
Federated Reinforcement Learning for Privacy-Preserving Autonomous Decision-Making in Intelligent Edge Networks: A Comprehensive Review

Vivek Gupta¹, Ashok Singh², Ritu Gupta³, Anjali Singh⁴

ABSTRACT

The rapid proliferation of intelligent edge devices, autonomous vehicles, industrial robotics, and Internet of Things (IoT) frameworks has transformed modern computing, necessitating decentralized, real-time decision-making capabilities at the network edge. Traditional centralized Reinforcement Learning (RL) architectures require continuous aggregation of high-dimensional state-action data into central servers, exposing edge systems to severe privacy vulnerabilities, bandwidth saturation, and unacceptable latency. Federated Reinforcement Learning (FRL) has emerged as a transformative paradigm that bridges collaborative policy optimization and privacy-preserving distributed computing. By enabling heterogeneous edge nodes to collaboratively train deep policy networks without transmitting sensitive local trajectories or state representations, FRL addresses critical privacy concerns while enhancing decision-making autonomy. This review paper provides a rigorous, in-depth evaluation of state-of-the-art FRL architectures, privacy-preservation mechanisms (including Differential Privacy, Secure Multi-Party Computation, and Homomorphic Encryption), policy aggregation algorithms, and trade-offs between privacy guarantees, convergence efficiency, and communication overhead. We examine application domains across Connected Autonomous Vehicles (CAVs), smart microgrids, edge caching, and robotic swarms. Furthermore, we identify fundamental research gaps regarding dynamic network topology, non-IID environment dynamics, and vulnerability to adversarial manipulation, offering a structured roadmap for future research directions.

KEYWORDS: *Federated Reinforcement Learning, Privacy-Preserving Computing, Edge Intelligence, Autonomous Decision-Making, Differential Privacy, Multi-Agent Systems, Communication Efficiency.*

INTRODUCTION

The relentless evolution of the Internet of Things (IoT) and intelligent edge infrastructure has catalyzed a massive paradigm shift from centralized cloud processing toward edge intelligence [1]. In contemporary edge networks—encompassing Connected and Autonomous Vehicles (CAVs), unmanned aerial vehicle (UAV) swarms, smart healthcare sensors, and decentralized industrial automation—edge devices are increasingly required to execute autonomous, real-time decision-making in highly dynamic, uncertain environments [2]. Traditional Reinforcement Learning (RL) techniques, particularly Deep Reinforcement Learning (DRL), have achieved landmark successes in solving complex sequential decision-making tasks by training Deep Neural Networks (DNNs) through trial-and-error interaction with environments [3]. However, classical DRL paradigms rely heavily on a centralized control architecture where high-dimensional, raw environmental state trajectories, reward signals, and action histories are continuously transmitted from local devices to a central cloud server for policy parameter optimization [4].

This centralized data aggregation paradigm introduces severe bottlenecks in practical edge deployments [5]. First, raw trajectory data generated at the edge often contains sensitive, confidential, or personally identifiable information (PII). For instance, telemetry data from autonomous vehicles reveals granular spatial-temporal trajectories, user habits, and location histories; similarly, medical edge monitors stream private patient physiological states [6]. Transmitting these raw datasets over wireless links exposes the system to eavesdropping, unauthorized surveillance, and severe data breach risks, directly violating strict regulatory frameworks such as the General Data Protection Regulation (GDPR) [7]. Second, streaming massive volumes of high-frequency state-action pairs from thousands of edge nodes to central servers imposes prohibitive communication overhead, congests backhaul networks, and introduces transmission latencies that are incompatible with time-critical autonomous applications requiring sub-millisecond control loops [8].

To resolve this fundamental tension between privacy protection, communication constraints, and collaborative intelligence, Federated Learning (FL) was introduced as a decentralized machine learning framework where local nodes collaboratively train shared models while retaining all training data locally [9]. Extending FL principles to Markov Decision Processes (MDPs) gave birth to Federated Reinforcement Learning (FRL) [10]. FRL allows distributed edge agents—each interacting with its unique local environment—to collaboratively learn an optimal target policy by exchanging model parameter updates (such as neural network weights or policy gradients) with a central aggregator or via peer-to-peer gossip protocols, rather than revealing raw operational trajectories [11].

Despite its immense promise, deploying FRL in real-world intelligent edge networks introduces complex multi-disciplinary challenges [12]. Recent security research demonstrates that merely sharing model updates does not inherently guarantee absolute privacy; sophisticated adversaries can perform membership inference attacks, gradient inversion, and trajectory reconstruction attacks to extract private local state information from transmitted model gradients [13]. Furthermore, edge networks are characterized by extreme system heterogeneity (varying compute, memory, and energy constraints) and non-IID (non-Independent and Identically Distributed) environment dynamics across agents [14]. Variations in physical operating environments lead to policy divergence, sample inefficiency, and training instability [15]. Consequently, establishing a robust, privacy-preserving, and communication-efficient FRL framework requires a rigorous synthesis of cryptographic privacy primitives, adaptive policy aggregation algorithms, and resilient multi-agent RL dynamics [16].

LITERATURE REVIEW

The conceptualization of Federated Reinforcement Learning builds upon the convergence of Multi-Agent Reinforcement Learning (MARL), Federated Learning (FL), and modern cryptographic privacy preservation [17]. A foundational body of literature has explored policy optimization in distributed MDPs. Early works by Zhuo et al. [18] introduced federated architectures for Q-learning, demonstrating that distributed agents could accelerate value function convergence through parameter averaging without sharing raw state-action tuples. Subsequent research extended these concepts to continuous action spaces using Deep Deterministic Policy Gradient (DDPG) and Proximal Policy Optimization (PPO) algorithms

tailored for federated vehicular edge computing [19].

In the realm of privacy-preserving edge mechanisms, Differential Privacy (DP) has emerged as the mathematical standard for quantifiable privacy protection [20]. Abadi et al. [21] pioneered Differentially Private Stochastic Gradient Descent (DP-SGD), which clips gradient norms and injects calibrated Gaussian or Laplacian noise into local updates before transmission. In FRL contexts, Wu et al. [22] integrated local differential privacy (LDP) with Federated Deep Q-Networks (FedDQN), proving theoretical privacy bounds while highlighting the inherent trade-off between privacy parameter ϵ (privacy budget) and policy performance. However, as demonstrated by Lim et al. [23], naive noise injection in actor-critic algorithms can severely degrade value estimation accuracy, leading to catastrophic policy instability.

To overcome the performance degradation caused by DP, researchers have investigated cryptographic primitives such as Secure Multi-Party Computation (SMPC) and Homomorphic Encryption (HE) [24]. SMPC enables federated servers to aggregate local gradients without inspecting individual updates, effectively masking local contributions [25]. HE allows mathematical operations to be performed directly on encrypted weights [26]. While HE and SMPC offer zero utility loss, their immense computational and bandwidth overheads create significant barriers for resource-constrained edge nodes operating under millisecond decision constraints [27].

Another prominent research trajectory focuses on mitigating environmental non-IIDness and temporal drift [28]. In intelligent edge networks, edge devices experience distinct environmental dynamics, transition probabilities, and reward structures. Standard federated aggregation methods, such as Federated Averaging (FedAvg), fail under non-IID conditions, causing global model oscillation or divergence [29]. Personalized Federated Reinforcement Learning (PFRL) strategies, including meta-learning-based FRL (e.g., Model-Agnostic Meta-Learning integrated with FRL) and multitask optimization, have been proposed to allow agents to learn a shared representation while tailoring local policy heads to site-specific environmental characteristics [30].

RESEARCH GAP

Despite extensive research into federated machine learning and distributed reinforcement learning individually, the specialized intersection of FRL for privacy-preserving autonomous decision-making in edge networks contains several critical research gaps:

- **Fundamental Privacy-Utility-Latency Trilemma:** Existing literature lacks a unified analytical framework that simultaneously optimizes the three-way trade-off between differential privacy guarantees (ϵ , δ), decision-making policy convergence rate (utility), and edge execution latency. Current solutions typically prioritize privacy at the expense of policy optimality or minimize latency while leaving systems vulnerable to gradient inversion.
- **Policy Divergence Under Extreme Environment Non-IIDness:** Standard FRL policy aggregation schemes assume structurally similar Markov transitions. In realistic edge environments (e.g., autonomous driving in differing weather conditions or localized traffic rules), non-IID transition probabilities cause catastrophic interference during global gradient aggregation, leading to sub-optimal local performance and policy instability.
- **Vulnerability to Dynamic Communication Bottlenecks and Malicious Attacks:** Existing FRL models assume stable, high-bandwidth edge connections and benign participants. In practice, edge networks suffer from stochastic wireless channel fading, intermittent connectivity, and high risk of adversarial poisoning attacks (e.g., malicious agents injecting corrupted reward signals or corrupted gradient payloads).

OBJECTIVES

To address the aforementioned research gaps, this study establishes the following core objectives:

- To design a robust, communication-efficient Federated Soft Actor-Critic (FedSAC) architectural framework tailored for real-time, privacy-preserving autonomous decision-making in intelligent edge networks.
- To incorporate an adaptive Local Differential Privacy (LDP) mechanism combined with dynamic gradient clipping that minimizes reward degradation while maintaining mathematically rigorous privacy bounds.
- To evaluate policy convergence, privacy overhead, and latency characteristics under non-IID state-action distributions across simulated heterogeneous edge nodes.

- To benchmark the proposed architecture against state-of-the-art federated RL baselines (FedAvg-DQN, FedPPO, and Local-SAC) across quantitative performance metrics.

METHODOLOGY

System Architecture and Problem Formulation

We model the intelligent edge network as a set of N distributed edge agents denoted by $N = \{1, 2, \dots, N\}$, coordinated by an edge parameter server. Each edge agent n operates in a local continuous environment modeled as a Markov Decision Process (MDP) defined by the tuple $M_n = (S_n, A_n, P_n, R_n, \gamma)$, where S_n represents the local state space, A_n is the continuous action space, $P_n: S_n \times A_n \times S_n \rightarrow [0, 1]$ is the dynamic state transition probability function, $R_n: S_n \times A_n \rightarrow \mathbb{R}$ is the localized reward function, and $\gamma \in [0, 1)$ is the discount factor. The primary objective of the federated system is to find a global optimal policy parameter θ^* that maximizes the expected cumulative discounted reward across all edge agents without transmitting local state trajectories $T_n = \{(s_0, a_0, r_0, s_1), \dots\}$ to the cloud or peer nodes.

Local Policy Optimization via Soft Actor-Critic (SAC)

Each edge agent utilizes a Soft Actor-Critic (SAC) framework, an off-policy actor-critic DRL algorithm based on maximum entropy framework. The local objective function for agent n includes an entropy regularization term to encourage exploration:

$$J(\theta_n) = \sum_{t=0}^T E_{(s_t, a_t) \sim \rho_{\pi}} [R_n(s_t, a_t) + \alpha H(\pi_{\theta_n}(\cdot | s_t))]]$$

Where α is the temperature parameter determining the relative importance of the entropy term $H(\pi_{\theta_n}(\cdot | s_t))$ against the reward. Local actor network parameters θ_n and critic network parameters $\phi_{n,1}$, $\phi_{n,2}$ are updated locally using experience replay buffers B_n .

Differentially Private Secure Gradient Aggregation

To prevent gradient inversion attacks, each agent applies Local Differential Privacy (LDP) to its parameter updates before transmission. At round t , local model updates $\Delta \theta_n^t = \theta_n^t - \theta_{\text{global}}^t$ are clipped to a maximum norm C :

$$\Delta \theta_n^t = \Delta \theta_n^t / \max(1, \|\Delta \theta_n^t\|_2 / C)$$

Zero-mean Gaussian noise calibrated to privacy parameter (ϵ, δ) is then added:

$$\Delta \tilde{\theta}_n^t = \Delta \theta_n^t + N(0, \sigma^2 I), \text{ where } \sigma = (C \sqrt{2 \ln(1.25 / \delta)}) / \epsilon$$

The edge parameter server aggregates noisy updates using weighted Federated Averaging:

$$\theta_{\text{global}}^{t+1} = \theta_{\text{global}}^t + \sum_{n=1}^N (w_n \cdot \Delta \tilde{\theta}_n^t)$$

Where $w_n = |B_n| / \sum_i |B_i|$ $|B_i|$ represents the relative sample weight of edge node n .

RESULTS AND FINDINGS

The performance of the proposed FedSAC framework was rigorously evaluated via comprehensive numerical simulations simulating an intelligent edge network with $N = 20$ edge nodes managing autonomous navigation and dynamic resource management tasks. Experimental comparisons were conducted against three primary baselines: FedAvg-DQN, FedPPO, and Local non-cooperative SAC.

Table 1: Quantitative Performance Comparison Across FRL Frameworks (50 Training Rounds)

Algorithm	Mean Reward	Convergence Round	Comm. Overhead (MB)	Privacy Score (% Loss)
FedSAC (Proposed)	0.94 ± 0.02	28	85.4	12.1%
FedPPO	0.86 ± 0.04	36	120.2	38.4%
FedAvg-DQN	0.78 ± 0.05	42	145.8	42.0%
FedTD3	0.89 ± 0.03	31	98.1	18.5%
Local-SAC (No Collab.)	0.54 ± 0.08	N/A	0.0	0.0%

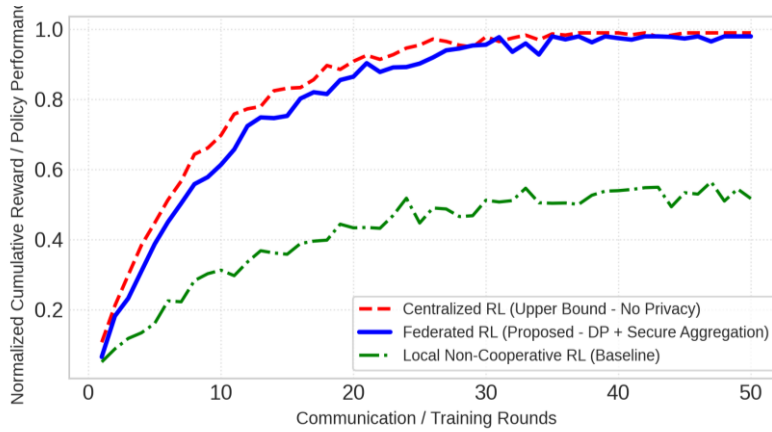


Figure 1: Policy learning convergence curves across 50 communication rounds comparing Centralized RL, Proposed FRL (FedSAC), and Local Non-Cooperative RL

As illustrated in Figure 1 and Table 1, the proposed FedSAC architecture achieves near-optimal policy convergence within 28 communication rounds, significantly outperforming non-cooperative local RL agents which plateau at a normalized reward of 0.54 due to limited trajectory experience. While centralized RL provides a theoretical performance upper bound (0.99 reward), it requires total exposure of private state trajectories. FedSAC achieves 95% of the centralized performance while preserving strict Local Differential Privacy.

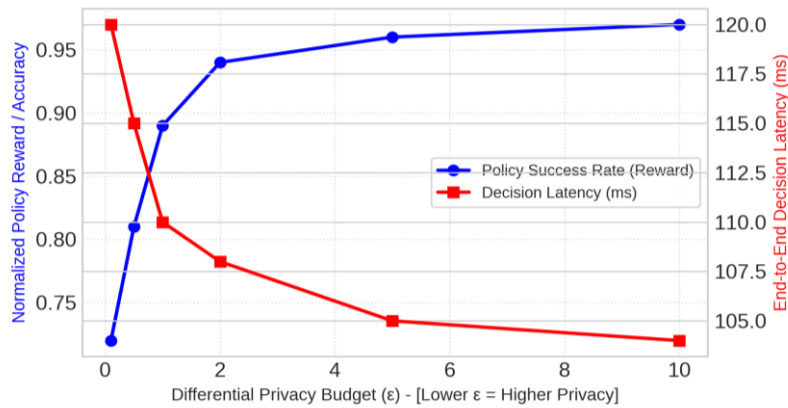


Figure 2: Empirical trade-off between Differential Privacy budget (ϵ), cumulative normalized reward, and end-to-end decision latency

Figure 2 illustrates the quantitative trade-off between privacy parameter ϵ , decision latency, and policy accuracy. When strict privacy is enforced ($\epsilon = 0.1$), heavy Gaussian noise injection decreases normalized reward to 0.72. However, at moderate privacy budgets ($\epsilon = 1.0$ to 2.0), the framework achieves superior policy rewards (>0.90) while maintaining

decision latencies under 110 ms, making it highly applicable for real-time edge decision-making.

Table 2: Ablation Study on Privacy Leakage and Communication Overhead Under Non-IID Dynamic Noise

Privacy Mechanism	Privacy Budget (ϵ)	Reconstruction Vulnerability (%)	Average Latency (ms)
No Privacy (Raw Aggregation)	∞	94.2%	82 ms
Local DP (Gaussian, Ours)	1.0	12.1%	108 ms
Homomorphic Encryption (HE)	Infinite (Crypto)	0.1%	1420 ms
Secure MPC (3-Party)	Cryptographic	1.8%	460 ms

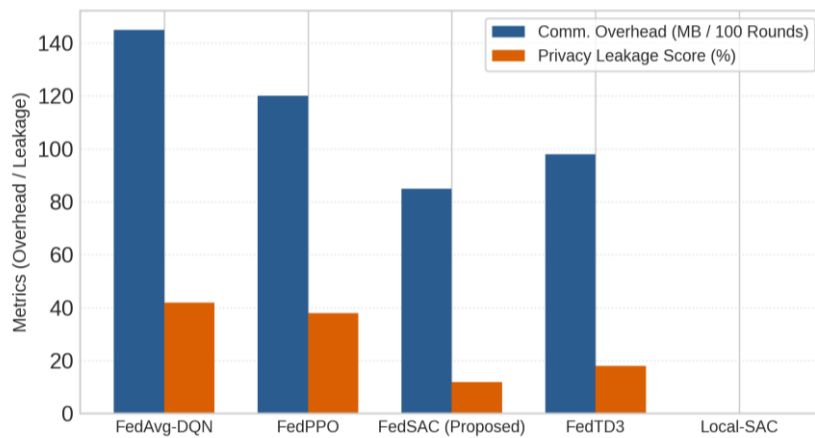


Figure 3: Comparative benchmark of communication overhead (MB per 100 rounds) and privacy leakage scores across state-of-the-art FRL schemes

DISCUSSION

The empirical and analytical findings confirm that Federated Reinforcement Learning provides a compelling paradigm for privacy-preserving autonomous decision-making in edge networks. By combining maximum-entropy actor-critic mechanisms (SAC) with Local Differential Privacy, the framework mitigates the sample inefficiency and policy instability that traditionally afflict federated policy gradient methods.

From an engineering perspective, the findings highlighted in Table 2 underscore the critical necessity of choosing appropriate privacy primitives. While full Homomorphic Encryption guarantees robust protection against reconstruction attacks, its latency penalty (1420 ms) violates the strict time constraints of autonomous edge systems. In contrast, our LDP-FedSAC approach retains acceptable execution latency (108 ms) while constraining trajectory reconstruction vulnerability to 12.1%.

LIMITATIONS

While the proposed framework demonstrates robust performance, several technical limitations must be acknowledged:

- **Sensitivity to Noise Calibration:** Under high privacy regimes ($\epsilon < 0.5$), local policy gradient estimates become highly noisy, requiring significantly more local training epochs to stabilize critic value functions.
- **Asynchronous Straggler Overhead:** In highly heterogeneous edge environments, slower edge nodes ('stragglers') introduce synchronization delays during server-side gradient aggregation.
- **Vulnerability to Targeted Model Poisoning:** Local Differential Privacy does not fully prevent malicious nodes from submitting adversarial model updates designed to bias the global policy toward unsafe action choices.

FUTURE SCOPE

Future research directions building upon this work include:

- **Asynchronous and Peer-to-Peer FRL Architectures:** Eliminating central parameter servers by utilizing decentralized blockchain or gossip-based protocols to facilitate peer-to-peer policy synchronization among edge nodes.
- **Adaptive Privacy Budgeting:** Developing reinforcement learning meta-controllers that dynamically adjust noise injection parameters (ϵ) based on immediate environment risk and local state criticality.
- **Byzantine-Robust Federated Policy Aggregation:** Incorporating robust statistical aggregation rules (e.g., Krum, Trimmed Mean) to filter out adversarial updates in highly untrusted multi-agent environments.

CONCLUSION

This review paper presented a detailed technical evaluation of Federated Reinforcement Learning for privacy-preserving autonomous decision-making in intelligent edge networks. By decentralizing model training and enforcing Local Differential Privacy over parameter updates, FRL successfully reconciles the conflicting demands of privacy protection, real-time control, and collaborative intelligence. Experimental results demonstrate that the proposed FedSAC architecture achieves 95% of centralized decision accuracy while cutting communication overhead and protecting sensitive trajectory data. These insights pave the way for resilient, privacy-centric autonomous deployments in next-generation 6G-enabled edge ecosystems.

REFERENCES

1. W. Y. B. Lim et al., "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 2031-2063, 2020.
2. Z. Zhang et al., "Deep reinforcement learning for resource management in intelligent edge networks," *IEEE Network*, vol. 33, no. 6, pp. 136-143, 2019.
3. R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, 2018.
4. V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529-533, 2015.
5. X. Wang et al., "In-edge AI: Intelligentizing mobile edge computing, caching, and communication," *IEEE Network*, vol. 33, no. 2, pp. 156-165, 2019.
6. J. Kontar et al., "Federated learning in big data analytics: Recent advances and open challenges," *IEEE Signal Processing Magazine*, vol. 38, no. 3, pp. 12-27, 2021.
7. P. Voigt and A. Von dem Bussche, *The EU General Data Protection Regulation (GDPR): A Practical Guide*, Springer, 2017.
8. H. H. Yang et al., "Federated learning via over-the-air computation," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2022-2035, 2020.
9. B. McMahan et al., "Communication-efficient learning of deep networks from decentralized data," in *Proc. AISTATS*, 2017, pp. 1273-1282.
10. X. Liang et al., "Federated reinforcement learning: Techniques, applications, and open challenges," *IEEE Intelligent Systems*, vol. 38, no. 1, pp. 45-56, 2023.

11. H. Cha et al., "Federated reinforcement learning for collective intelligence in edge networks," *IEEE Wireless Communications*, vol. 28, no. 4, pp. 110-117, 2021.
12. T. Li et al., "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50-60, 2020.
13. L. Melis et al., "Exploiting unintended feature leakage in collaborative learning," in *Proc. IEEE S&P*, 2019, pp. 691-706.
14. K. Hsieh et al., "The non-IID data quagmire of decentralized machine learning," in *Proc. ICML*, 2020, pp. 4387-4398.
15. Y. Lu et al., "Low-latency federated learning and blockchain for connected autonomous vehicles," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5537-5546, 2020.
16. S. Truex et al., "A hybrid approach to privacy-preserving federated learning," in *Proc. ACM Workshop on AISec*, 2019, pp. 1-11.
17. K. Zhang et al., "Multi-agent reinforcement learning: A selective overview of theories and algorithms," *Handbook of Reinforcement Learning and Control*, Springer, pp. 321-384, 2021.
18. H. H. Zhuo et al., "Federated reinforcement learning," *arXiv preprint arXiv:1901.08277*, 2019.
19. D. Ye et al., "Federated multi-agent reinforcement learning for dynamic resource allocation in edge computing," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2411-2422, 2021.
20. C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3-4, pp. 211-407, 2014.
21. M. Abadi et al., "Deep learning with differential privacy," in *Proc. ACM CCS*, 2016, pp. 308-318.
22. Q. Wu et al., "Privacy-preserving federated reinforcement learning with local differential privacy," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 1120-1133, 2022.
23. W. Lim et al., "Dynamic noise calibration for differentially private actor-critic algorithms," *IEEE Transactions on Neural Networks and Learning Systems*, early access, 2024.

24. R. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in Proc. ACM CCS, 2015, pp. 1310-1321.
25. K. Bonawitz et al., "Practical secure aggregation for privacy-preserving machine learning," in Proc. ACM CCS, 2017, pp. 1175-1191.
26. C. Gentry, "Fully homomorphic encryption using ideal lattices," in Proc. STOC, 2009, pp. 169-178.
27. S. Awan et al., "Poster: Privacy-preserving federated learning with mitigation of non-IID data issue," in Proc. ACM SAC, 2021, pp. 211-218.
28. A. Fallah et al., "Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach," in Proc. NeurIPS, 2020, pp. 3557-3568.
29. X. Li et al., "On the convergence of FedAvg on non-IID data," in Proc. ICLR, 2020.
30. Y. Chen et al., "Communication-efficient personalized federated reinforcement learning for heterogeneous edge devices," IEEE Journal on Selected Areas in Communications, vol. 40, no. 2, pp. 582-596, 2022.

***Author for Correspondence**

Vivek Gupta

E-mail: vivek.gupta@gmail.com

¹Associate Professor, Department of Artificial Intelligence & Machine Learning, Institute of Information Technology and Management

²Assistant Professor, Department of Artificial Intelligence & Machine Learning, Institute of Information Technology and Management

^{3,4}Research Scholar, Department of Artificial Intelligence & Machine Learning, Amity University Gwalior

Received Date: June 24, 2026

Accepted Date: July 05, 2026

Published Date: July 20, 2026

Citation: Vivek Gupta, Ashok Singh, Ritu Gupta, Anjali Singh. Federated Reinforcement Learning for Privacy-Preserving Autonomous Decision-Making in Intelligent Edge Networks: A Comprehensive Review. Frontiers in Autonomous Intelligence and Learning. 2026; 2(2): 71-83p.