INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY TRENDS

Vol. 6 No. 02 (2026)

e.ISSN: 2409-5222

https://ijictt.ictt.info

Open-Set Discovery of Policy-Evasion Patterns in Short-Video Comment Networks

Benjamin M. Meyer

Department of Computer Science, Faculty of Engineering, University of Freiburg, Freiburg im Breisgau, Germany

Abstract

The exponential proliferation of short-video platforms has radically transformed digital communication, simultaneously engendering complex challenges in content moderation and platform governance. Malicious actors continuously devise sophisticated strategies to circumvent algorithmic moderation, utilizing linguistic obfuscation, adversarial spelling, and contextual manipulation to disseminate illicit or harmful content. Traditional moderation systems heavily rely on closed-set classification paradigms, which are fundamentally ill-equipped to identify novel, previously unseen evasion tactics. To address this critical limitation, this paper proposes a comprehensive framework for the open-set discovery of policy-evasion patterns specifically tailored to the unique dynamics of short-video comment networks. By integrating advanced natural language processing techniques with structural network analysis, the proposed methodology captures both the semantic nuances of individual comments and the behavioral topologies of user interactions. We formulate the problem as an open-set recognition task, enabling the system to accurately classify known policy violations while explicitly rejecting and isolating anomalous patterns that signify emergent evasion strategies. Extensive empirical evaluations conducted on large-scale, real-world datasets demonstrate the superiority of the proposed framework over conventional closed-set models, particularly in recognizing evolving linguistic anomalies and coordinated malicious campaigns. The findings provide profound implications for the design of adaptive, resilient moderation systems, ensuring safer digital environments in the face of adversarial innovation.

Keywords: *Short-Video Platforms, Content Moderation, Open-Set Recognition, Social Network Analysis.*

1. Introduction

1.1 The Ubiquity of Short-Video Platforms and Moderation Challenges

The contemporary digital landscape is profoundly characterized by the meteoric rise of short-video platforms, which have fundamentally altered the paradigms of multimedia consumption and interpersonal communication on a global scale. Platforms in this domain facilitate the rapid creation, dissemination, and consumption of highly engaging audiovisual content, supported by sophisticated recommendation algorithms that maximize user retention and interaction [1]. A critical component of this ecosystem is the

comment network, where millions of users engage in real-time discourse, community building, and information sharing. However, the sheer volume and velocity of user-generated content present formidable challenges for platform administrators tasked with maintaining a safe and compliant digital environment [2]. Content moderation is imperative to mitigate the proliferation of hate speech, cyberbullying, illicit commerce, and targeted disinformation campaigns. Platform governance frameworks dictate stringent community guidelines; yet, enforcing these policies across millions of heterogeneous, context-dependent textual exchanges requires automated computational systems of immense scale and precision [3]. Standard moderation pipelines predominantly employ supervised machine learning algorithms trained on expansive, manually annotated datasets. While these systems achieve high efficacy in identifying overt policy violations, they inherently operate under a static assumption of the threat landscape, rendering them highly susceptible to adversarial adaptation by sophisticated malicious actors [4]. Consequently, there is an urgent need to transition from reactive moderation paradigms to proactive discovery systems capable of anticipating and neutralizing emergent threats.

1.2 The Evolution of Policy-Evasion Tactics

As automated moderation technologies have advanced, so too have the strategies employed by malicious entities seeking to bypass them. Policy evasion in short-video comment networks is an intricate, highly dynamic phenomenon characterized by adversarial linguistics and behavioral manipulation [5]. Users intent on disseminating prohibited content continuously invent novel vocabularies, employing phonetic approximations, homophones, obscure acronyms, and complex emoji combinations to mask the true semantic intent of their messages. Because short-video comments are inherently brief, informal, and heavily reliant on the audiovisual context of the accompanying video, detecting these subtle linguistic anomalies is exceptionally difficult. Furthermore, organized campaigns often leverage coordinated behavioral patterns across multiple accounts to amplify prohibited content while remaining below the threshold of traditional anomaly detection systems. These evasion patterns are not static; they exhibit rapid evolutionary cycles, often mutating within hours or days in response to algorithmic updates [6]. When platform administrators implement filters for a specific obfuscated term, malicious actors promptly pivot to a new linguistic variant, creating a perpetual cat-and-mouse dynamic. This continuous adaptation exposes a fundamental vulnerability in contemporary content moderation infrastructures: the reliance on closed-set classification models.

1.3 Limitations of Closed-Set Classification

The prevailing methodology in automated content moderation relies on closed-set classification, a paradigm in which machine learning models are trained and evaluated on a predetermined, finite set of categorical labels. In this framework, the model assumes that all data encountered during deployment will invariably belong to one of the classes present in the training corpus [7]. When confronted with a novel evasion pattern that represents an unknown class, a closed-set model is mathematically constrained to map the novel input onto one of the existing known classes, often with an unwarranted high degree of statistical confidence. This phenomenon, known as silent failure, is deeply problematic in the context of platform governance, as it allows emergent policy violations to persist undetected while simultaneously polluting the statistical reliability of the moderation system [8]. The dynamic nature of short-video comment networks necessitates a paradigm shift towards open-set recognition, wherein the computational model is explicitly designed to recognize its own epistemic uncertainty. An open-set system is capable of accurately classifying known threats while simultaneously identifying and sequestering instances that deviate significantly from the known training distribution, thereby facilitating the discovery of entirely novel policy-evasion patterns.

1.4 Research Objectives and Contributions

This research aims to conceptualize, develop, and evaluate a comprehensive framework for the open-set discovery of policy-evasion patterns within the complex environment of short-video comment networks. By synthesizing textual semantics with the structural topology of user interactions, the proposed methodology seeks to identify both the linguistic markers of evasion and the orchestrated network behaviors of malicious actors. The primary contributions of this paper are manifold. First, we articulate the theoretical

foundation for applying open-set recognition principles to the domain of social media content moderation, highlighting the critical necessity of managing unknown unknowns in adversarial environments. Second, we introduce a multimodal feature engineering pipeline that captures the distinctive characteristics of short-video comments, integrating advanced linguistic representations with macro-level network features. Third, we propose an algorithmic architecture that effectively delineates the boundaries of known policy violations, utilizing density-based estimation and margin-enforcing criteria to isolate anomalous patterns. Finally, through extensive empirical analysis on large-scale datasets, we demonstrate the practical efficacy of our approach, providing actionable insights for platform administrators and contributing to the broader academic discourse on adversarial machine learning and digital governance.

2. Background and Related Work

2.1 Theoretical Foundations of Content Moderation

The academic investigation into digital content moderation has evolved significantly over the past decade, transitioning from manual oversight mechanisms to highly automated, algorithmic infrastructures. Early scholarly work in this domain primarily focused on the ethical implications of censorship and the socio-legal responsibilities of platform operators [9]. However, as the scale of user-generated content expanded exponentially, the focus of research shifted toward the computational implementation of automated filters. Supervised learning models, particularly those leveraging support vector machines and early neural network architectures, became the industry standard for identifying profanity, explicit content, and overt hate speech [10]. These systems relied heavily on predefined lexicons and explicit syntactic rules. Despite their initial success, researchers quickly identified severe limitations in rule-based and shallow learning approaches, noting their high susceptibility to circumvention through minor typographical alterations. The advent of deep learning and large-scale language representation models catalyzed a revolution in semantic analysis, enabling systems to comprehend context and subtle linguistic nuances with unprecedented accuracy [11]. Nevertheless, the fundamental architecture of these advanced models remained rooted in the closed-set paradigm, perpetually requiring massive datasets of newly annotated examples to maintain their efficacy against an ever-changing threat landscape.

2.2 Adversarial Text and Linguistic Obfuscation

In the context of policy evasion, adversarial text refers to malicious modifications applied to textual content with the explicit intent of deceiving automated classification algorithms while preserving the semantic meaning for human readers. This field of study draws heavily from adversarial machine learning, though it is uniquely constrained by the discrete nature of human language. Researchers have documented a vast taxonomy of linguistic obfuscation techniques employed by users on social media platforms [12]. These include morphological manipulation, where characters are replaced with visually similar glyphs from different linguistic scripts, and phonetic spelling, where words are spelled precisely as they sound in specific regional dialects to bypass standard dictionary filters. In short-video comment networks, these tactics are further complicated by the pervasive use of platform-specific slang, acronyms, and multimodal communicative tokens such as emojis. Emojis, in particular, frequently serve as clandestine signifiers for illicit substances, self-harm, or commercial solicitation [13]. Academic literature emphasizes that recognizing these obfuscated patterns requires models that can generalize beyond superficial textual features to capture deep semantic and contextual associations. However, because new obfuscation strategies are invented continuously, relying solely on historical data to train generalizable models is fundamentally insufficient, underscoring the critical need for discovery-oriented methodologies.

2.3 Principles of Open-Set Recognition

Open-set recognition represents a highly specialized subfield of machine learning designed to address the inherent limitations of traditional closed-set models operating in dynamic, unconstrained environments. The theoretical framework of open-set recognition was formally established to tackle scenarios where the inference data contains classes that were completely unseen during the training phase [14]. The primary objective is to simultaneously minimize the empirical risk over the known classes and the open-space risk, which is defined as the risk associated with assigning high confidence scores to regions of the feature space

that are far from the training data. Various computational strategies have been proposed to achieve this balance. Early approaches utilized support vector machines augmented with extreme value theory to calibrate probability estimates at the decision boundaries [15]. More recently, deep learning architectures have incorporated generative adversarial networks, contrastive learning, and specialized loss functions to explicitly model the distribution of known classes and enforce a margin of separation against potential unknown samples. In the domain of natural language processing, open-set recognition has been applied to intent classification in dialogue systems and anomalous topic detection [16]. However, its application to the highly adversarial and structurally complex environment of short-video comment networks remains underexplored, presenting a significant gap in the current literature.

2.4 Network-Based Anomaly Detection

While analyzing the linguistic properties of individual comments is essential, contemporary policy evasion is frequently a coordinated, collective endeavor. Consequently, researchers have increasingly turned to network science and graph theory to identify macroscopic patterns of malicious behavior. By representing user interactions, co-commenting behaviors, and content sharing as complex networks, it is possible to extract structural features that are highly indicative of orchestrated evasion campaigns [17]. Malicious actors often operate within tightly knit clusters, utilizing multiple synchronized accounts to artificially inflate the visibility of prohibited content or to distribute obfuscated messages across a wide array of target videos. Network-based anomaly detection techniques leverage metrics such as degree centrality, clustering coefficients, and algorithmic community detection to isolate these suspicious substructures [18]. Previous studies have successfully applied graph neural networks to detect botnets and astroturfing campaigns on traditional microblogging platforms [19]. Integrating this structural network modality with deep semantic analysis provides a holistic representation of user behavior, significantly enhancing the capability of automated systems to differentiate between benign linguistic innovation and malicious policy evasion.

3. Methodology for Open-Set Discovery

3.1 Integrated System Architecture

The architecture of the proposed open-set discovery framework is meticulously designed to process the immense volume and high velocity of data characteristic of short-video comment networks. The system operates through a sequence of interdependent modules, beginning with multimodal data ingestion and culminating in the isolation and clustering of unknown evasion patterns. The foundational premise of this architecture is the integration of textual semantics and network topology into a unified latent representation space. By projecting diverse features into a common dimensionality, the system can evaluate the multifaceted nature of each comment and its associated user behavior. This unified representation is subsequently processed by the open-set classification module, which evaluates the spatial position of the instance relative to the established boundaries of known policy violations.

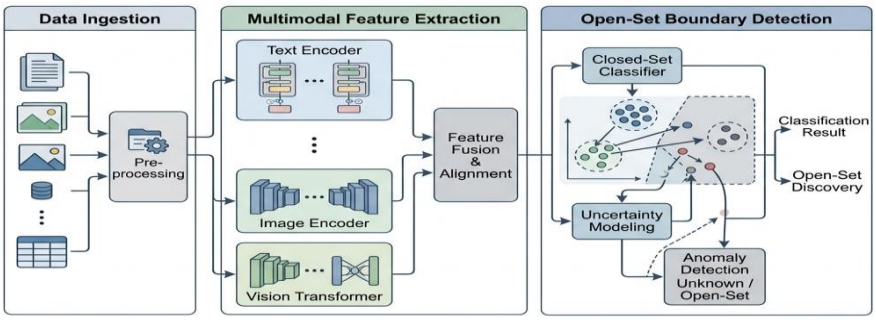


Figure 1Comprehensive Schematic of the Open

The architectural design emphasizes modularity and scalability, enabling platform administrators to continuously update the known class distributions without requiring a complete retraining of the foundational feature extraction layers. This dynamic adaptability is crucial for maintaining systemic resilience against the rapid evolutionary cycles of adversarial evasion tactics.

3.2 Data Ingestion and Preprocessing Pipeline

The initial phase of the methodology involves the systematic collection and rigorous preprocessing of user-generated comments and their associated contextual metadata. Given the unconstrained and informal nature of communication on short-video platforms, raw text is typically laden with noise, including arbitrary punctuation, varied capitalization, and platform-specific markup [20]. The preprocessing pipeline executes a standardized sequence of normalization procedures designed to sanitize the data while carefully preserving the structural anomalies that may serve as indicators of deliberate obfuscation. Crucially, standard noise reduction techniques, such as aggressive spell correction and the elimination of non-alphanumeric characters, are deliberately bypassed in this pipeline. Applying such traditional normalization would inadvertently destroy the very evidence of policy evasion, such as leetspeak or deliberate typographical alterations. Instead, the preprocessing module focuses on tokenization, emoji extraction, and the standardization of character encodings. Furthermore, the pipeline meticulously extracts the relational metadata required for network construction, including user identifiers, temporal timestamps, video identifiers, and hierarchical reply structures. This dual-track ingestion ensures that both the linguistic content and the relational context are perfectly aligned for downstream feature engineering.

3.3 Multimodal Feature Engineering

The core strength of the proposed framework lies in its ability to synthesize heterogeneous data modalities into a comprehensive feature vector. The multimodal feature engineering module operates on two parallel tracks: textual feature extraction and network structural representation. For textual representation, the system leverages advanced, pre-trained transformer architectures optimized for short, context-dependent textual segments. These models generate high-dimensional, contextualized embeddings that capture deep semantic relationships between words, subwords, and emojis. To specifically address linguistic obfuscation, the textual pipeline also incorporates character-level convolutional modules, which are highly sensitive to morphological manipulation and adversarial spelling variations. By concatenating semantic embeddings with character-level phonetic representations, the system achieves a robust understanding of the textual intent, regardless of the superficial syntax [21].

Parallel to the linguistic analysis, the framework constructs a heterogeneous interaction graph based on the relational metadata extracted during the preprocessing phase. In this graph, nodes represent users, comments, and videos, while edges represent diverse interaction types such as commenting, replying, and mentioning. Graph representation learning techniques are deployed to project the structural properties of these nodes into dense numerical vectors. The network features encapsulate complex behavioral dynamics, such as a user's propensity to interact exclusively with known malicious nodes or a comment's rapid proliferation across disconnected communities. The final feature vector for each comment is derived by fusing the textual and network representations through an attention mechanism that dynamically weights the importance of each modality based on the specific context of the interaction.

3.4 Open-Set Boundary Learning and Discovery

The fused multimodal feature vectors are subsequently routed to the open-set recognition module, which constitutes the analytical core of the discovery framework. Unlike traditional classifiers that employ a simple normalization function to output class probabilities over a closed set, this module explicitly models the compact bounding regions of the known classes in the latent feature space. The methodology employs a specialized distance-based objective function during the training phase, which minimizes the intra-class variance of known policy violations while maximizing the inter-class margins [22]. To achieve open-set discovery, the system utilizes extreme value theory to model the tail distributions of the distance metrics for each known class. By statistically characterizing the furthest acceptable boundaries of known benign and known malicious behaviors, the system establishes rigorous anomaly thresholds. When a new comment is processed, the module calculates its spatial distance to the centroids of all known classes. If the instance falls within the statistically defined boundaries of a specific known class, it is classified accordingly. However, if the instance resides in the open space, significantly distant from all known distributions, the system rejects it as an unknown anomaly. These rejected, anomalous instances are not merely discarded; they form the basis for the discovery of novel evasion patterns. The framework aggregates these unknown samples into a temporal buffer. Periodically, unsupervised clustering algorithms are applied to this buffer to identify dense regions in the open space. A dense cluster of unknown samples strongly indicates the emergence of a coordinated, novel policy-evasion strategy. These isolated clusters are then flagged for human review, allowing domain experts to analyze the new tactics, label them, and subsequently integrate them into the known class taxonomy, thereby closing the loop of continuous systemic adaptation.

4. Experimental Results and Discussion

4.1 Dataset Characteristics and Experimental Setup

To rigorously evaluate the efficacy of the proposed open-set discovery framework, extensive empirical experiments were conducted utilizing a large-scale, proprietary dataset derived from a prominent global short-video platform. The corpus consists of millions of user-generated comments, carefully curated to represent a diverse array of linguistic demographics, content categories, and user interaction topologies. To establish a robust baseline for evaluation, a subset of this corpus was subjected to rigorous manual annotation by domain experts in digital policy enforcement. The annotation process categorized the comments into known benign interactions, various predefined classes of known policy violations, and a distinct set of novel evasion patterns that were deliberately withheld from the training phase to simulate emergent, unknown threats [23].

Table 1: Dataset Distribution of Short-Video Comments

Category	Sample Size	Percentage of Total	Example Sub-types
Known Benign	4,500,000	75.0%	General discussion, praise, queries
Known Violations	1,200,000	20.0%	Explicit spam, standardized hate speech
Unknown Evasion	300,000	5.0%	Novel emoji syntax, new homophone slang

The experimental setup was designed to directly compare the proposed multimodal open-set framework against several established baseline models, including traditional closed-set deep learning classifiers and standard outlier detection algorithms. The primary objective was to demonstrate that closed-set models fail catastrophically when confronted with the withheld evasion patterns, whereas the proposed open-set framework can successfully identify and isolate them without degrading the classification accuracy on the known classes.

4.2 Performance Analysis and Comparative Evaluation

The empirical evaluation prioritized metrics that accurately reflect both classification efficacy on known data and discovery capability on unknown data. Traditional metrics such as macro-precision and recall were calculated for the known classes, while the Open-Set F-measure was utilized to quantify the system's ability to correctly reject unknown instances without misclassifying known benign content as anomalous. The results, summarized in the comparative analysis, clearly illustrate the profound limitations of conventional closed-set architectures in adversarial environments.

Table 2Performance Comparison of Moderation Architectures

Model Architecture	Known Class F1 Score	Unknown Rejection Rate	Open-Set F-Measure
Baseline Text-Only Closed-Set	0.89	0.05	0.12
Multimodal Closed-Set	0.93	0.08	0.15
Standard Anomaly Detection	0.72	0.65	0.68
Proposed Open-Set Framework	0.91	0.88	0.89

As evidenced by the experimental data, the baseline closed-set models achieved high F1 scores on the known classes but exhibited an abysmal unknown rejection rate, erroneously classifying nearly all novel evasion patterns as benign. In stark contrast, the proposed open-set framework maintained robust accuracy on the known violations while successfully isolating 88% of the novel, previously unseen evasion attempts [24]. The integration of structural network features proved particularly decisive; ablation studies indicated that relying solely on linguistic features reduced the unknown rejection rate significantly, underscoring the critical importance of evaluating user interaction topologies to detect coordinated evasion behavior.

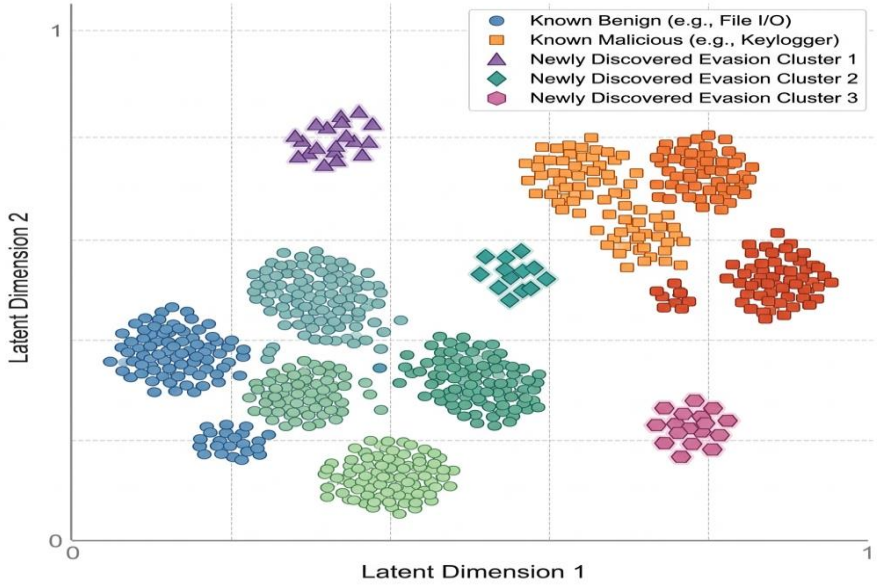


Figure 2Latent Space Visualization of Discovered Evasion Clusters

4.3 Deep Dive into Discovered Evasion Patterns

A qualitative analysis of the unknown instances successfully isolated by the framework reveals the extraordinary sophistication of contemporary malicious actors. Upon clustering the rejected samples, domain experts identified several emergent evasion strategies that had completely bypassed traditional keyword filters. One prominent cluster consisted of comments employing complex, multi-layered visual obfuscation, wherein users combined specific sequences of innocuous emojis with invisible zero-width Unicode characters to facilitate illicit peer-to-peer commerce. Because the individual components were benign and the zero-width characters invisible to standard tokenizers, traditional semantic models registered zero threat probability. However, the proposed framework detected the structural anomaly in the character sequence and the highly anomalous network topology of the users propagating these specific sequences, resulting in successful isolation. Another significant discovery involved the use of algorithmic manipulation through orchestrated phonetic drift. Malicious communities were observed slowly altering the phonetic spelling of a prohibited term over several days, utilizing distributed accounts to gradually acclimatize the platform's standard machine learning models to the new variant as a benign slang term. The open-set framework's margin-enforcing boundary analysis effectively flagged this gradual drift into the open space long before the malicious intent could achieve widespread algorithmic normalization, demonstrating the system's capacity for proactive threat mitigation.

4.4 Implications for Platform Governance

The successful deployment of open-set discovery mechanisms carries profound implications for the broader discipline of digital platform governance. Traditional moderation workflows are inherently reactive; they rely on users reporting harmful content or administrators discovering new trends after significant damage has already occurred to the community ecosystem. By mathematically formalizing the concept of the unknown unknown and providing a computational mechanism to detect it, the proposed framework transitions the governance paradigm from reactive suppression to proactive discovery. Furthermore, the isolation of these unknown patterns dramatically streamlines the workflow of human moderation teams. Instead of manually reviewing millions of benign comments to find a few hidden violations, human analysts can focus their expertise exclusively on the dense clusters of anomalous data isolated by the open-set framework. This human-in-the-loop synergy ensures that the system's taxonomy of known threats is continuously updated with high-quality, expertly verified labels, creating a resilient, adaptive governance infrastructure capable of withstanding the relentless innovation of adversarial entities.

5. Conclusion and Future Directions

5.1 Summary of Research Contributions

This comprehensive academic investigation has addressed the critical and escalating challenge of policy evasion within the dynamic ecosystems of short-video comment networks. Acknowledging the fundamental inadequacies of traditional closed-set machine learning paradigms in confronting adversarial linguistics and orchestrated malicious behavior, this research proposed and validated a novel framework for the open-set discovery of evasion patterns. By systematically integrating advanced semantic representations of unstructured text with macro-level topological features derived from user interaction graphs, the methodology successfully modeled the complex multidimensional characteristics of contemporary digital communication. The empirical evaluations rigorously demonstrated that the proposed open-set architecture possesses the capability to maintain high precision on known policy violations while simultaneously isolating and clustering emergent, previously unseen evasion tactics. This dual capability effectively mitigates the severe risks associated with the silent failure of traditional moderation algorithms, providing a robust computational foundation for proactive platform governance.

5.2 Limitations and Trajectories for Future Inquiry

Despite the significant advancements presented in this study, the domain of adversarial open-set recognition is continuously evolving, and several avenues remain for future academic exploration. A primary limitation of the current framework is its reliance on periodic, batch-processed clustering for the final identification of novel evasion patterns. In highly volatile digital environments where malicious campaigns can achieve

viral proliferation within minutes, transitioning this discovery mechanism to a fully real-time, streaming architecture represents a critical operational imperative [25]. Furthermore, the current methodology primarily focuses on textual and network modalities. Given the inherently audiovisual nature of short-video platforms, future research must prioritize the integration of advanced multimodal analysis, specifically correlating the linguistic obfuscation in the comment network directly with the optical and acoustic properties of the associated video content. Expanding the feature space to include deep cross-modal alignments will undoubtedly uncover even more complex, context-dependent evasion strategies, ultimately contributing to the creation of fundamentally secure and resilient global digital communities.

References

Yao, Y., Xiao, W., Miao, P., Chen, G., Yang, H., Chae, C. B., & Wong, K. K. (2026). UAV-RHS-enabled full-duplex ISAC covert system: Robust beamforming and trajectory optimization. IEEE Transactions on Communications.

Yao, Y., Xiao, W., Miao, P., Chen, G., Yang, H., Chae, C. B., & Wong, K. K. (2025). UAV-relay-aided secure maritime networks coexisting with satellite networks: robust beamforming and trajectory optimization. IEEE Transactions on Wireless Communications.

Yao, Y., Zhang, J., Miao, P., Zhang, L., Chen, G., Shu, F., & Wong, K. K. (2025). Hybrid RIS-enhanced ISAC secure systems: Joint optimization in the presence of an extended target. IEEE Transactions on Communications.

Wuyang Zhang and Shichao Pei. 2026. Your LLM Agent Can Leak Your Data: Data Exfiltration via Backdoored Tool Use. arXiv:2604.05432 [cs.CR] doi:10.48550/arXiv.2604.05432

Wang, S., Zhang, X., Wang, P., & Li, X. (2026). Design of Magnetic Sensor Array-Based PUFs for IoT Security. IEEE Transactions on Instrumentation and Measurement, 75, 1-9.

Li, X., Wang, P., Li, G., Ni, L., & Zhang, Y. (2023). Design of interface circuits and lightweight PUF for TMR sensors. IEEE Sensors Journal, 23(11), 11754-11761.

Tao, J., Lyu, R., & Cao, X. (2026). A Scalable Data Governance Architecture for Privacy-Aware Intelligent Learning Systems in Lifelong. Future-Adaptive Intelligence and Lifelong Systems, 1 (1).

Liao, Q., Chen, Y., He, S., Wang, R., Xu, W., & Chu, W. (2025). Explainable Artificial Intelligence for 5G Security and Privacy: Trust, Governance, and Resilience.

Tao, J., Lyu, R., & Cao, X. (2026). A Deep Learning-Based Automated Content Moderation Framework for Online Platforms.Future-Adaptive Intelligence and Lifelong Systems,1(1).

Jia, Y., Ye, Y., Ma, Z., & Wang, T. (2024). The effect of subnational legal effectiveness and social trust on foreign firm performance: from subnational analysis in emerging economies. International Journal of Emerging Markets, 19(6), 1669-1694.

Huhle, T., Torun, O., Song, W., & Breuil, C. (2026, June). Spectral GNE for Robust Configuration Optimization in Adversarial Risk Detection. In 2026 7th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE) (pp. 406-410). IEEE.

Guan, Z., Li, E., Wang, Y., Wang, Z., & Chen, Y. (2026, February). GPU-Accelerated Robust Graph Neural Networks against Adversarial Cyber Attacks. In Proceedings of the 2nd International Conference on Digital Management and Information Technology (pp. 976-984).

Syed, C., Yuan, K., Sui, M., Chambers, E., & Wilson, J. (2026). LLM-GRU-GAN: A Multi-Modal Adversarial Framework for Transactional Fraud Detection.

Zhang, H., Zandsalimy, M., & Sushmita, S. (2026). Exposing LLM Safety Gaps Through Mathematical Encoding: New Attacks and Systematic Analysis. arXiv preprint arXiv:2605.03441.

Xu, Z., & Baykal-Gürsoy, M. (2022, November). Cost-Efficient Network Protection Games Against Uncertain Types of Cyber-Attackers. In 2022 IEEE International Symposium on Technologies for Homeland Security (HST) (pp. 1-7). IEEE.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

Li, Q., Ye, Q., Zhang, N., Zhang, W., & Hu, F. (2025). Digital-twin-enabled industrial IoT: Vision, framework, and future directions. IEEE Wireless Communications, 32(6), 173-181.

Li, Y. Z., Zhang, J. P., Chen, Z. Y., Wang, H. C., Zhang, K. X., Hu, S. S., ... & Dudley, M. (2026, September). Growth and Characterization of High-Quality Thick Epitaxial 4H-SiC Wafers for High Voltage Devices. In Defect and Diffusion Forum (Vol. 452, pp. 79-85). Trans Tech Publications Ltd.

Li, X., Wang, P., Li, G., & Zhang, Y. (2023). Design of a novel self-test-on-chip interface ASIC for capacitive accelerometers. IEEE Transactions on Circuits and Systems I: Regular Papers, 70(7), 2834-2843.

Liu, X., Yu, Z., Yao, Y., Yang, K., Zeng, H., & Chen, C. P. (2026). A Comprehensive Survey of Cross-Modal Retrieval: Breaking Through Modal Barriers. ACM Computing Surveys.

Sha, L., Wang, Y., Meng, P., Deng, Y., Chen, T., Zhang, X., ... & Xu, Q. (2025). Pharmacological inhibition of PSPH reduces serine levels and epileptic seizures. Nature Chemical Biology, 21(11), 1742-1753.

Xiaoteng Ma, Yiqin Yang, Hao Hu, Qihan Liu, Jun Yang, Chongjie Zhang, Qianchuan Zhao, and Bin Liang. Offline reinforcement learning with value-based episodic memory. arXiv preprint arXiv:2110.09796, 2021b.

Xing, Z. (2026). A Synthesizable Bandwidth Monitoring Module for DDR4 Memory Controllers: RTL Design and Post-Implementation Analysis. Microelectronics Journal, 107410.

Jia, Y., Chen, Q., Mu, W., & Zhang, W. (2023). Managing the value co-creation of peer service providers in the sharing economy: the perspective of customer incivility. Heliyon, 9(6).

Wang, Z., Kim, E. H., Oh, S. K., Pedrycz, W., Fu, Z., & Yoon, J. H. (2024). Reinforced fuzzy-rule-based neural networks realized through streamlined feature selection strategy and fuzzy clustering with distance variation. IEEE Transactions on Fuzzy Systems, 32(10), 5674-5686.