pp 37–44, https://doi.org/10.1109/ICAITech66481.2025.11387329

Rojas N, Boettcher N, Játiva P, et al (2025) Technical evaluation of android parental control tools and pdns filtering services: The chilean context. IEEE Access 13:186523–186536. https://doi.org/10.1109/ACCESS.2025.3625851

Sachan R, Agarwal R, Shukla S (2023) Identifying malicious accounts in blockchains using domain names and associated temporal properties. Blockchain: Research and Applications 4(3):100136. https://doi.org/10.1016/j.bcra.2023.100136

Sánchez E, Tejada C, Quispe J (2026) Comparative evaluation of deep learning architectures and hybrid heuristics for automated gambling content detection. International Journal of Advanced Computer Science and Applications 17(1). https://doi.org/10.14569/IJACSA.2026.0170193

Santika S, Saputra M, Zahra A, et al (2025) Online gambling promotion detection in indonesian youtube comments using semi-supervised IndoBERT classification. Procedia Computer Science 269:1269–1278. https://doi.org/10.1016/j.procs.2025.09.068

Santosa B, Saliano A, Safira R, et al (2025) Comparative analysis of ai models for detecting online gambling promotion from trending youtube shorts' comments in indonesia. In: 2025 International Conference on Artificial Intelligence,

Blockchain, Cloud Computing, and Data Analytics (ICoABCD), https://doi.org/10.1109/ICoABCD67551.2025.11470891

Scholten O, Hughes N, Deterding S, et al (2019) Ethereum crypto-games: Mechanics, prevalence, and gambling similarities. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play, https://doi.org/10.1145/3311350.3347178

Schulmann H, Zhao S (2024) Zpredict: Ml-based ipid side-channel measurements. ACM Trans Priv Secur https://doi.org/10.1145/3672560

Smith E, Peters J, Reiter N (2024) Automatic detection of problem-gambling signs from online texts using large language models. PLOS Digital Health 3(9). https://doi.org/10.1371/journal.pdig.0000605

Stechschulte G, Wintner M, Hemmje M, et al (2024) In-database feature extraction to improve early detection of problematic online gambling behavior. IEEE Transactions on Computational Social Systems 11(5):6868–6881. https://doi.org/10.1109/TCSS.2024.3406501

Sun K, Meng K, Zheng Z (2022) GAME-BC: A graph attention model for exploring bitcoin crime. In: 2022 6th International Symposium on Computer Science and Intelligent Control (ISCSIC), pp 342–346, https://doi.org/10.1109/ISCSIC57216.2022.00077

Suzuki H, Nakamura R, Inagaki A, et al (2019) Early detection of problem gambling based on behavioral changes using shapelets. In: Proceedings of the 2019 IEEE/WIC/ACM International Conference on Web Intelligence, pp 367–372, https://doi.org/10.1145/3350546.3352549

Teichert T, Gainsbury S, Mühlbach C (2017) Positioning of online gambling and gaming products from a consumer perspective: A blurring of perceived boundaries. Computers in Human Behavior 75:757–765. https://doi.org/10.1016/j.chb.2017.06.025

Teppap P, Tipauksorn P, Surathong S, et al (2024) Automating hidden gambling detection in web sites: A beautifulsoup implementation. In: 2024 21st International Joint Conference on Computer Science and Software Engineering (JCSSE), pp 132–139, https://doi.org/10.1109/JCSSE61278.2024.10613687

Tran L, Wardle H, Colledge-Frisby S, et al (2024) The prevalence of gambling and problematic gambling: A systematic review and meta-analysis. The Lancet Public Health 9(8):e594–e613. https://doi.org/10.1016/S2468-2667(24)00126-9

Wang C, Xue P, Zhang M, et al (2022a) Identifying gambling websites with co-training. In: Proceedings of the International Conference on Software Engineering and Knowledge Engineering, SEKE, pp 598–603, https://doi.org/10.18293/SEKE2022-106

Wang C, Zhang M, Shi F, et al (2022b) A hybrid multimodal data fusion-based method for identifying gambling websites. Electronics 11(16):2489. https://doi.org/10.3390/electronics11162489

Wang C, Li Z, Yin J, et al (2023) IDTracker: Discovering illicit website communities via third-party service ids. In: 2023 53rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), pp 459–469, https://doi.org/10.1109/DSN58367.2023.00050

Wang H, Li Y, Huang R, et al (2025a) Detecting and understanding the promotion of illicit goods and services on twitter. In: Proceedings of the ACM on Web Conference 2025, pp 3389–3404, https://doi.org/10.1145/3696410.3714550

Wang J, Mao Q, Sun H, et al (2025b) Gamblers or delegatees: Identifying hidden participant roles in crypto casinos. In: Proceedings of the ACM on Web Conference 2025, pp 1004–1015, https://doi.org/10.1145/3696410.3714689

Wang L, Shang Q (2024) The research on a rapid verification method for malicious domain names. In: Proceedings of the 2023 4th International Conference on Big Data Economy and Information Management, pp 166–174, https://doi.org/10.1145/3659211.3659240

Wang X, Yu H, Guo J, et al (2025c) Towards fraud detection via fine-grained classification of user behavior. IEEE Transactions on Big Data 11(4):1994–2007. https://doi.org/10.1109/TBDATA.2024.3517313

Wen L, Zhang M, Wang C, et al (2024) MEDAL: A multimodality-based effective data augmentation framework for illegal website identification. Electronics 13(11):2199. https://doi.org/10.3390/electronics13112199

Wen S, Zhao Z, Yan H (2018) Detecting malicious websites in depth through analyzing topics and web-pages. In: Proceedings of the 2nd International Conference on Cryptography, Security and Privacy, https://doi.org/10.1145/3199478.3199500

Wu M, Hong G, Mai W, et al (2025) Exposing the hidden layer: Software repositories in the service of seo manipulation. In: 2025 IEEE/ACM 47th International Conference on Software Engineering (ICSE), pp 2100–2112, https://doi.org/10.1109/ICSE55347.2025.00147

Wu S, Wu Z, Chen S, et al (2021) Community detection in blockchain social networks. Journal of Communications and Information Networks https://doi.org/10.23919/JCIN.2021.9387705

Xiong J (2022) Recognition of illegal websites based on similarity of sensitive features of mixed elements. In: 2022 International Conference on Computation, Big-Data and Engineering (ICCBE), https://doi.org/10.1109/ICCBE56101.2022.9888158

Xiong J, Wei M, Lu Z, et al (2024) Adversarial hidden link threats in meta computing. In: 2024 International Conference on Meta Computing (ICMC), pp 23–34, https://doi.org/10.1109/ICMC60390.2024.00010

Xiong J, Wei M, Lu Z, et al (2025) Assessing the effectiveness of crawlers and large language models in detecting adversarial hidden link threats in meta computing. High-Confidence Computing https://doi.org/10.1016/j.hcc.2024.100292

Zhang S, Yin J, Li Z, et al (2022) Node-imbalance learning on heterogeneous graph for pirated video website detection. In: 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD), https://doi.org/10.1109/CSCWD54268.2022.9776224

Zhang Y, Fu X, Yang R, et al (2023) DRSDetector: Detecting gambling websites by multi-level feature fusion. In: 2023 IEEE Symposium on Computers and Communications (ISCC), pp 1441–1447, https://doi.org/10.1109/ISCC58397.2023.10217923

Zhang Z, Han D, Wu S, et al (2025) Identification and detection of illegal gambling websites and analysis of user behavior. Computer Science and Information Systems 22(3):859–879. https://doi.org/10.2298/CSIS240930019Z

Zhao R (2023) The chameleon on the web: An empirical study of the insidious proactive web defacements. In: Proceedings of the ACM Web Conference 2023, pp 2241–2251, https://doi.org/10.1145/3543507.3583377

Zhao Y, Yu L, Sun Y, et al (2024) No source code? no problem! demystifying and detecting mask apps in ios. In: Proceedings of the 32nd IEEE/ACM International Conference on Program Comprehension, pp 358–369, https://doi.org/10.1145/3643916.3644419