Recent Advances in Federated Learning and Privacy-Preserving Artificial Intelligence: A Systematic Literature Review

[1]Priya Kumawat, [2]Dr. Vishal Shrivastava

[1]M.TECH. SCHOLAR, [2]PROFESSOR
DEPARTEMNT OF CSE, ARYA COLLEGE OF ENGINEERING AND IT, JAIPUR, INDIA

‘

Abstract

The proliferation of data-driven artificial intelligence (AI) is increasingly hindered by concerns regarding data privacy, regulatory compliance (e.g., GDPR), and data silos. Federated Learning (FL) has emerged as a transformative paradigm allowing decentralized model training while preserving data locality. This paper provides a systematic literature review of the state-of-the-art in FL and privacy-preserving AI. We categorize recent advances into communication efficiency, Byzantine robustness, and privacy-enhancing technologies (PETs) like Differential Privacy (DP), Homomorphic Encryption (HE), and Secure Multi-Party Computation (SMPC). We identify critical research gaps and outline future directions, including the integration of FL with large foundation models and regulatory-aware machine learning.

1. Introduction

Modern machine learning (ML) thrives on vast amounts of centralized data. However, the centralization of sensitive data—personal, medical, or financial—poses significant privacy risks and violates data sovereignty laws. Federated Learning (FL) shifts the paradigm by bringing the computation to the data rather than the data to the computation. Initially proposed by Google, FL has evolved beyond simple stochastic gradient descent (SGD) to a complex field encompassing security, privacy, and system optimization.
This paper reviews the advancements in FL and privacy-preserving AI, analyzing how these technologies address the "trilemma" of privacy, accuracy, and communication efficiency.

2. Theoretical Foundations

2.1 The Federated Learning Lifecycle

The standard FL lifecycle consists of four phases:
Selection: The server selects a subset of clients.
Broadcast: The server sends the global model to selected clients.
Local Computation: Clients train the model on local private data.
Aggregation: The server collects updates and computes a global policy (e.g., FedAvg).

2.2 Privacy-Preserving Mechanisms

Privacy-Preserving AI (PPAI) utilizes cryptographic and statistical techniques to prevent the server or malicious actors from reconstructing training data via gradient leakage. Key techniques include:
Differential Privacy (DP): Introducing calibrated noise to gradients to provide mathematical guarantees against inference attacks.
Homomorphic Encryption (HE): Performing arithmetic operations on encrypted data without decryption.

Secure Multi-Party Computation (SMPC): Ensuring that no single party can see individual updates, only the final summation.

3. Communication Efficiency

Communication overhead is the primary bottleneck in FL. When models scale to millions of parameters, the energy and bandwidth costs of transmitting updates become prohibitive.

3.1 Gradient Compression and Sparsification

Recent advances, such as *FedPAQ* [10], utilize periodic averaging and bit-quantization to reduce communication. Studies suggest that only 1-5% of gradients are necessary to maintain convergence, provided error compensation mechanisms are applied [12].

3.2 Knowledge Distillation (KD)

To overcome heterogeneity, Federated Knowledge Distillation (FedKD) allows clients to share model outputs (logits) rather than heavy model parameters, significantly reducing the payload size [15].

4. Privacy and Security Challenges

FL is not inherently privacy-preserving. Gradient inversion attacks—where an adversary reconstructs raw data from gradients—have been demonstrated even in deep neural networks [18].

4.1 Differential Privacy (DP)

DP remains the industry standard for user-level privacy. However, the "privacy-utility trade-off" remains a challenge. Recent literature proposes Adaptive DP, where the noise scale is dynamically adjusted based on the training iteration, mitigating degradation of model accuracy [22].

4.2 Robustness to Poisoning

Malicious clients can perform "model poisoning" (submitting malicious updates). Byzantine-robust aggregation algorithms, such as *Krum* and *Median-based filtering*, have been integrated into FL frameworks to detect and exclude outliers [25].

5. Emerging Trends: FL and Large Language Models (LLMs)

The current frontier is Federated Tuning of Large Language Models (FedLLM). Fine-tuning massive models locally is computationally expensive. Research into Parameter-Efficient Fine-Tuning (PEFT) methods, such as LoRA (Low-Rank Adaptation), has enabled the federated adaptation of LLMs on edge devices without updating the full model weight matrix [28].

6. Discussion and Future Directions

Despite the progress, the following challenges remain:

Statistical Heterogeneity (Non-IID data): Traditional FL fails when data distributions vary significantly across devices.

System Heterogeneity: Diverse hardware capabilities (e.g., IoT vs. Smartphones) lead to the "straggler problem."

Regulatory Compliance: Auditing federated systems to satisfy GDPR "right to be forgotten" is currently an active, largely unsolved, research area.

7. Conclusion

FL has matured from a conceptual framework into a robust field. The integration of PETs with efficient communication strategies represents the next phase of secure AI. As research pivots toward LLMs and decentralized edge intelligence, the synergy between performance and privacy will remain the critical determinant of success.

References

[1] McMahan, B., et al. (2017). "Communication-efficient learning of deep networks from decentralized data." *AISTATS*.

[2] Yang, Q., et al. (2019). "Federated Machine Learning: Concept and Applications." *ACM TIST*.

[3] Kairouz, P., et al. (2021). "Advances and Open Problems in Federated Learning." *Foundations and Trends in ML*.

[4] Bonawitz, K., et al. (2019). "Towards secure and private federated learning." *IEEE*.

[5] Li, T., et al. (2020). "Federated Optimization in Heterogeneous Networks." *MLSys*.

[6] Dwork, C. (2006). "Differential Privacy." *ICALP*.

[7] Gentry, C. (2009). "Fully homomorphic encryption using ideal lattices." *STOC*.

[8] Wang, J., et al. (2020). "Federated learning with matched averaging." *ICLR*.

[9] Zhao, Y., et al. (2018). "Federated learning with non-IID data." *ICLR*.

[10] Reisizadeh, A., et al. (2020). "FedPAQ: A communication-efficient federated learning method." *AISTATS*.

[11] Aono, Y., et al. (2017). "Privacy-preserving deep learning via homomorphic encryption." *IEEE WIFS*.

[12] Lin, Y., et al. (2018). "Deep gradient compression." *ICLR*.

[13] Abadi, M., et al. (2016). "Deep learning with differential privacy." *CCS*.

[14] Truex, S., et al. (2019). "Hybrid differentially private federated learning." *ACM CCS*.

[15] Li, D., & Wang, J. (2021). "FedKD: Federated knowledge distillation." *ACM Multimedia*.

[16] Choudhury, O., et al. (2019). "Answering queries on healthcare data using differential privacy." *JAMIA*.

[17] Fung, C., et al. (2020). "Mitigating sybil attacks in federated learning." *arXiv*.

[18] Zhu, L., et al. (2019). "Deep leakage from gradients." *NeurIPS*.

[19] Geiping, J., et al. (2020). "Inverting gradients: How easy is it to break privacy?" *NeurIPS*.

[20] Hynes, N., et al. (2017). "Efficient deep learning on multi-source data." *NeurIPS*.

[21] Konečný, J., et al. (2016). "Federated learning: Strategies for improving communication." *arXiv*.

[22] Wei, K., et al. (2020). "User-level privacy-preserving federated learning." *IEEE TDSC*.

[23] Geyer, R. C., et al. (2017). "Differentially private federated learning: A client-level perspective." *arXiv*.

[24] Bhagoji, A. N., et al. (2019). "Analyzing federated learning through the lens of evasion attacks." *arXiv*.

[25] Blanchard, P., et al. (2017). "Machine learning with adversaries: Byzantine-tolerant gradient descent." *NeurIPS*.

[26] Mironov, I. (2017). "Rényi differential privacy." *IEEE CSF*.

[27] Li, X., et al. (2021). "Federated Learning with Non-IID Data: A Survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[28] Hu, E. J., et al. (2021). "LoRA: Low-Rank Adaptation of Large Language Models." *ICLR*.

[29] Hard, A., et al. (2018). "Federated learning for mobile keyboard prediction." *arXiv*.

[30] Nguyen, T. D., et al. (2022). "Federated learning for privacy-preserving AI: A review." *IEEE Access*.