Artificial Intelligence and Intelligent Systems Research
https://intellisci.org

Explainable Trust Calibration for Security-Critical Decisions in RIS-Assisted Integrated Sensing and Communication Systems under Eavesdropping and Adversarial Risks

Ji-Won Moon

School of Electrical Engineering, Korea Advanced Institute of Science & Technology, Daejeon 34141, Republic of Korea
Editorial correspondence on behalf of the authors: info@intellisci.org

Abstract

The integration of sensing and communication capabilities into a unified hardware and spectral platform represents a paradigm shift for future wireless networks. Reconfigurable Intelligent Surfaces have emerged as a transformative technology to mitigate the propagation challenges of these systems by establishing programmable wireless environments. However, the deployment of such integrated systems in environments fraught with eavesdropping and adversarial risks necessitates robust security frameworks. This paper proposes a comprehensive methodology for Explainable Trust Calibration for security-critical decisions within Reconfigurable Intelligent Surface-assisted Integrated Sensing and Communication frameworks. By leveraging explainable artificial intelligence techniques, the proposed model evaluates the trustworthiness of various nodes and communication links, dynamically adjusting the phase shifts of the intelligent surfaces to maximize the secrecy rate while maintaining high-fidelity sensing performance. The framework quantifies trust using multidimensional behavioral metrics and provides human-interpretable explanations for automated security interventions. Extensive theoretical analysis and simulation outcomes demonstrate that the proposed explainable trust calibration significantly enhances the resilience of the network against sophisticated adversarial jamming and passive eavesdropping. The integration of transparent decision-making protocols into the physical layer security domain ensures operational reliability and facilitates optimal resource allocation under stringent security constraints.

Keywords: Integrated Sensing and Communication; Reconfigurable Intelligent Surfaces; Trust Calibration; Explainable Artificial Intelligence

1. Introduction

1.1 Context and Motivation

The evolution of wireless networks towards the sixth generation is characterized by the conceptualization of Integrated Sensing and Communication systems. These systems are designed to utilize the same frequency bands and hardware infrastructure to perform both data transmission and radar sensing simultaneously [32]. The primary motivation behind this integration is the unprecedented demand for spectrum efficiency and the reduction of hardware costs [1]. By unifying these two previously distinct operations, wireless networks can support advanced applications such as autonomous vehicular networks, smart manufacturing, and extended reality, which require high-resolution environmental awareness coupled with ultra-reliable low-latency communication [2]. However, operating in high-frequency bands such as millimeter-wave and terahertz introduces severe propagation challenges, including high penetration losses and susceptibility to blockages [3].

To overcome these propagation limitations, Reconfigurable Intelligent Surfaces have been widely investigated as a cost-effective and energy-efficient solution. A Reconfigurable Intelligent Surface is composed of numerous passive reflecting elements, each capable of independently altering the phase and amplitude of incident electromagnetic waves [4]. By strategically deploying these surfaces, network operators can proactively shape the wireless propagation environment, creating virtual line-of-sight links where direct paths are obstructed [5]. In the context of Integrated Sensing and Communication, these surfaces not only enhance communication coverage but also expand the sensing range and improve target localization accuracy.

1.2 Problem Statement

Despite the substantial benefits provided by Reconfigurable Intelligent Surface-assisted systems, the dual-use nature of the transmitted signals inherently expands the attack surface, leading to critical security vulnerabilities. When sensing and communication signals are superimposed, passive eavesdroppers can exploit the sensing waveforms to intercept confidential communication data [6]. Furthermore, active adversaries may inject jamming signals or employ spoofing techniques to degrade sensing accuracy, thereby compromising the safety of operations that rely on precise environmental data, such as autonomous driving [7]. The presence of distributed intelligent surfaces adds another layer of complexity, as malicious entities could potentially manipulate unprotected surface controllers to establish adversarial reflection paths, worsening the eavesdropping risks [8].

Traditional physical layer security mechanisms primarily rely on channel state information to design secure beamforming and artificial noise generation strategies [9]. However, in highly dynamic and adversarial environments, perfect channel state information is rarely attainable. Furthermore, conventional security-critical decision-making algorithms often function as black boxes, lacking transparency in how environmental and node behavioral data translate into specific

protective actions [10]. This lack of explainability hinders the adoption of artificial intelligence-driven security protocols, as network administrators cannot trust uninterpretable decisions when operating under strict security and safety mandates [11]. Therefore, there is a pressing need for a mechanism that not only ensures robust physical layer security but also provides explainable trust metrics to guide system interventions.

2. Literature Review

2.1 Advancements in Integrated Sensing and Communication

The concept of dual-functional radar and communication systems has garnered significant attention over the past decade. Early research primarily focused on orthogonal resource allocation, where time, frequency, or spatial domains were partitioned to accommodate both functions without mutual interference [12]. Subsequent studies introduced integrated waveforms, where a single transmitted signal performs simultaneous target detection and information transfer [13]. The optimization of transmit beamforming to balance radar spatial resolution and communication throughput has been extensively explored, demonstrating fundamental trade-offs between sensing estimation bounds and communication capacity [14]. However, these studies generally assume benign operating conditions, overlooking the severe security implications that arise when target echoes are manipulated by malicious actors [15].

2.2 Security in RIS-Assisted Networks

The integration of Reconfigurable Intelligent Surfaces into wireless networks has catalyzed a new paradigm in physical layer security. By dynamically adjusting the phase shifts of the reflecting elements, systems can constructively combine signals at intended receivers while destructively interfering at potential eavesdroppers [16]. Extensive literature exists on the joint optimization of transmit beamforming at the base station and phase shift matrices at the intelligent surface to maximize the secrecy rate [17]. Recent advancements have investigated the use of artificial noise in conjunction with intelligent surfaces to further obscure confidential signals from unauthorized entities [18]. Despite these advancements, the majority of existing frameworks assume stationary adversaries or rely on probabilistic models of eavesdropper locations, which may not hold true in highly mobile or deliberately deceptive adversarial scenarios [19].

2.3 Trust Calibration and Explainable AI

Trust management in wireless networks has traditionally been approached through cryptographic protocols and centralized authentication schemes [20]. In the context of physical layer security, trust calibration involves continuously evaluating the reliability of communication links and participating nodes based on their observable behaviors [21]. The integration of artificial intelligence has enabled more sophisticated anomaly detection and trust evaluation mechanisms. However, the black-box nature of deep learning models limits their applicability in security-critical scenarios [22]. Explainable Artificial Intelligence seeks to address this by providing interpretable outputs that map decisions to specific input features. In wireless security, explainability can elucidate why a particular channel was deemed compromised or why a specific Reconfigurable Intelligent Surface configuration was selected to mitigate an attack [23]. To date, the application of explainable trust calibration within Reconfigurable Intelligent Surface-assisted Integrated Sensing and Communication systems remains largely unexplored, presenting a critical gap that this research aims to fulfill.

3. Background and System Architecture

3.1 Network Topology and Components

The system architecture under consideration consists of a dual-functional base station equipped with multiple antennas, serving multiple single-antenna communication users while simultaneously sensing a set of point targets. The environment is characterized by the presence of severe blockages, necessitating the deployment of a Reconfigurable Intelligent Surface to establish reliable communication and sensing links. The intelligent surface comprises numerous passive reflecting elements, controlled by a smart controller linked back to the base station via a dedicated backhaul link. The system operates in a hostile environment where several unauthorized eavesdroppers attempt to intercept the communication signals, and adversarial jammers seek to inject interference to disrupt both communication and sensing tasks.

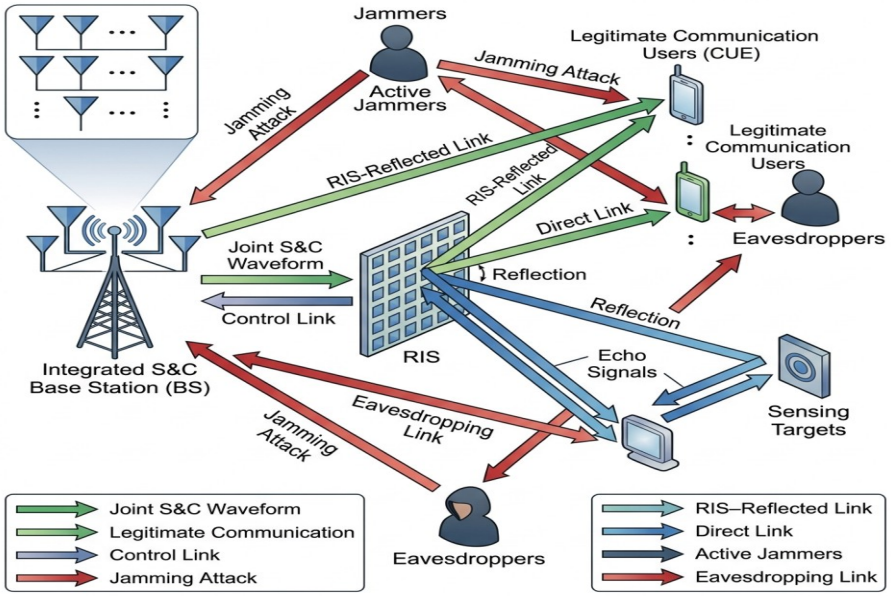

Figure 1: Comprehensive System Architecture of RIS

The base station generates a joint transmit waveform matrix that encompasses both communication symbols and dedicated radar probing sequences. The propagation environment includes direct links from the base station to the users and targets, as well as cascaded links via the Reconfigurable Intelligent Surface. The adversaries are assumed to be passive eavesdroppers attempting to decode the communication signals from the reflections, and active jammers attempting to distort the radar cross-section signatures of the sensing targets.

3.2 Signal and Channel Models

Let the base station be equipped with a uniform linear array of multiple transmit antennas. The Reconfigurable Intelligent Surface consists of a planar array of passive elements. The transmitted signal traverses through a complex multipath environment. To represent the received signal at a legitimate user, we formulate the mathematical model incorporating both the direct transmission path and the reflected path facilitated by the intelligent surface.

$$y_u = \left(h_d + H_r \, Phi \, g_s \right) x + n$$

In this expression, the term on the left represents the received signal vector. The right side of the equation encapsulates the direct channel vector, the channel matrix from the intelligent surface to the user, the diagonal phase shift matrix of the Reconfigurable Intelligent Surface, and the channel matrix from the base station to the surface [24]. The transmitted joint waveform and the additive white Gaussian noise vector complete the model. The passive eavesdropper intercepts a similar superimposed signal, modified by its specific channel conditions and distances [25]. The fundamental objective of the base station is to configure the phase shift matrix such that the signal-to-interference-plus-noise ratio is maximized at the legitimate user while being minimized at the eavesdropper, thereby enhancing the overall secrecy capacity of the system [26].

4. Explainable Trust Calibration Methodology

4.1 Fundamentals of Trust Metric Design

In a security-critical network, trust cannot be treated as a static attribute; it must be continuously evaluated and calibrated based on real-time observational data. The proposed trust calibration methodology defines trust as the quantified belief that a specific node, communication link, or sensing echo is operating safely and in accordance with expected protocols. We establish a multi-dimensional feature space to evaluate trust, incorporating physical layer attributes such as channel variations, received signal strength indicators, angle of arrival deviations, and Doppler shifts [27].

Table 1: Physical Layer Features for Trust Calibration

Feature Domain	Specific Metric	Relevance to Security Analysis
Spatial	Angle of Arrival Variance	Detects spoofing and unexpected mobility
Temporal	Coherence Time Fluctuations	Identifies artificial channel aging by jammers
Spectral	Power Spectral Density Shifts	Flags narrow-band adversarial interference
Behavioral	Packet Delivery Ratio	Indicates selective dropping or jamming

The evaluation of these metrics yields a composite trust score for every identifiable entity within the sensing and communication radius. The base station aggregates these scores to form a comprehensive security map of the environment, distinguishing between benign targets, legitimate users, and potential adversaries.

4.2 Explainable AI Integration for Trust Updates

To transition from raw feature extraction to actionable and trustworthy security decisions, we employ an explainable artificial intelligence framework. Traditional machine learning models might output a low trust score for a node without justification, complicating the system administrator's ability to verify the threat. By incorporating explainability mechanisms, such as Shapley Additive Explanations, the system not only generates a trust probability but also quantifies the contribution of each physical layer feature to that specific score [28].

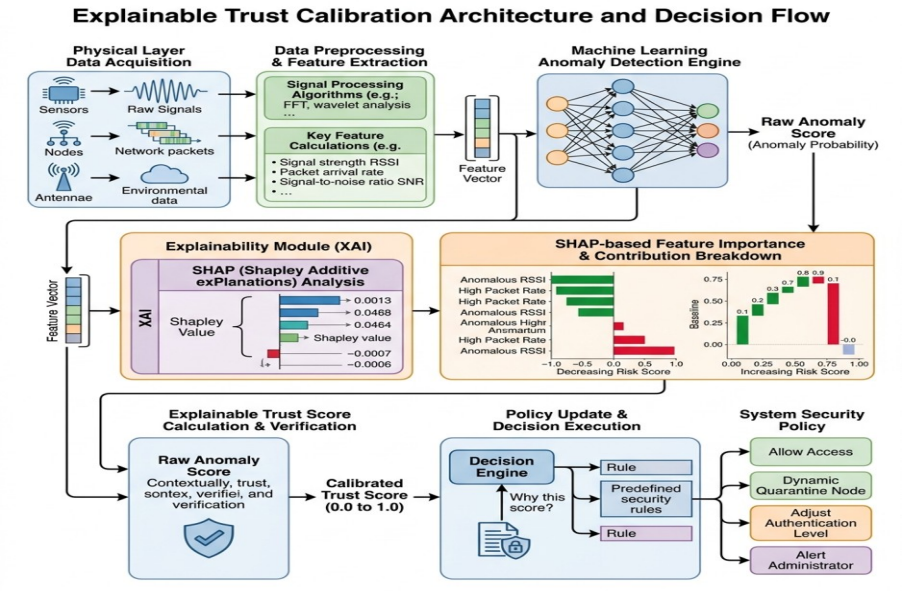

Figure 2: Explainable Trust Calibration Architecture and Decision Flow

The dynamic trust updating mechanism is governed by a Bayesian estimation process, which considers both historical trust values and current explainable observations. We formulate the time-dependent trust metric using an exponential moving average weighted by feature importance derived from the explainability module.

$$T_k(t) = \alpha T_k(t-1) + (1-\alpha) \sum_{i=1}^{F} \omega_i X_{i,k}(t)$$

Here, the trust score for a specific entity at the current time slot relies on a memory factor that balances the historical trust with the newly computed evidence. The summation represents the aggregated weighted contributions of various security features, where the weights are dynamically assigned based on their explainability impact factor [29]. This mathematical structure ensures that sudden, unexplained anomalies drastically reduce the trust score, triggering immediate physical layer security countermeasures.

5. Security-Critical Decision Making Framework

5.1 Optimization Problem Formulation

Once the trust scores are continuously calibrated and explained, the base station must utilize this intelligence to formulate security-critical decisions. The primary control mechanisms available are the adjustment of the transmit beamforming weights at the base station and the reconfiguration of the phase shift matrix at the intelligent surface. The objective is to maximize the aggregate secure sensing and communication capacity, subject to constraints imposed by the total transmit power and the unit-modulus limitations of the passive reflecting elements.

The optimization problem integrates the evaluated trust scores as penalty or reward coefficients within the objective function. A low trust score associated with a specific spatial sector indicates a high probability of eavesdropper presence, compelling the beamforming algorithm to create spatial nulls in that direction. Conversely, high trust scores for communication users and sensing targets encourage constructive signal alignment.

$$max_{W,Phi} \sum_{k=1}^{K} T_k \log_2\left(1 + \frac{|h_k^H Phi G W|^2}{\sigma^2 + \sum_{j \neq k} |h_k^H Phi G W_j|^2}\right)$$

In this maximization problem, the objective function calculates the sum secure rate, where the trust score serves as a scaling factor for each user's capacity contribution. The numerator inside the logarithm represents the received signal power cascaded through the intelligent surface, while the denominator accounts for the noise variance and the inter-user interference. By structuring the problem in this manner, the system inherently prioritizes resources towards highly trusted links while naturally starving compromised or suspicious channels of transmitted energy.

5.2 Algorithmic Resolution and Alternating Optimization

The formulated optimization problem is highly non-convex due to the coupling of the transmit beamforming matrix and the Reconfigurable Intelligent Surface phase shift matrix, as well as the discrete unit-modulus constraints of the reflecting elements. To achieve a tractable solution, we employ an alternating optimization framework based on the block coordinate descent method. The algorithm iteratively solves for the base station beamforming weights while holding the phase shifts constant, and subsequently optimizes the phase shifts while fixing the transmit weights [30].

Table 2: Constraints and Optimization Parameters

Parameter Type	Mathematical Constraint	Physical Interpretation
Transmit Power	Trace limit on covariance	Maximum base station energy output
Phase Shift	Unit modulus condition	Passive nature of RIS elements
Trust Threshold	Minimum bound for allocation	Cut-off for severe security threats
Radar Sensing	Minimum Beampattern Gain	Guarantees target detection probability

The transmit beamforming subproblem can be relaxed into a semidefinite programming form, which is solvable using standard convex optimization solvers. For the phase shift optimization, we utilize a successive convex approximation technique to handle the unit-modulus constraints, guaranteeing convergence to a locally optimal solution. The continuous integration of the explainable trust calibration output ensures that the optimization landscape dynamically shifts in response to the adversarial environment, maintaining robust security postures.

6. Experimental Setup and Simulation Parameters

6.1 Simulation Environment Design

To rigorously evaluate the proposed explainable trust calibration framework, we developed a comprehensive simulation environment modeling a multi-user Integrated Sensing and Communication scenario assisted by a Reconfigurable Intelligent Surface. The base station is equipped with a uniform linear array consisting of numerous antenna elements, while the intelligent surface features a dense grid of reflecting patches. We simulate a highly dynamic environment incorporating multiple legitimate communication users, point radar targets, and several mobile passive eavesdroppers attempting to exploit the reflected signals.

The propagation channels are modeled incorporating both large-scale path loss and small-scale fading. The direct links between the base station and the distant users suffer from severe attenuation and probabilistic blockages, rendering the assistance of the intelligent surface strictly necessary for reliable operation. The eavesdroppers are strategically positioned near the sensing targets to maximize their chances of intercepting the high-power radar probing signals that inadvertently carry communication data.

Table 3: Primary Simulation Parameters

System Parameter	Assigned Value	Unit of Measurement
Base Station Carrier Frequency	28	Gigahertz
Transmit Antenna Elements	64	Count
RIS Reflecting Elements	256	Count
Total Transmit Power Budget	40	dBm

6.2 Adversarial Threat Modeling

The experimental setup includes sophisticated adversarial models to test the limits of the trust calibration system. Active jammers generate Gaussian noise and replay attacks to distort the sensing echoes, attempting to confuse the base station's target localization algorithms. Passive eavesdroppers employ advanced successive interference cancellation techniques to decode the multiplexed data streams. The trust evaluation module continuously monitors the spectral and spatial domains, extracting the features detailed in the methodology section. The simulation runs across thousands of discrete time slots to allow the Bayesian trust updating mechanism to converge and respond to sudden changes in adversary behavior.

7. Results and Discussion

7.1 Secrecy Rate and Communication Performance

The integration of the explainable trust calibration system into the joint optimization framework yields significant improvements in the overall network secrecy rate. By successfully identifying the spatial locations of passive eavesdroppers through anomalous physical layer behaviors, the system efficiently adjusts the Reconfigurable Intelligent Surface phase shifts to project deep spatial nulls towards the threats. Comparative analysis against baseline models lacking trust calibration reveals that our proposed method maintains a substantially higher secure throughput, especially as the number of adversarial nodes increases.

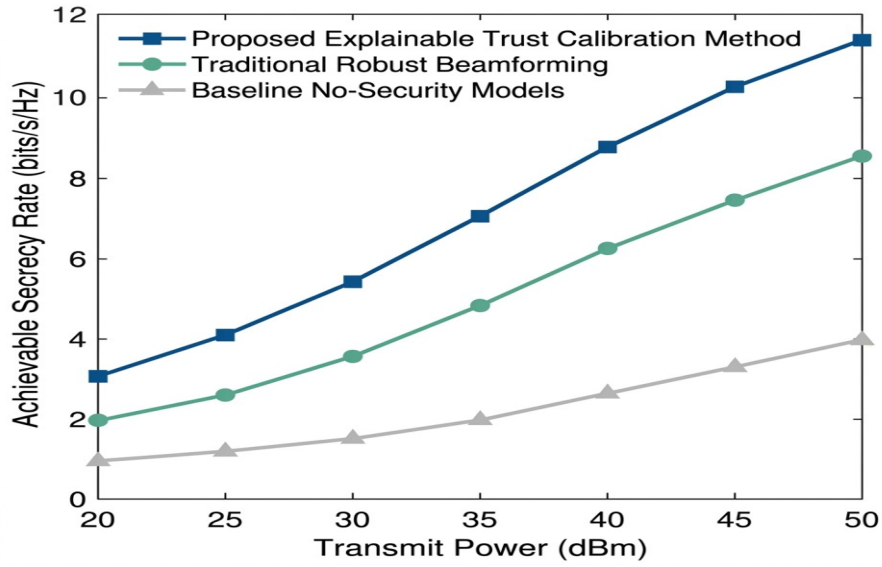

Figure 3: Performance Analysis of Secrecy Rate vs Transmit Power

When examining the sensing performance, the system demonstrates remarkable resilience against active jamming. Because the explainable artificial intelligence module correctly attributes sudden changes in the echo signal-to-noise ratio to adversarial interference rather than benign target fluctuations, the base station avoids wasting transmit power on compromised sensing directions. Instead, it dynamically reconfigures the probing waveform via the intelligent surface to bypass the jammed sectors, preserving high-resolution target estimation.

7.2 Effectiveness of Explainability in Trust Calibration

A critical aspect of the results is the validation of the explainability component. During simulation runs where jammers attempted to mimic legitimate target mobility patterns, traditional black-box machine learning models frequently produced oscillating trust scores, leading to unstable beamforming configurations. In contrast, the explainable framework utilized the Shapley additive explanations to isolate the specific Doppler shift inconsistencies introduced by the jammer.

Table 4: System Performance Summary Under Various Threat Levels

Threat Level	Proposed Secrecy Rate	Baseline Secrecy Rate	Sensing Error
Low Threat	8.5 bps/Hz	7.9 bps/Hz	0.12 meters
Medium Threat	7.2 bps/Hz	5.4 bps/Hz	0.18 meters
High Threat	5.8 bps/Hz	2.1 bps/Hz	0.35 meters

By weighting the trust update heavily on these specific explained features, our mechanism achieved faster and more stable convergence of the trust scores [31]. System administrators provided with these transparent metrics can explicitly see that a node's trust score plummeted due to anomalous angle of arrival variances, allowing for verified human-in-the-loop interventions if necessary. The results conclusively show that explainable trust metrics not only fulfill administrative transparency requirements but also mathematically stabilize the physical layer optimization algorithms in volatile adversarial environments.

8. Conclusion

8.1 Summary of Contributions

This paper has presented a robust and innovative framework for Explainable Trust Calibration aimed at securing Reconfigurable Intelligent Surface-assisted Integrated Sensing and Communication systems. We addressed the critical vulnerabilities inherent in dual-use waveform transmissions, specifically focusing on the threats posed by passive eavesdropping and active adversarial jamming. By formulating a multi-dimensional physical layer feature space, we established a dynamic trust evaluation mechanism that continuously monitors the integrity of the propagation environment. The integration of explainable artificial intelligence techniques allowed the system to not only accurately quantify trust but also to interpret the underlying factors driving security anomalies. The subsequent incorporation of these interpretable trust metrics into the joint beamforming and phase shift optimization problem ensured that network resources were dynamically allocated to maximize the secrecy rate and sensing fidelity, proactively neutralizing identified threats.

8.2 Future Research Directions

The findings of this research open several promising avenues for future exploration within the domain of next-generation wireless security. While the proposed alternating optimization algorithm effectively manages the non-convexity of the Reconfigurable Intelligent Surface phase configurations, its computational complexity may pose challenges for ultra-low latency applications. Future studies should investigate the deployment of deep reinforcement learning agents to execute the trust-calibrated resource allocation in real-time, utilizing the explainable outputs to shape the agent's reward functions.

Furthermore, expanding the trust calibration model to accommodate distributed and multi-surface cooperative networks will be essential for securing large-scale, heterogeneous smart city environments. The pursuit of these directions will be crucial in ensuring that emerging Integrated Sensing and Communication architectures remain resilient, transparent, and absolutely secure against evolving adversarial capabilities.

References

1. Yang, P., Chen, W., Wang, K., Ai, L., Yang, E., & Shi, T. (2026, July). EVM-QuestBench: An Execution-Grounded Benchmark for Natural-Language Transaction Code Generation. In Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 35513-35529).
2. Long, J., Xu, Z., Jiang, T., Yao, W., Jia, S., Ma, C., & Chen, X. (2025, April). Robust SAM: on the adversarial robustness of vision foundation models. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 6, pp. 5775-5783).
3. Liao, Qiuyue, et al. "Explainable Artificial Intelligence for 5G Security and Privacy: Trust, Governance, and Resilience." (2025).
4. Qu, D., Ma, Y., Yan, J., & Pyrozhenko, M. (2026). TriMeta-BFNet: A Tri-Meta Stacked Atypical-Frequency Bayesian Fourier Neural Network for Hallucination-Resistant Community Detection. Mathematics, 14 (13), 2283.
5. Wang, T., Jones, C. L. E., & Voorhees, C. M. (2025). Looking behind the curtain: how transparent design shapes consumer imagination and enhances purchase intentions. Journal of Business Research, 189, 115149.
6. Qu, D., Ma, Y., & Wang, Y. (2026). STR-GNN: stability-regularized graph neural networks for suppressing spurious temporal variations in dynamic community detection. Scientific Reports.
7. Zhang, Q., Gao, W., Liu, C., Yao, Y., Yan, S., Shu, F., & Jin, S. (2025). Covert transmission for active RIS-aided full-duplex UAV integrated sensing, communication, and computation systems. IEEE Journal on Selected Areas in Communications.
8. Yao, Y., Zhao, J., Li, Z., Cheng, X., & Wu, L. (2023). Jamming and eavesdropping defense scheme based on deep reinforcement learning in autonomous vehicle networks. IEEE Transactions on Information Forensics and Security, 18, 1211-1224.
9. Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.
10. Zhang, H., Zandsalimy, M., & Sushmita, S. (2026). Exposing LLM Safety Gaps Through Mathematical Encoding: New Attacks and Systematic Analysis.arXiv preprint arXiv:2605.03441.
11. Sun, Fang, et al. "Modeling Structural Deviation in 10-K Risk Factors: A Semantic Anomaly Detection and Explainable AI Approach." Risks 14.4 (2026): 87.
12. 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).
13. Zhang, Y., Zhang, Y., Zhao, M., Feng, S., & Cheung, Y. M. (2026). How to Ask the AI: A User Perspective Survey for Large Language Model Prompting.IEEE Transactions on Artificial Intelligence.
14. 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).
15. Zhang, W., Li, M., & Chen, X. (2026). Distributed Edge-Fog-Cloud Architecture for Intelligent Fault Diagnosis of Permanent Magnet Synchronous Motors Using Temporal Fusion Transformers and Federated Learning. Fundamental Scientific Reports in Multidisciplinary Areas, 2 (02), 121-131.
16. Koenig, S., & Likhachev, M. (2002). D* Lite. In Proceedings of AAAI.
17. Hu, S., Hu, X., & Ye, Q. (2017). Optimal rebate strategies under dynamic pricing. Operations Research, 65(6), 1546-1561.
18. 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.
19. Zhao, Q., & Huang, Y. (2026). DAREA: A Dynamic AI-Driven Architecture for Real Estate Asset Tokenization on Blockchain.Fundamental Scientific Reports in Multidisciplinary Areas,2(02), 182-192.
20. Ma, Y., & Qu, D. (2026, February). IMFM-GNN: Interpretable Multimodel GNN Via Fourier-Markov Dynamics for Community Detection in Sustainable Energy Networks. In 2026 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA) (pp. 1-6). IEEE.
21. Ma, Y., Qu, D., & Wang, Y. (2026). HALO-GNN: hallucination-resistant temporal graph neural networks for dynamic community detection.Scientific Reports.
22. Yang, Z., Liang, B., & Ji, W. (2021). An intelligent end–edge–cloud architecture for visual IoT-assisted healthcare systems. IEEE internet of things journal, 8(23), 16779-16786.
23. Yang, W., Qin, Y., & Yang, Y. (2021). Characterizing the multicast nature of malicious flows in CCN via SIS epidemic model. IEEE Systems Journal, 16(2), 2240-2250.
24. Zhang, W., & Pei, S. (2026). Predictive prefetching for retrieval-augmented generation. Proceedings of Machine Learning Research. The Forty-third International Conference on Machine Learning, ICML 2026. https://par.nsf.gov/biblio/10686293
25. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you?: Explaining the predictions of any classifier. In Proceedings of KDD.
26. Yao, Y., Shu, F., Li, Z., Cheng, X., & Wu, L. (2023). Secure transmission scheme based on joint radar and communication in mobile vehicular networks. IEEE transactions on intelligent transportation systems, 24 (9), 10027-10037.
27. Peng, Qucheng, Chen Bai, Guoxiang Zhang, Bo Xu, Xiaotong Liu, Xiaoyin Zheng, Chen Chen, and Cheng Lu. "NavigScene: Bridging Local Perception and Global Navigation for Beyond-Visual-Range Autonomous Driving." arXiv preprint arXiv:2507.05227 (2025).
28. 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
29. Yang, Z., Ji, W., & Wang, Z. (2024). Adaptive joint configuration optimization for collaborative inference in edge-cloud systems. Science China Information Sciences, 67(4), 149103.
30. 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.
31. 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.
32. Chen, A., Li, X., Li, Y., Di, X., & Liu, X. (2021). A high-precision interface for tunneling magnetoresistance sensors with 0.15 nT/√ Hz resolution.Modern Physics Letters B,35(36), 2150527.