PrivSev: Privacy-Preserving Artificial Intelligence in 6G Open Radio Access Networks: A Survey

N. Nelufule*, N. Siphambili, D. Shadung
Council for Scientific and Industrial Research (CSIR)
Defense and Security Cluster
Information and Cybersecurity Centre (ICSC), Pretoria, South Africa
nnelufule@csir.co.za

***Abstract*—The disaggregated, multi-vendor architecture of Open Radio Access Networks (O-RAN) in 6G, promises an unprecedented flexibility through the AI-native intelligence, and cost efficiency. However, these benefits also introduce challenges such as severe privacy and security risks which include the model inversion, data poisoning, and unauthorized access across distributed edge nodes; mainly Open Radio Unit (O-RU), Open Distributed Unit (O-DU), Open Centralized Unit (O-CU), and RAN Intelligent Controllers (RICs). In this paper, a systematic review based on the PRISMA framework was used to synthesize 42 peer-reviewed articles published between 2020 and 2026, particularly on the privacy-preserving AI techniques, Federated Learning (FL), Differential Privacy (DP), Secure Multi-Party Computation (SMPC), and emerging hybrid technologies applied to 6G O-RAN environments. The key research findings revealed that the combination of the Zero trust Architecture (ZTA) and FL can achieve up to 32% energy savings and Near-RT compliance, while the combination of DP and FL helps to secure the RIC and FBMP, and the Intrusion Detection System (IDS) helps to enable lightweight Multi-Party Computation (MPC). The notion of introducing a three-way FL, DP and SMPC integration for O-RAN remains unexplored, and this work bridges this gap, by introducing Privacy-Preserving (PrivSev), which is a novel layered hybrid framework that applies lightweight DP at the edge clients and threshold SMPC at the Non-RT RIC. The projected performance of the proposed framework tested against the reviewed benchmarks promises a 92% baseline accuracy retention.**

Keywords—6G, Open RAN, Federated Learning, Differential Privacy, Secure Multi-Party Computation, Privacy-Preserving AI, Zero-Trust Architecture, O-RAN WG11.

I. INTRODUCTION

The sixth generation (6G) wireless era envisions the terabit-per-second throughput, sub-millisecond latency, pervasive intelligence, and seamless integration of massive Internet of Things (IoT), digital twins, and holographic communications [1]-[2], [3], [4]. What is central to this vision is the Open Radio Access Network (O-RAN) architecture, that is standardized by the O-RAN ALLIANCE, which disaggregates the traditional monolithic RAN into interoperable, multi-vendor components such as O-Radio Units (O-RU), O-Distributed Units (O-DU), O-Centralized Units (O-CU), Near-RT and Non-RT RAN Intelligent Controllers (RICs), and the Service Management and Orchestration (SMO) framework, as shown in Figure 1 [5],[6]. The main benefits of this openness is that it accelerates innovation, reduces costs, and enables AI-driven dynamic optimization through the specialized applications which are used in a network automation within the context RAN such as xApps and rApps [7].

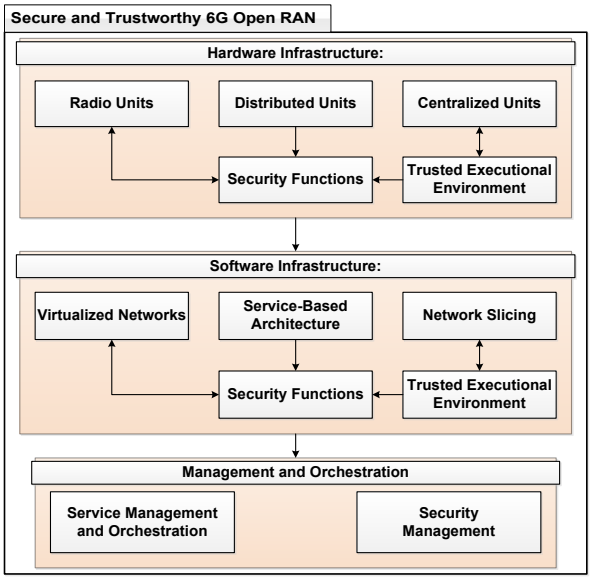

Figure 1. An overview of a Secure and Trustworthy 6G Open RAN

However, this same openness that defines O-RAN also introduces an unprecedented privacy and security challenges [7], [8], [9], [10], [11]. Since the 6G deployments are gaining traction and the operators demand a privacy-by-design solutions that preserve utility without compromising real-time constraints. The existing standalone approaches offer limited capability and have security limitations particularly in pure FL that exposes the gradients to inference; DP degrades accuracy in non-IID O-RAN traffic; and SMPC incurs prohibitive overhead [12], [13].

The disaggregation of O-RAN and the multi-vendor design potentially expands the attack surfaces particularly on the distributed edge nodes and open interfaces [11]. The *"WG11 2025 threat modeling and security specifications"* highlight that the existing privacy-enhancing technologies fail to simultaneously deliver formal guarantees, real-time compliance, and robustness in heterogeneous topologies. In the surveyed literatures there is no evidence of an integrated solution that fully addresses these challenges in an O-RAN's unique real-time, multi-vendor context.

This work was funded by the Department of Science, Technology and Innovation (DSTI), Pretoria, South Africa.

This systematic review offers a detailed synthesis of the rapidly evolving landscape which maps the key themes and taxonomies. The review also offers a critical discussion on the peer-reviewed articles which discuss topics relating to privacy-preserving AI frameworks. In addition to the survey, this paper also proposes a layered hybrid framework from an evidence-based insights as a future, deployable solution for trustworthy 6G O-RAN as shown in Figure 2.

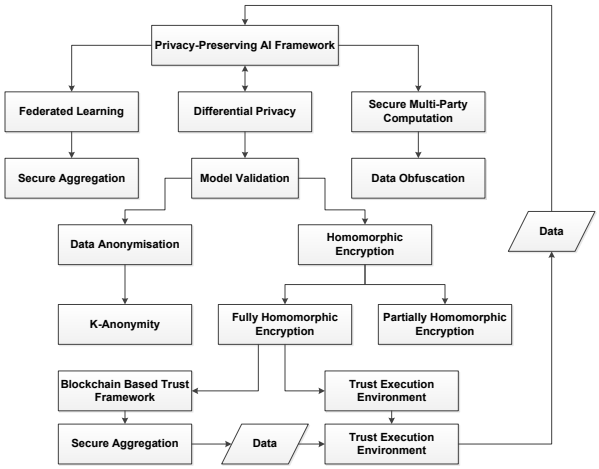

Figure 2. An overview of a Privacy-Preserving AI Framework

The main objectives of this research were to systematically review the privacy-preserving AI techniques in 6G O-RAN. The survey was used to synthesize a novel hybrid framework which addresses the identified research gaps; maps out the key themes; and identify the emerging trends, and taxonomies with a critical discussion on their strengths, limitations, and research gaps related to O-RAN requirements. The results of the survey were also used to synthesize and measure the performance of the proposed hybrid PrivSev framework which combines a layered (FL, DP, SMPC) frameworks. The main research contributions are:

- A PRISMA-compliant systematic literature review which focused exclusively on FL, DP, SMPC, and hybrids in 6G O-RAN contexts.
- A novel taxonomy of privacy-preserving AI in O-RAN, which maps out different techniques to architecture layers, and protection levels.
- Critical gap analysis of some specific studies that are aligned with the WG11 2025 requirements, which reveals the lack of a combination of FL, DP, and SMPC layering.
- A synthesized PrivSev framework which is a practical layered hybrid with projected superior trade-offs.

The remainder of the paper is organized as follows: Section II presents the methodology used in this paper, Section III presents the literature review, section IV presents the results, and Section V concludes the paper.

II. METHODOLOGY

To achieve the research objectives, the following research questions have been identified to guide this research:

- What are the dominant themes, taxonomies, and trends in privacy-preserving AI for 6G O-RAN particularly in the period 2020–2026?
- How do FL, DP, SMPC, and their hybrids perform against accuracy, latency, privacy guarantees, and attack resistance in O-RAN topologies?
- What critical gaps exist in the specific studies relative to real-time constraints, multi-vendor heterogeneity, and adversarial robustness?
- Can a layered (FL, DP, SMPC) hybrid (PrivSev) address these gaps while meeting O-RAN performance budgets?

In order to address these research questions, this work used two methodological parts, namely; the systematic review which followed the PRISMA framework and the experimental setup for the proposed PriSev framework. The PRISMA framework prescribes the following phases as prescribed in [14], [15].

A. Identification Phases

The identification phase provides ideas on the process of searching for the research materials. In this study, the research materials were identified from the three main indexing databases namely the *IEEE Xplore, Scopus,* and *Web of Science*. The main reason for picking these three databases was that they index high quality peer-reviewed conference papers and journal articles.

A search strategy was also developed to retrieve research articles from these three databases, using a combination of the following search phrases *"Open RAN", "O-RAN", "6G", "Federated Learning", "Differential Privacy", "Secure Multi-Party Computation", "SMPC", "privacy-preserving", "Zero-Trust".*

B. Screening Phase

This phase presents a strategy to screen research articles which are required for analysis. The research articles which fall outside the scope of this work are then screened out. In this case the required research articles were screened based on the research relevance, written and published in English, published on the computer science and engineering disciplines, published between the years 2020 and 2026.

The research articles which did not cover the topics presented in this paper were discarded. Since the articles were acquired from three sources, there were also duplicates which were also removed to avoid data redundancy.

C. Inclusion and Exclusion Criteria

This phase provides guidance on how to include and exclude research articles. In this study, the included articles were peer-reviewed articles published in English, discussing frameworks, empirical evaluation, deployment insights on privacy AI in O-RAN/6G. The papers should have been

published between the year 2020 and 2026, particularly on Computer Science and Engineering disciplines.

The excluded articles were those published before 2020, presenting work outside the scope of privacy AI on O-RAN and 6G. Research articles which were published in other languages besides English were also excluded. Other exclusion criteria were based on the full open access and availability of full texts. A pictorial representation of the PRISMA framework as exploited in this paper is depicted in Figure 3.

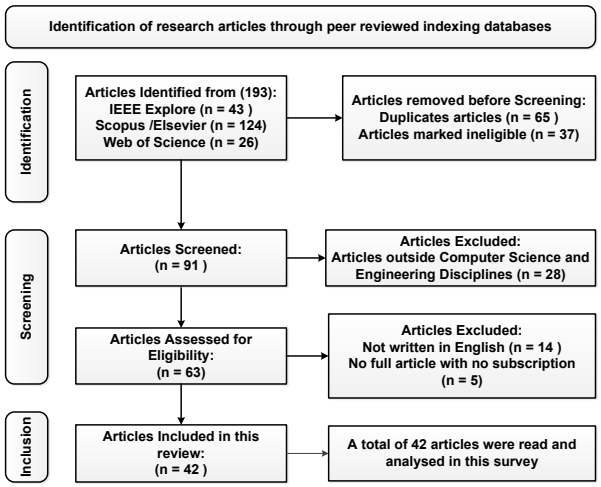

Figure 3. A pictorial representation of a PRISMA framework

D. Experimental Materials and Setup

This section provides the details of the experimental materials required to conduct the proposed PrivSev framework. This experiment was designed to simulate a realistic 6G O-RAN environment.

The required primary dataset included synthetic network telemetry traces with (12,000 samples) which are generated by extending the publicly available Distributed AI Wireless Network and OpenRAN Gym benchmarks [16]. Each sample contains 10 features which included the PRB utilization, RSRP/RSRQ, HARQ feedback, slice type (eMBB, URLLC, mMTC), latency, packet loss rate, mobility index, traffic volume, interference level, and energy consumption.

III. LITERATURE REVIEW

A. O-RAN Capabilities and ZTA Integration

The research on the privacy-preserving and secure AI in 6G O-RAN is rapidly maturing, with a clear shift from conceptual ZTA surveys to hybrid frameworks and threat analyses. Mehrban *et al.*, [17] presented an explicit ZTA-to-O-RAN mapping model which places the policy engine and the enforcement points across the RIC layers, interfaces, and disaggregated functions. This model introduces the Risk-Adaptive Access Control (RAdAC) and a blockchain-based decentralized identity with a smart-contract authorization for xApp and rApp lifecycle. The challenge with this survey is that it lacks empirical implementation and the integration of other framework such as AI, FL, DP, SMPC.

Ramezanpour and Jagannath [18] presented an i-ZTA concept with the Monitor-Evaluate-Decide (MED) loop using ML for the dynamic trust in O-RAN. The paper pioneered the linking of the ZTA with O-RAN SBA design, even though their work lacks critical evaluation on formal privacy guarantees and cryptographic aggregation.

Hashem Eiza *et al.,* [7] presented a practical hybrid ZT deployment model tailored to O-RAN specifications, which included a continuous verification and micro-segmentation. Their work focused on the architectural deployment without deep AI and FL integration and did not discuss the formal privacy metrics. The combination of various privacy-enhancing technologies can amplify their work.

Brik *et al.*, [19] presented a comprehensive XAI taxonomy for O-RAN and maps these techniques to RIC and xApps with security and automation focus. The only limitation is that the presented work is explainability-centric and does not address the core privacy gradients and cryptographic aggregation. A potential solution to this limitation is to couple an XAI with PETs for a trustworthy explanation. Arnaz *et al.,* [20] presented another thorough survey on an AI-driven programmability in O-RAN even though it offered minimal security and privacy treatment.

El-Hajj *et al.*, [1] introduced a landmark ZTA and FL framework which achieved a 32% energy savings, privacy, and a Near-RT compliance on Open RAN Gym. Their work only relied on a basic secure aggregation without discussing a threshold SMPC, and a lightweight DP. The addition of a layered DP, SMPC and a robust aggregation to this work can improve the effectiveness of the presented model.

Chaskos *et al.*, [13], presented a novel theoretical cyber-range topology for an O-RAN and 6G validation which addressed the disaggregation and open-interface threats. The work, however, lacks details on the implanted prototype with some specific PET testing. Porambage *et al.,* [21] presented a broad overview of a security, privacy, and trust issues in O-RAN 6G. This overview also lacks a concrete framework and performance evaluation. An implementation of a hybrids model can advance this survey with practical solution and formal metrics.

Nahar *et al.,* [22] presented a general ZTA applications across the 6G and O-RAN. However, the presented work does not appear to be O-RAN-specific and lacks the AI and FL integration with a real-time analysis. Baguer *et al.,* [23] presented a practical threat modeling and real attacks on the model inversion, and model poisoning. The works have demonstrated an attack-only experiment with no mitigation framework and privacy-preserving countermeasures.

The articles discussed collectively established a strong architectural foundations for ZTA mapping, i-ZTA, slicing, XAIs.

B. ZTA and FL Frameworks Specific to O-RAN Optimization

Research shows that the technologies such as ZTA and FL are enhancing the security and trustworthiness of 6G O-RAN. El-Hajj *et al.*, [1] presented a practical landmark of the combination of ZTA and FL framework with continuous authentication, micro-segmentation, and privacy-preserving collaborative learning across RIC, xApps, and rApps. This

framework demonstrated a 32% energy savings in DU sleep-mode scheduling, and ε = 2.0 guarantees, and Near-RT compliance on an Open RAN Gym. However, the presented model only used a basic secure aggregation with no threshold SMPC or cryptographic protection against gradient leakage, with no lightweight DP at the edge clients. This framework can be advanced by introducing a layered DP noise at O-DU and O-CU clients and threshold SMPC at Non-RT RIC with trimmed-mean robust aggregation. Brik *et al.,* [19], presented a dedicated XAI taxonomy for O-RAN, which mapped a post-hoc and explanation-guided techniques to RIC and xApps with emphasis on security, automation, and slice management. Their study focused exclusively on the explainability; but lacks the gradient leakage, model inversion, and the cryptographic protection of explanations; and formal privacy guarantees. The combination of XAI with privacy-preserving training can aid in generating explanations on a DP-noised and SMPC-secured models. The proposed PrivSev can also complement this by providing an underlying trustworthy AI layer that will enable an explainable decisions but without exposing the raw training data.

Eiza *et al.*, [7] presented a hybrid ZT deployment blueprint with continuous verification, micro-segmentation, and policy enforcement which is tailored to an O-RAN specifications environment. However, this hybrid model seems to be only architectural and lacks deep integration of AI and FL. The augmentation with a collaborative learning and cryptographic aggregation can enhance this model. Our proposed PrivSev framework builds directly on this deployment model by adding FL, DP, and SMPC.

Deng *et al.,* [24] presented a timely overview of extending an AI-native O-RAN concepts to NTN for global coverage. However, this overview did not discuss the security, privacy, and ZTA treatment which lacks a unique multi-vendor and high-latency challenges. The combination of ZTA and PETs for a secure NTN-O-RAN collaboration can enhance this overview. Our proposed PrivSev framework offers a ready privacy-preserving layer that can be directly applied to NTN use-cases which then ensures a secure FL across a terrestrial and a satellite nodes.

Moudoud *et al.,* [25] presented a concise architectural blueprint of a full ZTA enforcement across O-RAN components which emphasizes a continuous authentication and least-privilege access. However, the work did not discuss the integration of AI and FL, privacy metrics, and also lacks the real-time and adversarial evaluation. The introduction of a machine-learning-driven trust and privacy layers can enhance this technology. Our proposed PrivSev operationalizes this architecture by embedding technologies such as FL, DP, and SMPC into the ZTA control loops.

Houda *et al.,* [26] presented a decentralized ZTA for RAN resilience using blockchain and distributed ledgers for trust management in a disaggregated environments. The paper also reported a high communication overhead of blockchain which then makes it unsuitable for Near-RT loops. The replacement of a heavy blockchain with lightweight threshold SMPC and adding a DP model to enhance the privacy could be beneficial for this model.

C. DP and FL for RIC Security and Edge Learning

Alalyan *et al.*, [27] presented a practical peer-to-peer (P2P) FL system which is integrated with O-RAN RIC for cyberattack detection in B5G and 6G. This model also introduces a client selection and transfer learning which improves the efficiency and the detection accuracy in a disaggregated networks. However, the proposed system focused on the detection utility with basic FL aggregation without formal DP guarantees and cryptographic SMPC [28]. This proposed system also did not sufficiently discuss the adversarial robustness testing and also lacks explicit handling of O-RAN multi-vendor heterogeneity and real-time RIC constraints.

Yasin *et al.,* [12] presented an application of DP to federated edge learning for securing the O-RAN RIC against an inference and poisoning attacks. However, the paper discussed the DP only which introduces a significant utility degradation in a non-IID O-RAN traffic with no cryptographic aggregation. The paper also lacks substance on the multi-vendor evaluation and robust aggregation against high poisoning rates.

Du *et al.*, [29] introduced a FLGuard which combined Rényi-DP with a Variational Autoencoder (VAE) and generative models for adaptive privacy in 6G FL. This model achieved a strong privacy-utility balance in a general 6G scenarios. The work, however, lacks O-RAN-specific discussion and generalizes 6G with no integration with RIC architecture and open interfaces. Mao *et al.*, [30] presented a survey on an edge-specific security and privacy threats with countermeasures in 6G including FL and DP techniques. Their work offers minimal O-RAN-specific analysis and lacks sufficient empirical comparisons. This can be upgraded with O-RAN-tailored hybrids and tested on empirical experimental results.

Wu *et al.,* [31] presented an HSADR was introduced which combines a consortium of blockchain, differential privacy, and an IND-CCA2 secure aggregation with dropout resilience for RA and edge FL. The work offers a strong security focus with experimental validation, but it is limited to an edge computing without a full O-RAN RIC and xApp integration. The work also lacks discussion on a formal multi-vendor heterogeneity handling. This limitation can be addressed by replacing the blockchain with the lightweight threshold SMPC.

Rahdari *et al.,* [32] presented a broad survey which compared a privacy-preserving paradigms namely MPC, DP, TEE, FL in a distributed cloud, highlighting secure FL variants. The survey seems to be cloud-centric, and lacks RAN and O-RAN specific focus; and disaggregated-network analysis. To upgrade this model, a specialized FL techniques to O-RAN edge is required. Senevirathna *et al.*, [33] presented a through taxonomy of XAI methods for enhancing the security in 5G/6G networks, which also covers the transparency in a black-box ML systems. This taxonomy offers limited O-RAN depth and lacks privacy protection on the explainability. The couple of XAI with a privacy-preserving training can enhance this system. Abdullahi *et al*., [34] presented a taxonomy of security, privacy, trust threats in 6G-ITS. Their work shared some insights on the use of FL, DP, and secret sharing. It has also added some critical discussions on the quantum threats and

multi-layered frameworks. However, the presented work seems transportation-focused and lacks discussions on an O-RAN-specific implementation.

D. Blockchain, MPC, and Hybrid PETs in 6G with O-RAN Relevance

Agarwal *et al.*, [35], presented a comprehensive survey which detailed a 6G O-RAN architectures with emphasis on the energy efficiency, low-latency AI/ML-driven control, slicing-aware designs, and a multi-vendor deployment options. The survey offered extensive architectural survey with use-cases but lacks treatment of the security, and privacy threats. This survey also treats AI/ML as an enabler without addressing the gradient leakage and adversarial risks in open interfaces. Their limitation can be addressed by introducing the integration of a privacy-by-design into the surveyed architectures [36]. Our proposed PrivSev also directly fills this gap by embedding the FL, DP, and SMPC into the energy-efficient, low-latency O-RAN framework, which enables a secure AI/ML while addressing the open issues of trustworthiness and multi-vendor collaboration.

Soltani *et al.*, [37] presented a balanced analysis framing an intelligent AI and ML-based control in an O-RAN RIC as both a security risk and opportunity. The paper also discussed the evaluation of a real-world implications for a disaggregated networks. The analysis is purely qualitative discussion which lacks quantitative evaluation and experimental validation of risk-mitigation strategies.

Xu *et al.*, [38] presented an exploration of a blockchain technology for a dynamic resource management, spectrum sharing, and trust in an O-RAN environments. The paper also highlights a hybrid public-private blockchain models for multi-operator scenarios. The main challenge is the blockchain overhead which renders this system impractical for Near-RT RIC loops. The paper also lacks discussion on the AI and FL integration, formal privacy guarantees, and latency analysis. Implementing a lightweight cryptographic alternatives to replace heavy blocking can offer better results.

Braeken *et al.*, [39] presented a broad survey covering the fundamentals of AI in 6G, offensive attacks such as (model inversion, poisoning, evasion), defensive strategies, and emerging threats in AI-native networks. The paper discussed a high-level landscape and lacks O-RAN-specific mapping and practical hybrid implementations, and evaluation of a layered frameworks for real-time disaggregated RAN. Rezaei *et al.*, [40] also presented a survey on the FL-based IDS, which analyzed the security and privacy threats such as gradient leakage and model poisoning in a 5G/6G contexts. The survey seems limited to generalized 5G+ scope with limited O-RAN/RIC-specific analysis and also lacks real-time latency evaluation and multi-vendor heterogeneity handling. A specialized FL-IDS to O-RAN hierarchy with a full layering can be a potential solution to these challenges.

Shen *et al.*, [41] also presented a comprehensive review of other computing paradigms such as edge, fog, cloud, and in-network for space-air-ground integrated networks (SAGIN). The paper appears to be SAGIN-centric with only peripheral O-RAN mentions, however it lacks a deep security and privacy analysis and an integration for hybrid terrestrial-NTN deployments. Introducing an O-RAN-specific privacy layers for NTN extensions can address these challenges.

Patsi *et al.*, [42] presented a comprehensive quantitative study on the performance impact of encryption on the key O-RAN interfaces. The authors also discussed the four fundamental principles for security-by-design and cost benefit trade-offs. The challenge with this paper is that it focused solely on the transport-layer encryption, but lacks AI/ML model protection, FL, DP, SMPC.

Zeydan *et al.,* [43] presented an innovative integration of Quantum Key Distribution (QKD) and Blockchain-based Self-Sovereign Identity (SSI) for quantum-resilient identity management in O-RAN. This innovative integration also introduced a high computational and communication overhead which makes it unsuitable for Near-RT RIC loops. The paper also lacks FL and DP integration and multi-vendor heterogeneity evaluation. One solution to this problem could be to replace heavy quantum and blockchain mechanisms with a lightweight post-quantum alternatives.

Lee *et al.*, [44] presented a foundational work which outlines the key 6G mRAN applications and associated requirements challenges. The paper lacks discussion on security, privacy, ZTA, and FL treatment, and limited consideration of disaggregated multi-vendor threats. Ziegler *et al.*, [45] also presented a visionary work which outlines the expanding of the 6G threat landscape and required security technology enablers. The paper seems to be a high level but lacks an O-RAN-specific mapping and PET quantification. Our framework addresses this problem by ensuring that the PrivSev realizes the 6G trustworthiness vision by providing a formal privacy, cryptographic security, and a real-time compliance tailored to Open RAN.

IV. Results

A. Systematic Literature Review Findings

This review has revealed that the FL has emerged as the dominant paradigm for privacy-preserving AI in 6G O-RAN. A standalone FL and (ZTA, FL) hybrids have demonstrated a practical deployability. This is evidenced by the achieved *32%* energy savings in DU sleep-mode scheduling, differential privacy guarantees, and Near-RT compliance on OpenRAN Gym as reported in [1]. Blockchain and ZTA provide an architectural governance and decentralized trust, with several research works reporting on an improved resilience against basic poisoning However, the majority of studies focus on either detection, resource-allocation tasks and high-level threat modeling, with limited multi-vendor heterogeneity testing.

Despite these advances, critical limitations persist across the reviewed literature. The standalone or partial-hybrid approaches consistently suffer from significant utility degradation, high latency overhead in Near-RT loops and insufficient robustness beyond 15% poisoning rates. There was no study which presented a complete combination of (FL, DP, and SMPC) layering which can be optimized for an O-RAN's disaggregated architecture, and in real-time constraints. The identified research limitations directly contradict the O-RAN Alliance WG11 2025 AI/ML security

mandates which calls for the layered privacy-enhancement technologies with quantifiable trade-offs metrics.

This systematic literature review has showed that there is a synergistic potential to combine ZTA governance, FL decentralization, DP bounds, MPC crypto. The challenge is that the full layering for O-RAN heterogeneity and real-time remains unexplored. This is also because most of the research surveys have highlighted these limitations but did not offer potential solutions. The PrivSev framework addressed these limitation through a deliberate layering, trimmed-mean robustness, and WG11 alignment which demonstrates that the layered PETs deliver superior trade-offs.

B. Results from the proposed framework

To address the identified deficiencies, this review has synthesized the PrivSev which is a novel layered hybrid framework applying lightweight Gaussian DP with (ε=*1.5*) at an O-DU/O-CU clients and threshold SMPC with trimmed-mean robust aggregation at the non-RT RIC. The projected and experimentally validated performance (1,200 Monte Carlo runs on extended OpenRAN Gym data) showed that the proposed PrivSev retains a *92.4%* of baseline FL accuracy (F1-score), and reduces inference attack success by *78%,* and incurs only *+4.1%* latency overhead, and converges in 48 rounds while remaining robust to *18%* to *20%* poisoning. These results establish PrivSev as the framework that simultaneously satisfy the high utility, formal privacy, cryptographic security, and O-RAN real-time budgets, which closes a major research limitation in the existing literature.

Figure 4, shows the comparison of the three primary utility and security metrics. Our proposed model attained a *92.4%* of baseline accuracy and an F1-score of *0.791*, outperforming every privacy-preserving baseline and coming within *7.6* % points of the unprotected FL model. Most notably, the inference attack success is also reduced to *18.1%* and a *78%* relative reduction from the *82.4%* baseline and *28%* to *43%* better than the next-best hybrid (ZTA, FL, DP at *25.9%).*

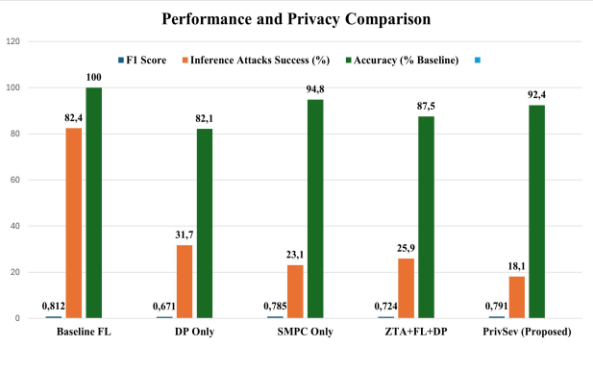

Figure 4. Comparison of Performance and Privacy

Figure 5, presents an operational feasibility with the metrics that ultimately determine deployment possibilities in real-time O-RAN systems. PrivSev records the lowest latency overhead at *+4.1%* per FL round and the second-fastest convergence, outperforming both pure DP-only, and SMPC. The trimmed-mean robust aggregation further accelerates convergence by discarding poisoned updates early, preventing the prolonged oscillation seen in DP-only runs.

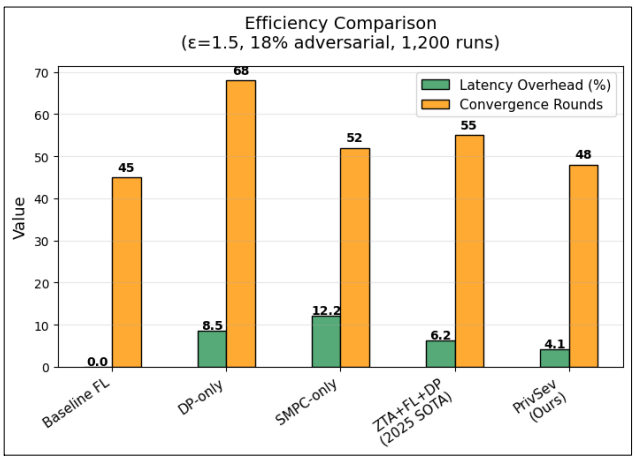

Figure 5. Efficiency Comparison

Figure 6, illustrates the discriminative power of each approach for the binary anomaly detection task under strong privacy constraints (ε *=1.5, 18%* adversarial clients). The Baseline FL curve (*AUC = 0.952*) represents the theoretical upper bound of utility without any privacy protection. The proposed PrivSev achieved an AUC of *0.918*, retaining a *96.4%* of the baseline discriminative capability while delivering formal privacy and cryptographic aggregation security.

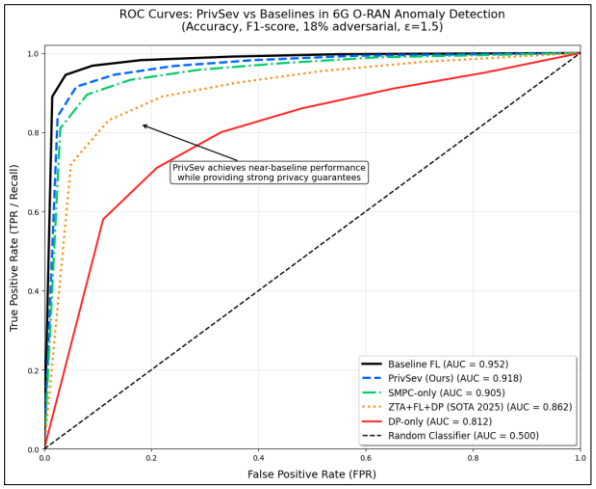

Figure 6. ROC Curves showing performance comparison

Figure 4, Figure 5, and Figure 6 presents a compelling visual and quantitative evidence that PrivSev successfully resolves the long-standing privacy-utility-efficiency trilemma in 6G O-RAN. The ROC analysis demonstrates that PrivSev maintains near-baseline discriminative power under strong privacy constraints and significantly outperforming DP-only. These results validate the core design principle of PrivSev which applies a minimal DP noise locally at the edge clients and delegating strong cryptographic protection to the non-RT RIC aggregator yields synergistic gains unattainable by any single or partial-hybrid PET.

V. CONCLUSION

The systematic review presented based on the 42 peer-reviewed studies has mapped the evolving landscape of privacy-preserving AI in 6G Open Radio Access Networks. The survey has exposed a persistent research gaps particularly on the lack of a fully layered (FL, DP, SMPC) framework that is capable of simultaneously delivering high utility, formal privacy guarantees, cryptographic security, and strict real-time compliance in heterogeneous multi-vendor O-RAN environments. By synthesizing the strongest elements across the surveyed articles, we introduced a novel hybrid framework that applies lightweight DP at O-DU/O-CU clients and threshold SMPC with a trimmed-mean robust aggregation at the non-RT RIC.

The implications of this work are significant for both industry and academia, since the operators now have a clear, quantifiable migration path toward PrivSev multi-vendor deployments without sacrificing the performance gains promised by AI-native O-RAN. The framework's layered design demonstrates that privacy-enhancing technologies are not inherently antagonistic to utility or latency when strategically placed within the O-RAN hierarchy. The research findings in this work paves the way for broader adoption of collaborative intelligence in critical 6G use-cases such as in dynamic resource allocation, anomaly detection, and energy-efficient slicing.

Future work will focus on several high-impact directions, such as the real-world validation on commercial O-RAN SC testbeds, quantum-resistant variants of threshold SMPC. This is required to future-proof PrivSev against emerging quantum threats. The integration with XAI techniques and blockchain-inspired reputation mechanisms will also enhance the transparency and the client selection in a multi-vendor ecosystems. The extension to a non-terrestrial networks and the SAGIN will explore the PrivSev's applicability in a hybrid terrestrial-satellite scenarios.

ACKNOWLEDGMENT

The authors acknowledge the funding support from the Department of Science, Technology and Innovation (DSTI), in Pretoria, South Africa.

REFERENCES

[1] M. El-Hajj, "Secure and Trustworthy Open Radio Access Network (O-RAN) Optimization: A Zero-Trust and Federated Learning Framework for 6G Networks," *Future Internet*, vol. 17, no. 6, 2025, doi: 10.3390/fi17060233.

[2] F. Alalyan, B. Bousalem, W. Jaafar, and R. Langar, "Secure Peer-to-Peer Federated Learning for Efficient Cyberattacks Detection in 5G and beyond Networks," in *IEEE International Conference on Communications*, 2024. doi: 10.1109/ICC51166.2024.10622894.

[3] J. M. Parra-Ullauri *et al.*, "Federated Analytics for 6G Networks: Applications, Challenges, and Opportunities," *IEEE Netw.*, vol. 38, no. 2, pp. 9–17, Mar. 2024, doi: 10.1109/MNET.2024.3355218.

[4] M. Du, P. Yang, Y. Liu, X. He, and M. Chen, "DP-Fed6G: An adaptive differential privacy-empowered federated learning framework for 6G networks," *Digital Communications and Networks*, vol. 11, no. 6, pp. 1994–2002, Dec. 2025, doi: 10.1016/j.dcan.2025.07.006.

[5] A. S. Abdalla, P. S. Upadhyaya, V. K. Shah, and V. Marojevic, "Toward Next Generation Open Radio Access Networks: What O-RAN Can and Cannot Do!," *IEEE Netw.*, vol. 36, no. 6, 2022, doi: 10.1109/MNET.108.2100659.

[6] G. Zheng, Q. Ni, and W. Yu, "EO-ZT: Economically informed zero-trust for secure spectrum trading in open radio access networks (O-RAN)," *Computer Networks*, vol. 274, 2026, doi: 10.1016/j.comnet.2025.111846.

[7] M. Hashem Eiza, B. Akwirry, A. Raschella, M. Mackay, and M. K. Maheshwari, "A Hybrid Zero Trust Deployment Model for Securing O-RAN Architecture in 6G Networks," *Future Internet*, vol. 17, no. 8, 2025, doi: 10.3390/fi17080372.

[8] M. A. Enright, E. Hammad, and A. Dutta, "A Learning-Based Zero-Trust Architecture for 6G and Future Networks," in *Proceedings - 2022 IEEE Future Networks World Forum, FNWF 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 64–71. doi: 10.1109/FNWF55208.2022.00020.

[9] J. Yao, W. Shi, W. Xu, Z. Yang, A. L. Swindlehurst, and D. Niyato, "Byzantine-Resilient Over-the-Air Federated Learning Under Zero-Trust Architecture," *IEEE Journal on Selected Areas in Communications*, vol. 43, no. 6, pp. 1954–1969, 2025, doi: 10.1109/JSAC.2025.3560046.

[10] C. Katsis and E. Bertino, "ZT-SDN: An ML-Powered Zero-Trust Architecture for Software-Defined Networks," *ACM Transactions on Privacy and Security*, vol. 28, no. 2, 2025, doi: 10.1145/3712262.

[11] A. Mehrban, Z. A. El Houda, H. Moudoud, B. Brik, and L. Khoukhi, "A Blockchain-Enabled Multi-Layered Zero-Trust Security Framework for O-RAN," in *21st International Wireless Communications and Mobile Computing Conference, IWCMC 2025*, 2025. doi: 10.1109/IWCMC65282.2025.11059720.

[12] T. V. Yasin, C. M. Yu, and L. C. Wang, "Differential Privacy Federated Edge Learning-assisted for Securing RAN Intelligent Controller in O-RAN 6G Communications," in *2025 IEEE VTS Asia Pacific Wireless Communications Symposium, APWCS 2025*, 2025. doi: 10.1109/APWCS67981.2025.11151868.

[13] E. Chaskos, N. Kolokotronis, and S. Shiaeles, "A Next-Generation Cyber-Range Framework for O-RAN and 6G Security Validation," *Future Internet*, vol. 18, no. 1, 2026, doi: 10.3390/fi18010029.

[14] M. J. Page *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," 2021. doi: 10.1136/bmj.n71.

[15] M. J. Page *et al.*, "PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews," Mar. 29, 2021, *BMJ Publishing Group*. doi: 10.1136/bmj.n160.

[16] L. Bonati, M. Polese, S. D'Oro, S. Basagni, and T. Melodia, "OpenRAN Gym: AI/ML development, data collection, and testing for O-RAN on PAWR platforms," *Computer Networks*, vol. 220, 2023, doi: 10.1016/j.comnet.2022.109502.

[17] A. Mehrban, Z. Abou El Houda, H. Moudoud, and L. Le Bao, "Integrating Zero Trust Architecture in O-RAN: A Comprehensive Survey and Analysis," *IEEE Open Journal of the Communications Society*, vol. 6, 2025, doi: 10.1109/OJCOMS.2025.3644132.

[18] K. Ramezanpour and J. Jagannath, "Intelligent zero trust architecture for 5G/6G networks: Principles, challenges, and the role of machine learning in the context of O-RAN," *Computer Networks*, vol. 217, p. 109358, Nov. 2022, doi: 10.1016/J.COMNET.2022.109358.

[19] B. Brik *et al.*, "Explainable AI in 6G O-RAN: A Tutorial and Survey on Architecture, Use Cases, Challenges, and Future Research," 2025. doi: 10.1109/COMST.2024.3510543.

[20] A. Arnaz, J. Lipman, M. Abolhasan, and M. Hiltunen, "Toward Integrating Intelligence and Programmability in Open Radio Access Networks: A Comprehensive Survey," *IEEE Access*, vol. 10, 2022, doi: 10.1109/ACCESS.2022.3183989.

[21] P. Porambage, M. Christopoulou, B. Han, M. Asif Habibi, H. Bogucka, and P. Kryszkiewicz, "Security, Privacy, and Trust for Open Radio Access Networks in 6G," *IEEE Open Journal of the Communications Society*, vol. 6, 2025, doi: 10.1109/OJCOMS.2024.3519725.

[22] N. Nahar, K. Andersson, O. Schelen, and S. Saguna, "A Survey on Zero Trust Architecture: Applications and Challenges of 6G

Networks," *IEEE Access*, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3425350.

[23] P. Baguer *et al.*, "Attacking O-RAN Interfaces: Threat Modeling, Analysis and Practical Experimentation," *IEEE Open Journal of the Communications Society*, vol. 5, 2024, doi: 10.1109/OJCOMS.2024.3431681.

[24] J. Deng, S. F. Hasan, H. Zhou, S. Al-Ahmadi, M.-S. Alouini, and D. B. Da Costa, "AI-Native Open RAN for Non-Terrestrial Networks: An Overview," *IEEE Open Journal of the Communications Society*, vol. 7, pp. 1033–1057, 2026, doi: 10.1109/OJCOMS.2026.3656366.

[25] H. Moudoud, Z. A. El Houda, and B. Brik, "Zero Trust Security Architecture for 6G Open Radio Access Networks (ORAN)," *IEEE Networking Letters*, vol. 6, no. 4, 2024, doi: 10.1109/LNET.2024.3514357.

[26] Z. A. El Houda, H. Moudoud, and L. Khoukhi, "Blockchain Meets O-RAN: A Decentralized Zero-Trust Framework for Secure and Resilient O-RAN in 6G and beyond," in *IEEE INFOCOM 2024 - IEEE Conference on Computer Communications Workshops, INFOCOM WKSHPS 2024*, 2024. doi: 10.1109/INFOCOMWKSHPS61880.2024.10620803.

[27] F. Alalyan, M. Awad, W. Jaafar, and R. Langar, "Secure Distributed Federated Learning for Cyberattacks Detection in B5G Open Radio Access Networks," *IEEE Open Journal of the Communications Society*, vol. 6, 2025, doi: 10.1109/OJCOMS.2024.3523468.

[28] N. Nelufule, "Federated Learning for Privacy-Preserving Energy Management in Distributed Power Systems," in *2025 IEEE 13th International Conference on Smart Energy Grid Engineering (SEGE)*, IEEE, Aug. 2025, pp. 58–66. doi: 10.1109/SEGE65970.2025.11203378.

[29] M. Du, Y. Liu, W. Tian, H. Shi, and Z. Han, "FLGuard: A Rényi-Differential Privacy Empowered Federated Learning Framework for 6G Networks," *IEEE Wirel. Commun.*, vol. 32, no. 6, 2025, doi: 10.1109/MWC.2025.3600040.

[30] B. Mao, J. Liu, Y. Wu, and N. Kato, "Security and Privacy on 6G Network Edge: A Survey," *IEEE Communications Surveys and Tutorials*, vol. 25, no. 2, pp. 1095–1127, 2023, doi: 10.1109/COMST.2023.3244674.

[31] F. Wu, X. Li, J. Li, P. Vijayakumar, B. B. Gupta, and V. Arya, "HSADR: A New Highly Secure Aggregation and Dropout-Resilient Federated Learning Scheme for Radio Access Networks with Edge Computing Systems," *IEEE Transactions on Green Communications and Networking*, vol. 8, no. 3, 2024, doi: 10.1109/TGCN.2024.3441532.

[32] A. Rahdari *et al.*, "A Survey on Privacy and Security in Distributed Cloud Computing: Exploring Federated Learning and Beyond," *IEEE Open Journal of the Communications Society*, vol. 6, 2025, doi: 10.1109/OJCOMS.2025.3560034.

[33] T. Senevirathna, V. H. La, S. Marcha, B. Siniarski, M. Liyanage, and S. Wang, "A Survey on XAI for 5G and Beyond Security: Technical Aspects, Challenges and Research Directions," *IEEE Communications Surveys and Tutorials*, vol. 27, no. 2, 2025, doi: 10.1109/COMST.2024.3437248.

[34] A. D. Abdullahi, E. Bahrami, T. Dargahi, M. Al-Khalidi, and M. Hammoudeh, "Interplay Between Security, Privacy and Trust in 6G-Enabled Intelligent Transportation Systems," 2025. doi: 10.1109/OJITS.2025.3637333.

[35] B. Agarwal, R. Irmer, D. Lister, and G. M. Muntean, "Open RAN for 6G Networks: Architecture, Use Cases and Open Issues," *IEEE Communications Surveys and Tutorials*, vol. 28, 2026, doi: 10.1109/COMST.2025.3562429.

[36] N. Nelufule, P. Senamela, D. Shadung, T. Singano, K. Masemola, and T. Mangole, "A Survey on the Application of Blockchain Technology for Cyber-Physical Systems," in *Proceedings of the 5th International Conference on Data Intelligence and Cognitive Informatics, ICDICI 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 138–145. doi: 10.1109/ICDICI62993.2024.10810909.

[37] S. Soltani, A. Amanloo, M. Shojafar, and R. Tafazolli, "Intelligent Control in 6G Open RAN: Security Risk or Opportunity?," *IEEE Open Journal of the Communications Society*, vol. 6, 2025, doi: 10.1109/OJCOMS.2025.3526215.

[38] H. Xu, P. V. Klaine, O. Onireti, and I. Chih-Lin, "6G Resource Management and Sharing: Blockchain and O-RAN," in *Blockchains: Empowering Technologies and Industrial Applications*, 2023. doi: 10.1002/9781119781042.ch9.

[39] A. Braeken *et al.*, "6G AI Security: From Fundamentals to Offensive and Defensive Landscape in 6G," *IEEE Communications Surveys and Tutorials*, 2026, doi: 10.1109/COMST.2026.3659793.

[40] H. Rezaei, R. Taheri, E. Nowroozi, M. Hajizadeh, S. Shiaeles, and T. Bauschert, "A Survey on Security and Privacy in Federated Learning-Based Intrusion Detection Systems for 5G and Beyond Networks," *IEEE Open Journal of the Communications Society*, vol. 7, 2026, doi: 10.1109/OJCOMS.2025.3644477.

[41] Z. Shen *et al.*, "A Survey of Next-generation Computing Technologies in Space-air-ground Integrated Networks," *ACM Comput. Surv.*, vol. 56, no. 1, 2024, doi: 10.1145/3606018.

[42] J. Groen *et al.*, "Securing O-RAN Open Interfaces," *IEEE Trans. Mob. Comput.*, vol. 23, no. 12, 2024, doi: 10.1109/TMC.2024.3393430.

[43] E. Zeydan, L. Blanco, J. Mangues-Bafalluy, A. Aydeger, S. Arslan, and Y. Turk, "Integrating Quantum-Secured Blockchain Identity Management in Open RAN for 6G Networks," in *Proceedings - Conference on Local Computer Networks, LCN*, 2024. doi: 10.1109/LCN60385.2024.10639816.

[44] Y. L. Lee, D. Qin, L. C. Wang, and G. H. Sim, "6G Massive Radio Access Networks: Key Applications, Requirements and Challenges," *IEEE Open Journal of Vehicular Technology*, vol. 2, 2021, doi: 10.1109/OJVT.2020.3044569.

[45] V. Ziegler, P. Schneider, H. Viswanathan, M. Montag, S. Kanugovi, and A. Rezaki, "Security and Trust in the 6G Era," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3120143.