Journal of Harbin Engineering University
ISSN: 1006-7043
Vol 47 No. 07
July 2026

A Comprehensive Review of Federated Learning: Privacy, Multimodality, and Robustness in Heterogeneous Environments

Adarsh R Menon[1], Aditya Venkatesh[2], Aarav Adarsh[3], A Haveesh Kumar[4], Chetana Srinivas[5]

[1, 2, 3, 4, 5] Department of Computer Science and Engineering (AIML) PES University, Bengaluru, Karnataka, India

Abstract

Federated Learning (FL) enables collaborative model training while preserving data privacy by keeping raw data locally on client devices. Despite its advantages, practical FL deployments face major challenges due to statistical heterogeneity, multimodal data distributions, missing modalities, and the trade–off between privacy, utility and efficiency. Existing methods typically address these challenges in isolation, resulting in performance degradation, excessive overhead, or weakened privacy guarantees. This paper presents a comprehensive review of recent advances in federated learning through a pillar-based analysis. A technical taxonomy is introduced to categorize aggregation strategies, privacy mechanisms, and modality handling approaches. Key research gaps are identified, highlighting the absence of unified frameworks that simultaneously address heterogeneity, multimodality, and strong privacy. Finally, future research directions toward practical and trustworthy federated learning systems are discussed.

Keywords: Federated learning, privacy preservation, multi-modal learning, non-IID data, homomorphic encryption, differential privacy.

1. Introduction

Construction industry work Federated Learning (FL) has emerged as a decentralized machine learning paradigm that enables multiple clients to collaboratively train a shared model without transferring raw data to a central server [1], [2]. This approach is particularly attractive for privacy-sensitive domains such as healthcare, finance, and social media analytics [3]. However, real-world FL deployments are characterized by extreme data heterogeneity, multimodal inputs, and partial data availability, which significantly degrade model performance [4], [5].

2. Pillar-Based Literature Review

A. Federated Learning under Data Heterogeneity

Data heterogeneity is a fundamental challenge in FL, where client data distributions differ significantly in size, class balance, and statistical properties [4]. Several studies propose heterogeneity-aware aggregation techniques and client clustering strategies to mitigate model drift [11]. While these approaches improve convergence, they often assume unimodal data and lack integrated privacy mechanisms.

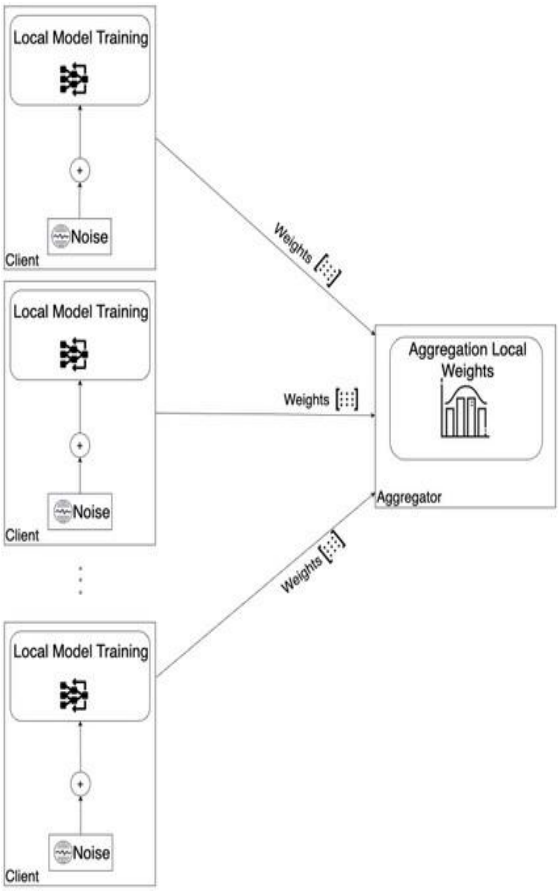

Fig. 1. Differential Privacy-based Federated Learning workflow.

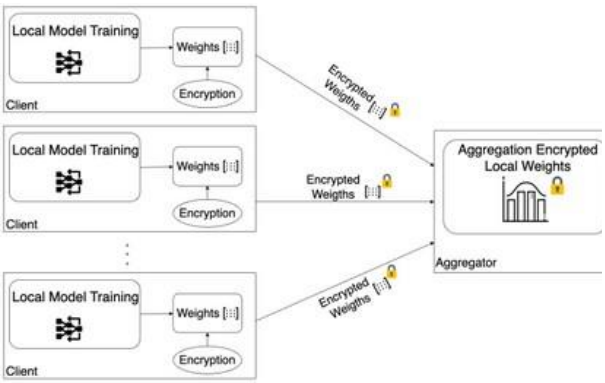

Fig. 2. Homomorphic Encryption-based Federated Learning workflow.

B. Privacy-Preserving Federated Learning

Figure 1 illustrates a Differential Privacy-based Federated Learning workflow, where calibrated noise is injected into local model updates to protect sensitive training data [7].

Figure 2 depicts a Homomorphic Encryption-based Federated Learning workflow, in which model updates are aggregated in encrypted form without exposing plaintext gradients to the server [7].

Privacy-preserving FL approaches primarily rely on Differential Privacy and Homomorphic Encryption [12], [13]. DP provides statistical privacy guarantees by adding calibrated

noise to model updates, whereas HE enables secure aggregation over encrypted parameters [14]. Hybrid DP+HE methods attempt to balance privacy and efficiency, but often incur substantial computational and communication overhead [6], [8], [15].

C. Multimodal Federated Learning

Multimodal FL extends traditional FL to settings where clients possess heterogeneous data modalities such as text, images, and sensor signals [16]. Recent approaches employ contrastive learning, prototype alignment, and ensemble strategies to address missing modalities [5]. However, these methods rarely incorporate strong privacy mechanisms due to increased complexity.

D. Trust, Security, and Robustness

Beyond privacy, FL systems must be robust against adversarial attacks such as poisoning and inference attacks [10]. Robust aggregation rules and anomaly detection techniques have been proposed, yet most security-focused works assume unimodal data and overlook multimodal and heterogeneous environments.

E. Formalized Threat Model and Adversarial Behavior

The robustness of a Federated Learning (FL) framework is defined by its ability to withstand specific adversarial behaviors [10].

1) ***Honest-but-Curious Aggregators:*** Honest-but-curious (semi-honest) aggregators strictly follow the prescribed FL protocol but attempt to infer sensitive information from inter-mediate model updates through passive statistical or inference-based analysis [6]. Although they do not actively manipulate the training process, their access to gradient information poses potential privacy risks.

2) ***Malicious Clients and Collusion:*** Active adversaries may engage in model poisoning by transmitting carefully crafted gradients designed to corrupt the global model [10]. In addition, malicious clients may conduct Membership Inference Attacks (MIA) to determine whether a particular individual's data was included in the training set. Collusion among multiple adversarial clients can further amplify these threats.

3) ***Byzantine Fault Tolerance:*** To ensure resilience in open and decentralized FL environments, systems must account for Byzantine participants who may upload arbitrary or conflicting updates [10]. This necessitates robust aggregation rules that extend beyond simple averaging and incorporate mechanisms to detect and mitigate anomalous behavior.

3. Technical Analysis and Taxonomy

Federated learning (FL) systems can be systematically analyzed through a technical taxonomy that captures the key design dimensions influencing performance, privacy, and deployability [2], [9]. Based on the surveyed literature, existing FL approaches are categorized along four fundamental dimensions: aggregation strategy, privacy mechanism, modality handling, and system efficiency [17]. These dimensions are largely orthogonal and collectively determine the robustness of an FL framework under realistic conditions.

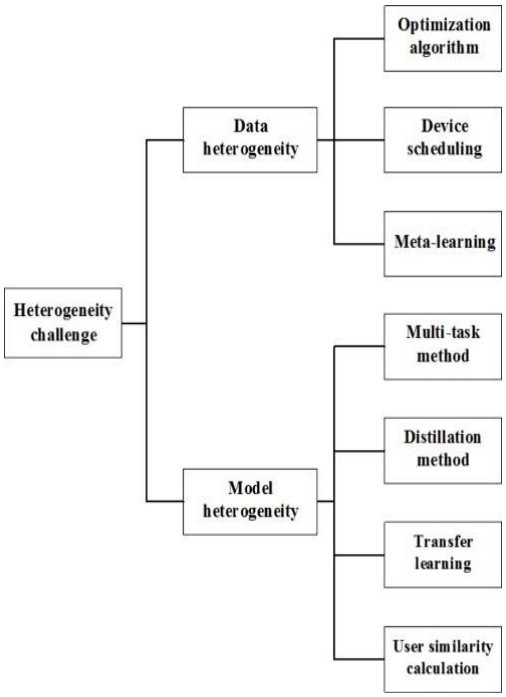

Fig. 3. Taxonomy of heterogeneity challenges in federated learning.

A. Heterogeneity Handling

Beyond statistical non-IID data, real-world federated systems must also contend with deeper forms of heterogeneity that influence model performance and convergence [4]. Figure 3 summarizes a high-level taxonomy of heterogeneity challenges in federated learning, adapted from [9].

1) *Statistical vs. Structural Heterogeneity:* Statistical heterogeneity refers to differences in data distributions across clients, whereas structural heterogeneity arises when clients employ different model architectures or possess varying data modalities [4]. Both forms require distinct mitigation strategies.

2) *Quality-Aware Client Selection:* Aggregation logic can be enhanced by incorporating client reliability indicators that weight updates based on local data quality and historical contribution to global convergence [11].

3) *Feature Alignment in Multimodality:* In multimodal settings, modality-aware weighting ensures that clients with missing data types do not negatively bias the global representation, which is particularly important for complex tasks such as misinformation detection [5], [16].

B. Aggregation Strategy

Aggregation defines how a central server combines model updates received from distributed clients [1]. Let there be K clients, where client k produces a local model update wk after training on its local dataset of size nk.

1) *Standard Aggregation (FedAvg):* The most commonly used aggregation method is Federated Averaging (FedAvg), defined as:

$$w^{(t+1)} = \sum_{k=1}^{K} \frac{n_k}{\sum_{j=1}^{K} n_j} w_k^{(t)}$$

(1)

FedAvg assumes that client updates are unbiased and drawn from similar data distributions [1]. While effective under IID settings, this assumption breaks down in the presence of statistical heterogeneity, leading to client drift and unstable convergence [4].

DSGD vs DFedAvgM: In traditional decentralized stochastic gradient descent (DSGD), each client communicates with its neighbors after one single training step. In contrast, de-centralized federated averaging with momentum (DFedAvgM) allows clients to perform multiple local training iterations before communication, thereby reducing communication over-head while maintaining convergence [18].

2) *Heterogeneity-Aware Aggregation:* To mitigate non-IID effects, heterogeneity-aware methods introduce adaptive aggregation weights [11], [18]:

$$w^{(t+1)} = \sum_{k=1}^{K} \alpha_k w_k^{(t)}, \quad \sum_{k=1}^{K} \alpha_k = 1$$

(2)

where *αk* may depend on distribution similarity, gradient divergence, or clustering-based statistics. These approaches improve convergence stability but are often designed for unimodal data and lack integrated privacy awareness.

3) *Drift and Modality-Aware Aggregation:* In multimodal settings, aggregation may further incorporate modality-specific weights [5], [16]:

$$w^{(t+1)} = \sum_{k=1}^{K} \sum_{m \in M_k} \beta_{k,m}, w_{k,m}^{t+1}$$

(3)

where M_k denotes the set of modalities available at client k. Such strategies address modality imbalance but significantly increase system complexity.

C. Privacy Mechanism

Privacy mechanisms aim to prevent leakage of sensitive information from model updates exchanged during training [12], [13].

1) *Differential Privacy (DP):* A mechanism M satisfies (ε, δ)-differential privacy if:

$$\Pr[M(D) \in S] \leq e^{\varepsilon} \Pr[M(D') \in S] + \delta$$

(4)

where datasets D and D' differ by a single data point [12]. In FL, DP is commonly implemented by perturbing local updates:

$$\tilde{\mathbf{w}}_k = \mathbf{w}_k + N(0, \sigma^2 I)$$

(5)

Although DP provides formal privacy guarantees, excessive noise injection can degrade model accuracy, especially under non-IID and multimodal conditions [19].

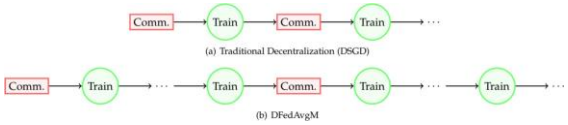

Fig. 4. Comparison of communication and training patterns in traditional DSGD and DFedAvgM.

2) *Homomorphic Encryption (HE):* Homomorphic Encryption enables computation directly on encrypted updates [13]. Given encryption $E(\cdot)$ and decryption $D(\cdot)$:

$$D\ E(\mathbf{w}_1) \oplus E(\mathbf{w}_2) = \mathbf{w}_1 + \mathbf{w}_2$$

(6)

HE preserves exact model correctness but introduces substantial computational and communication overhead, limiting scalability [14].

3) *Hybrid DP–HE Approaches:* Hybrid methods combine DP and HE to balance privacy and efficiency [6], [8], [15]:

$$\tilde{\mathbf{w}}_k = E(\mathbf{w}_k + N(0, \sigma^2 I))$$

(7)

However, these approaches are typically applied as independent security layers and are not jointly optimized with aggregation under heterogeneity.

4) *Advanced Privacy–Utility Optimization:* Standalone Differential Privacy (DP) often suffers from an availability cost, where excessive noise injection can severely degrade model utility [8]. By integrating DP with Homomorphic Encryption (HE), noise can be applied more selectively, since the HE layer prevents the server from observing even the noisy gradients in

A major challenge in HE-based FL is that computational noise within ciphertexts grows with the complexity of operations [13]. To mitigate this, selective encryption strategies have been proposed in which only the final aggregation step is encrypted rather than the entire training pipeline [8]. This optimization significantly improves system efficiency while maintaining strong privacy guarantees. An example of such differential privacy with homomorphic encryption to selectively protect local updates while maintaining efficiency, as illustrated in Fig. 5 [8].

D. Modality Handling

This dimension categorizes FL methods based on the type and completeness of data modalities across clients [5], [16].

1) *Unimodal Federated Learning:* Each client possesses a single modality, and the optimization objective is:

$$\min_{\boldsymbol{w}} \sum_{k=1}^{K} \mathbb{E}_{(x_k, y_k) \sim D_k} [\ell(f(x_k; \boldsymbol{w}), y_k)]$$

(8)

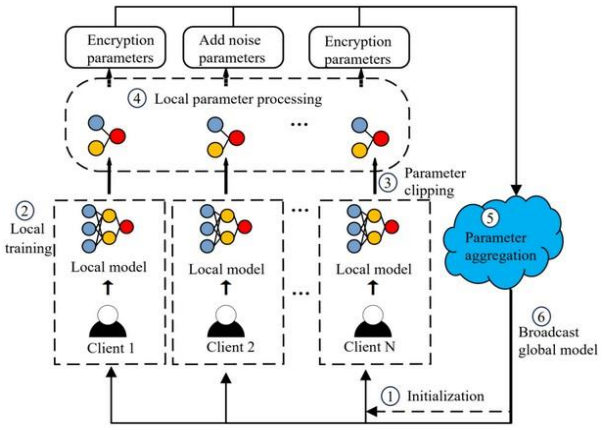

Fig. 5. Architecture of ADPHE-FL combining adaptive differential privacy and homomorphic encryption.

2) *Multimodal Federated Learning:* Clients possess multiple modalities $x_k = \{x_k^{(1)}, \dots, x_k^{(M)}\}$, yielding:

$$\min_{\boldsymbol{w}} \sum_{k=1}^{K} \sum_{m=1}^{M} \mathbb{E}\left[\ell\left(f_m\left(x_k^{(m)}; \boldsymbol{w}\right), y_k\right)\right] \quad (9)$$

3) *Missing-Modality-Aware Learning:* To handle missing modalities, contrastive objectives are frequently employed [5], [16]:

$$\mathcal{L}_{\text{con}} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_k \exp(\text{sim}(z_i, z_k)/\tau)} \quad (10)$$

where z_i and z_j represent modality-specific embeddings. While effective, such techniques complicate privacy-preserving training.

E. System Efficiency

System efficiency determines the feasibility of deploying FL methods at scale [17].

1) *Communication Efficiency Techniques:* Communication bottlenecks can be mitigated through gradient sparsification and quantization [1], [17]. Sparsification transmits only the most significant gradients, while quantization reduces numerical precision before encryption and transmission.

Another strategy is intensive local processing, where clients perform more local SGD iterations before communicating with the server. This reduces the total number of global rounds, provided the aggregation mechanism can handle the resulting model drift [18].

2) *Communication Cost:* The communication cost per round is approximately:

$$C = K \times |\mathbf{w}| \quad (11)$$

where $|\mathbf{w}|$ is the model size. HE significantly increases $|\mathbf{w}|$ due to ciphertext expansion [14].

3) *Computational Overhead:* Local training complexity can be approximated as:

$$O(E \cdot n_k \cdot d) \quad (12)$$

where E is the number of local epochs and d is the model dimension. Cryptographic operations further increase computational cost [8].

4) *Scalability:* Scalability is constrained by client count, model size, and privacy budget [17]. Existing privacy-preserving multimodal FL systems often fail to scale beyond small client populations.

F. Taxonomy Insight

This taxonomy highlights a critical structural limitation in current federated learning research: most approaches optimize only a subset of defined dimensions [9]. Consequently, existing systems remain fragile when faced with the combined challenges of heterogeneity, multimodality, strong privacy requirements, and practical efficiency.

4. Research Gaps

Despite extensive research, several critical gaps remain. Current FL frameworks fail to simultaneously address Non-IID data, different data types, and missing information [4], [5]. Model drift increases and accuracy deteriorates. Introducing a Heterogeneity-Aware Aggregation algorithm directly addresses model drift. Instead of just averaging everything like FedAvg, this method uses adaptive weighted aggregation that accounts for the characteristics of each client's local data [11], [18]. Additionally, no practical solution achieves strong, provable privacy guarantees without severely sacrificing model accuracy or system

efficiency [7], [8]. Increasing privacy budget in Differential privacy sacrifices model accuracy whereas Homomorphic encryption carries massive computational overhead [13], [14]. Existing hybrid privacy approaches are often unoptimized and do not directly address aggregation bottlenecks due to heterogeneity, causing the aforementioned privacy-utility-efficiency trilemma [6], [8].

5. Practical Applications of Hybrid Federated Learning

Federated Learning is particularly transformative in domains governed by strict privacy regulations such as GDPR and HIPAA, and in environments characterized by multimodal and non-IID data [3].

A. Digital Healthcare and Mental Health Analytics

1) *Privacy-Preserving Diagnosis:* Healthcare institutions can collaboratively train diagnostic models for sensitive conditions such as depression or anxiety without sharing patient-identifiable records [3], [20].

2) *Multimodal Integration:* Hybrid FL frameworks enable the fusion of clinical text notes and medical imaging, supporting more holistic patient assessments while maintaining privacy [3].

B. Social Media Misinformation Tracking

1) *Cross-Platform Detection: Federated systems allow multiple social platforms to collaboratively detect misinformation without centralizing proprietary user data [16].*

2) *Content Veracity Analysis: Using datasets such as Fakeddit, models can analyze both visual and textual content in a decentralized, privacy-first manner [5], [16].*

C. Industrial Internet of Things (IIoT) and Edge Computing

1) *Predictive Maintenance:* In smart manufacturing, sensors across *different* machines generate non-IID data *due* to varying operational conditions. Heterogeneity-aware aggregation ensures robust global performance [18].

1) *Secure Edge Intelligence:* Hybrid DP–HE protocols protect sensitive industrial parameters during frequent edge-to-cloud model updates [8].

D. Finance and Anti-Money Laundering (AML)

1) *Secure Multi-Institution Collaboration:* Financial institutions can collaboratively train fraud detection models while keeping transaction patterns encrypted during secure aggregation [14], [15].

6. Future Research Directions and Open Challenges

A. Adaptive Privacy Allocation

Future systems should adopt "Privacy-on-Demand" architectures where the privacy budget ε is dynamically adjusted based on data sensitivity, modality type, and client reliability [8].

B. Communication-Efficient Cryptography

Ciphertext expansion in Homomorphic Encryption remains a bottleneck [13]. Future research should integrate gradient sparsification and quantization directly into encrypted computations [17].

C. Resilience Against Advanced Poisoning Attacks

With DP and HE defenses maturing, adversaries are shifting toward backdoor attacks that subtly corrupt global models [10]. Developing heterogeneity-aware aggregation rules to distinguish legitimate data skew from malicious poisoning is critical.

D. Unified Multimodal Benchmarking

There is a lack of standardized benchmarks that simultaneously test non-IID data, missing modalities, and privacy leakage [5], [16]. Datasets such as Fakeddit and SMHD should be used to establish common evaluation protocols.

E. Post-Quantum Privacy

As quantum computing advances, many current HE schemes may become vulnerable [13]. Exploring lattice-based crypto-graphic primitives resistant to quantum attacks is essential for long-term privacy.

7. Discussion

The analysis presented in this review highlights that the core challenges of federated learning (statistical heterogeneity, multimodal data imbalance, missing modalities, and strong privacy requirements) are deeply interconnected [4], [9]. Addressing these challenges in isolation has resulted in fragmented solutions that fail under realistic deployment

conditions. Future federated learning systems must therefore move toward integrated designs that jointly optimize aggregation robustness, privacy guarantees, and system efficiency.

One promising research direction is the development of *heterogeneity-aware federated learning frameworks* that explicitly adapt aggregation and privacy mechanisms to client-level data characteristics [11], [18]. Instead of relying on static aggregation weights or uniform privacy budgets, future systems can dynamically adjust aggregation coefficients based on distributional similarity, gradient divergence, or modality availability across clients.

References

[1] H. B. McMahan et al., "Communication Efficient Learning of Deep Networks from Decentralized Data," in *Proc. AISTATS*, Fort Lauderdale, FL, USA, Apr. 2017, pp. 1273–1282.

[2] T. Li et al., "Federated Learning: Challenges, Methods, and Future Directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, 2020.

[3] S. S. Khalil, N. S. Tawfik, and M. Spruit, "Federated learning for privacy-preserving depression detection with multilingual language mod-els in social media posts," *Patterns*, vol. 5, no. 7, p. 100990, Jul. 2024.

[4] S. Vahidian et al., "Rethinking Data Heterogeneity in Federated Learn-ing," *IEEE Trans. Artif. Intell.*, vol. 5, no. 3, pp. 1386–1397, Mar. 2024.

[5] G. Bao et al., "Multimodal Federated Learning with Missing Modality via Prototype Mask and Contrast," arXiv:2312.13508, 2023.

[6] S. Truex et al., "A Hybrid Approach to Privacy-Preserving Federated Learning," in *Proc. ACM AISec*, London, UK, Nov. 2019, pp. 19–30.

[7] A. Catalfamo, M. Fazio, A. Celesti, and M. Villari, "Privacy-Preserving in Federated Learning: A Comparison between Differential Privacy and Homomorphic Encryption across Different Scenarios," in *Proc. IEEE ICSTW*, Naples, Italy, May 2025, pp. 451–459.

[8] T. Wu, Y. Deng, Q. Zhou, X. Chen, and M. Zhang, "ADPHE-FL: Federated learning method based on adaptive differential privacy and homomorphic encryption," *Peer-to-Peer Netw. Appl.*, vol. 18, no. 1, p. 141, Apr. 2025.

[9] J. Wen et al., "A survey on federated learning: challenges and applications," *Int. J. Mach. Learn. Cybern.*, vol. 14, pp. 513–535, Nov. 2022.

[10] Y. Zhang et al., "A Survey of Trustworthy Federated Learning with Perspectives on Security, Robustness, and Privacy," arXiv:2302.10637, 2023.

[11] M. S. Islam et al., "FedClust: Optimizing Federated Learning on Non-IID Data through Weight-Driven Client Clustering," in *Proc. IPDPSW*, May 2024.

[12] 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.

[13] A. Acar et al., "A Survey on Homomorphic Encryption Schemes: Theory and Implementation," *ACM Comput. Surv.*, vol. 51, no. 4, pp. 79:1–79:35, Jul. 2018.

[14] F. Castro, D. Impedovo, and G. Pirlo, "An Efficient and Privacy-Preserving Federated Learning Approach Based on Homomorphic En-cryption," *IEEE Open J. Comput. Soc.*, vol. 6, pp. 336–346, 2025.

[15] P. Correia, I. Silva, I. Amorim, E. Maia, and I. Prac¸a, "Feder-ated Learning: An approach with Hybrid Homomorphic Encryption," arXiv:2509.03427, 2025.

[16] Q. Yu et al., "Multimodal Federated Learning via Contrastive Represen-tation Ensemble," in *Proc. ICLR*, Kigali, Rwanda, May 2023.

[17] C. W. Ling et al., "Communication Efficient and Privacy-Adaptable Mechanism for Federated Learning," arXiv:2501.12046, 2025.

[18] T. Sun, D. Li, and B. Wang, "Decentralized Federated Averaging," *IEEE TPAMI*, vol. 45, no. 4, pp. 4289–4301, Apr. 2023.

[19] X. Xiao et al., "A Differential Privacy-Based Mechanism for Preventing Data Leakage in Large Language Model Training," *Acad. J. Sociol. Manage.*, vol. 3, no. 2, pp. 33–45, 2025.

[20] S. Kumar, “Ensuring Data Privacy in Teen and Children Mental Health Analytics,” *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 10, no. 6, pp. 1334–1341, Dec. 2024.

[21] J. Bai et al., “Federated Fine-tuning of Large Language Models under Heterogeneous Tasks and Client Resources,” arXiv:2402.11505, 2024.