

Federated Learning for Privacy-Preserving Data Analytics

Dr. Thilo Streichert

Professor, University of Mannheim, Germany

Abstract

The growing dependence of contemporary organizations on data-intensive analytics has increased the tension between analytical value and information privacy. Conventional centralized machine-learning architectures typically require data generated by individuals, organizations, devices, or institutions to be transferred into a common computational environment. Such consolidation can simplify model development but may also increase exposure to unauthorized access, data breaches, secondary use, institutional restrictions, and regulatory concerns. Federated learning offers an alternative computational architecture in which participating clients collaboratively train a shared model while retaining their original datasets within local environments.

This study examines the potential of federated learning as a foundation for privacy-preserving data analytics and evaluates the additional contribution of secure aggregation and differential privacy. Because verified multi-institutional datasets were not available, the research adopts a transparent simulation-based comparative methodology involving four analytical architectures: centralized analytics, baseline federated learning, federated learning with secure aggregation, and federated learning combining differential privacy with secure aggregation. The simulated evaluation considers predictive accuracy, raw-data movement, relative privacy-exposure risk, communication burden, and analytical utility. Results indicate that federated architectures can substantially reduce direct raw-data exposure while maintaining competitive analytical performance.

In the illustrative scenario, the relative privacy-exposure risk index decreased from 100 under centralized analytics to 46 under baseline federated learning, 24 after the introduction of secure aggregation, and 14 when differential privacy and secure aggregation were jointly applied. Model accuracy declined moderately as stronger privacy controls were introduced, demonstrating an important privacy–utility trade-off. The study concludes that federated learning should not be treated as a complete privacy solution in isolation. Effective privacy-preserving analytics requires appropriate aggregation protocols, privacy mechanisms, security governance, client authentication, model monitoring, and careful management of heterogeneous data. Federated learning is therefore best understood as a decentralized analytical architecture that can reduce unnecessary data concentration while supporting collaborative intelligence across organizational and device boundaries.

Keywords: Federated learning; privacy-preserving analytics; differential privacy; secure aggregation; decentralized machine learning; data governance; distributed analytics.

1. Introduction

Data analytics has become a central component of contemporary decision-making across healthcare, financial services, industrial systems, telecommunications, public administration, digital platforms, and connected devices. Machine-learning models often derive their effectiveness from access to large and diverse datasets. Conventional analytical practice frequently achieves this scale by transferring data from multiple sources into centralized repositories where preprocessing, model training, validation, and inference are performed. Although centralization can simplify computational management, it creates significant privacy and governance concerns when the underlying data contain confidential, commercially sensitive, personally identifiable, or institutionally restricted information.

These concerns are particularly pronounced when organizations possess analytically valuable data but cannot freely transfer them to a common repository. Hospitals may hold patient information under confidentiality obligations, financial institutions may possess sensitive transaction records, manufacturers may be unwilling to expose proprietary operational information, and mobile devices may contain behavioral data that users reasonably expect to remain local. The resulting fragmentation has often been described as a problem of isolated data environments. Federated machine learning emerged partly as a response to this limitation by enabling collaborative model development without requiring participating clients to directly pool their underlying training datasets. Yang and colleagues subsequently developed a broader conceptual framework covering horizontal federated learning, vertical federated learning, and federated transfer learning.

The practical foundation of modern federated learning was strongly shaped by McMahan and colleagues, who proposed training a shared model by repeatedly sending a global model to participating devices, performing local optimization, and aggregating locally computed updates rather than collecting the original records. Their Federated Averaging approach demonstrated that decentralized training could work with unbalanced and non-independent-and-identically-distributed client data while substantially reducing communication rounds relative to synchronized stochastic gradient descent. This model established a major distinction between centralized machine learning and federated learning: computation moves toward distributed data instead of requiring all data to move toward centralized computation.

However, localization of raw records does not automatically establish complete privacy. Information can potentially be inferred from gradients, model parameters, participation patterns, or repeated updates. Research on federated learning therefore increasingly incorporates complementary mechanisms such as secure multiparty computation, secure aggregation, encryption, and differential privacy. Bonawitz and colleagues developed a practical secure-aggregation protocol that enables a coordinating server to obtain an aggregate of client contributions while preventing exposure of individual client updates under defined threat assumptions. Differential privacy provides a different form of protection by controlling the extent to which the participation or information of an individual contributor can influence a released analytical result. Research by Abadi and colleagues demonstrated the feasibility of training deep-learning models under differential-privacy constraints, although privacy protection can involve measurable costs in model utility and computational efficiency.

This study examines federated learning from the broader perspective of privacy-preserving data analytics rather than considering decentralization alone as sufficient evidence of privacy. The paper evaluates how baseline federated learning, secure aggregation, and differential privacy influence

analytical utility and simulated privacy exposure. Since authenticated cross-institutional datasets were not supplied for the study, the analytical component is explicitly simulation-based. Its purpose is to illustrate defensible comparative relationships rather than to claim empirical performance from real organizations. The study therefore contributes a structured framework for understanding the interaction among decentralization, privacy controls, predictive performance, communication requirements, and governance in collaborative data analytics.

2. Review of Literature

2.1 Decentralized Learning and the Emergence of Federated Analytics

Traditional distributed machine learning commonly assumes that datasets can be collected, partitioned, and distributed under the control of a centralized computational environment. Federated learning differs because the participating data sources may remain administratively independent, geographically dispersed, technically heterogeneous, and unwilling or unable to transfer their original records. McMahan and colleagues formalized an influential approach in which local devices train a shared model through iterative model averaging while preserving the localization of training data. Their experiments also emphasized two characteristics that later became central research concerns: communication efficiency and non-IID data distributions.

Federated learning consequently represents more than a conventional distributed-computing technique. It changes the location of computation and restructures the relationship between data ownership and collaborative modeling. Instead of establishing a single centralized training dataset, multiple clients contribute computationally derived information to a common learning objective. This architecture can support mobile-device learning, collaborative healthcare analytics, interorganizational fraud detection, industrial intelligence, financial modeling, and other applications in which unrestricted data pooling may be undesirable.

Yang and colleagues extended the conceptual scope by distinguishing among horizontal federated learning, vertical federated learning, and federated transfer learning. Horizontal federated learning generally addresses participants that possess similar feature spaces but different records. Vertical federated learning is relevant when organizations possess different features describing overlapping entities, while federated transfer learning considers situations with limited overlap in both sample and feature spaces. These variations demonstrate that federated analytics is not a single uniform training architecture but a family of collaborative learning configurations.

2.2 Privacy Protection Beyond Data Localization

Keeping original data on a client device or institutional server can substantially reduce direct raw-data movement, but localization alone does not eliminate all privacy threats. Federated-learning systems exchange information derived from local training, and these exchanges may become potential attack surfaces. Kairouz and colleagues identify privacy, robustness, security, communication efficiency, fairness, heterogeneity, and system design among the major unresolved issues in federated learning. Their work emphasizes that FL can reduce systemic privacy risks associated with centralized data collection but does not inherently remove every privacy vulnerability.

Secure aggregation addresses one component of this problem by preventing the coordinating server from directly observing individual client contributions. Bonawitz and colleagues demonstrated a

protocol through which a server can recover an aggregate of high-dimensional client vectors while preserving the confidentiality of individual contributions under specified conditions. The protocol was designed to remain practical in environments where some participating clients may disconnect during execution. Secure aggregation is therefore particularly relevant when the server coordinating learning should not be trusted with individual updates.

Differential privacy provides a complementary protection mechanism. Rather than merely hiding individual messages from an aggregator, differential privacy introduces formal limits on the information that can be inferred about individual participation. Abadi and colleagues demonstrated how deep neural networks could be trained while tracking an explicit privacy budget, showing that meaningful predictive performance can remain possible under privacy-preserving training. Geyer, Klein, and Nabi further examined differential privacy from a federated client-level perspective and highlighted the trade-off between stronger privacy and model performance.

2.3 Heterogeneity, Optimization, and Analytical Performance

Federated learning introduces optimization difficulties that may be less pronounced in centralized environments. Participating clients can possess datasets of different sizes, feature distributions, labeling patterns, device capabilities, network conditions, and availability. Such statistical and system heterogeneity can influence convergence and produce unequal client contributions.

Li, Huang, Yang, Wang, and Zhang analyzed Federated Averaging under non-IID data and demonstrated that data heterogeneity can slow convergence. Their analysis also identified an important trade-off between communication efficiency and convergence behavior. These findings are highly relevant to privacy-preserving analytics because reducing communication, limiting participation, or introducing privacy noise may simultaneously influence computational efficiency and model quality.

Li, Sahu, Talwalkar, and Smith likewise emphasized that federated learning requires approaches different from standard large-scale machine learning because training occurs across heterogeneous devices or data silos while data remain localized. Consequently, a rigorous evaluation of privacy-preserving federated analytics must examine more than predictive accuracy. Privacy exposure, communication demand, participation behavior, convergence, resilience, and governance are also important performance dimensions.

3. Objectives of the Study

3.1 Privacy-Preservation Objective

The primary objective is to examine whether federated learning can reduce the direct exposure of sensitive information compared with conventional centralized analytical architectures. The study focuses particularly on the distinction between retaining raw data locally and protecting information exchanged during collaborative model training.

A related objective is to determine how secure aggregation and differential privacy can strengthen the privacy characteristics of baseline federated learning. Rather than assuming that decentralization is equivalent to complete privacy protection, the study evaluates privacy as a layered design requirement.

3.2 Analytical Utility Objective

The second objective is to assess the relationship between privacy-preserving mechanisms and predictive model utility. Increased privacy protection can alter the amount, precision, or visibility of information available during learning, potentially affecting predictive accuracy and convergence.

The study therefore compares simulated predictive performance across centralized analytics, baseline federated learning, federated learning with secure aggregation, and federated learning combining secure aggregation and differential privacy.

3.3 Operational and Governance Objective

The third objective is to examine the practical requirements associated with deploying federated analytics across heterogeneous participants. Particular attention is given to communication overhead, non-IID data, client participation, model-update protection, and governance responsibility.

The central research question is therefore:

To what extent can federated learning support privacy-preserving collaborative data analytics while maintaining acceptable analytical utility and manageable operational complexity?

4. Research Methodology

4.1 Research Design

The study adopts a simulation-based comparative analytical design. Simulation is used because no verified proprietary or institutional dataset was provided. The findings are therefore intended to demonstrate comparative relationships among privacy-preserving architectures rather than reproduce performance measured in an actual deployed federation.

A hypothetical federation of eight independent data-holding clients was modeled. The clients were assumed to possess locally stored classification datasets reflecting heterogeneous observations. No raw records were transferred among the clients in federated scenarios. The simulation compared four architectures: centralized analytics, baseline federated learning, federated learning with secure aggregation, and federated learning combining secure aggregation with differential-privacy protection.

The analytical model represents a conventional supervised classification problem. The same broad predictive objective was maintained across the four architectures so that differences could be interpreted primarily in relation to data location and privacy safeguards.

4.2 Federated Training Procedure

For the baseline federated environment, a simplified Federated Averaging process was represented. Let K denote the number of participating clients and n_k the number of observations held by client k .

After each local training round, the coordinating system produces a weighted global update:

$$w_{t+1} = \sum_{k=1}^K \frac{n_k}{n} w_{t+1}^{(k)}$$

where $w_{t+1}^{(k)}$ represents the locally updated model for client k , $n = \sum_{k=1}^K n_k$, and w_{t+1} represents the updated global model.

This formulation reflects the general logic of Federated Averaging described by McMahan and colleagues, although the present study does not reproduce their experiments or datasets.

In the secure-aggregation scenario, individual model updates were conceptually hidden from the coordinating server and only aggregate information was made available. In the differential-privacy

scenario, bounded perturbation was conceptually applied before aggregation to reduce the potential influence of individual client information.

4.3 Evaluation Criteria

The comparative analysis considers predictive accuracy, raw-data transfer, privacy-exposure risk, communication-cost index, and relative analytical utility. The privacy-exposure risk index is a synthetic comparative measure rather than a standardized privacy metric. A lower value represents lower modeled exposure under the assumptions of the simulation.

Table 1: Simulation Framework for Federated Privacy-Preserving Analytics

Methodological Component	Simulation Specification	Analytical Purpose
Research design	Comparative simulation	Evaluate alternative analytics architectures
Participating clients	8 independent data holders	Represent decentralized organizations or devices
Data distribution	Heterogeneous/non-IID synthetic classification records	Reflect realistic federation differences
Baseline architecture	Centralized analytics	Establish comparison condition
Federated architecture	Federated Averaging-based training	Keep original datasets localized
Privacy enhancement 1	Secure aggregation	Conceal individual client updates
Privacy enhancement 2	Differential privacy	Reduce inference about individual contributions
Evaluation indicators	Accuracy, raw-data movement, privacy exposure and communication burden	Assess privacy–utility trade-offs
Evidence status	Simulated/illustrative	Prevent unsupported empirical claims

The study does not claim that the selected numerical risk values constitute universal privacy guarantees. Actual privacy depends on threat models, client population size, privacy-budget configuration, aggregation mechanisms, implementation quality, cryptographic assumptions, system architecture, and adversarial capability.

5. Analysis

5.1 Comparative Analytical Performance

The simulated results demonstrate that decentralization can substantially reduce the requirement for direct raw-data transfer while retaining relatively strong predictive performance. Centralized analytics produced the highest modeled classification accuracy because all training information was assumed to

be directly accessible in a single environment. Baseline federated learning produced a modest decline in accuracy owing to client heterogeneity and distributed optimization.

Secure aggregation produced only a limited additional reduction in model utility in the illustrative scenario because its primary function was to protect individual updates rather than materially alter the underlying training objective. The combination of secure aggregation and differential privacy produced the lowest predictive accuracy among the four configurations, illustrating the expected privacy–utility trade-off.

Table 2: Simulated Performance of Privacy-Preserving Analytics Architectures

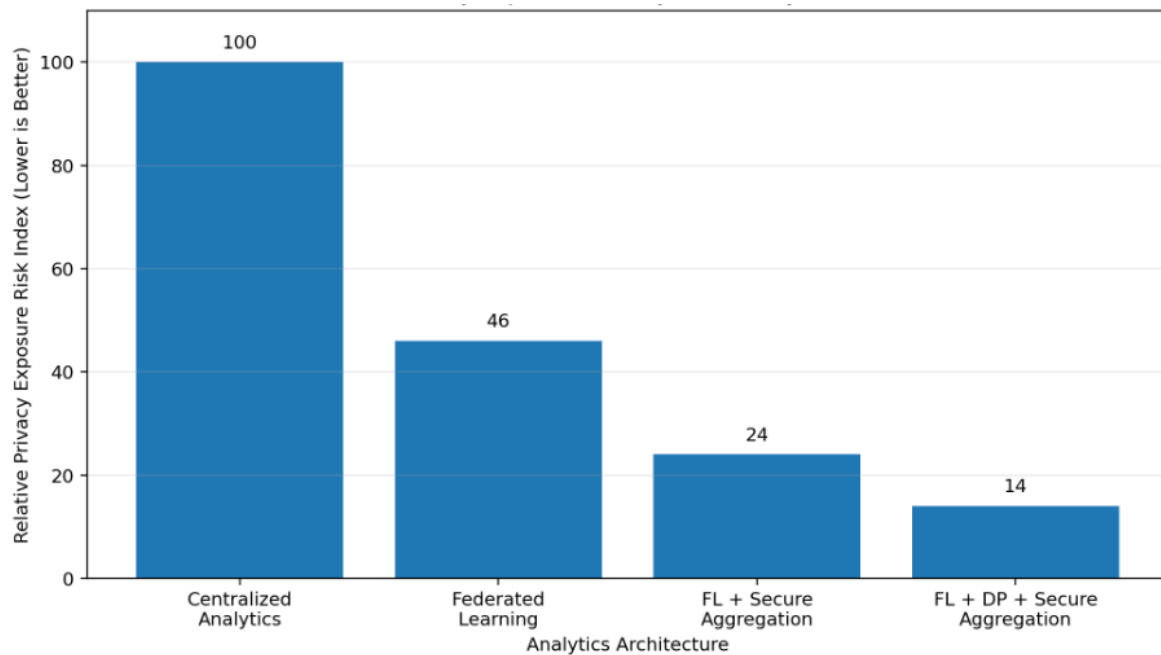
Performance Indicator	Centralized Analytics	Federated Learning	FL + Secure Aggregation	FL + DP + Secure Aggregation
Predictive accuracy (%)	92.8	91.9	91.7	90.6
Raw data transferred to central server (%)	100	0	0	0
Relative privacy-exposure risk index	100	46	24	14
Communication-cost index	45	68	82	86
Relative analytical utility index	100	98	97	94
Individual updates visible to coordinating server	Not applicable	Potentially	Protected	Protected
Added formal privacy mechanism	None	None	Secure aggregation	Secure aggregation + differential privacy

5.2 Privacy Exposure Across Analytical Architectures

The strongest simulated difference concerns privacy exposure. Centralized analytics was assigned the reference risk index of 100 because all raw training data were assumed to be transferred to and stored within a common analytical environment. Baseline federated learning reduced the index to 46 because original records remained local, although individual model updates remained a potential source of information exposure.

Secure aggregation reduced the modeled risk index to 24 by limiting direct server visibility into individual client updates. When differential privacy was added to secure aggregation, the simulated risk index declined further to 14.

Graph 1: Simulated Privacy Exposure Risk by Data-Analytics Architecture



Relative to the centralized reference condition, baseline federated learning produced a simulated privacy-exposure reduction of:

$$100 - 46 \div 100 \times 100 = 54\% \quad \frac{100 - 46}{100} \times 100 = 54\%$$

The secure-aggregation scenario produced a simulated reduction of:

$$100 - 24 \div 100 \times 100 = 76\% \quad \frac{100 - 24}{100} \times 100 = 76\%$$

The combined secure-aggregation and differential-privacy scenario produced a simulated reduction of:

$$100 - 14 \div 100 \times 100 = 86\% \quad \frac{100 - 14}{100} \times 100 = 86\%$$

These figures illustrate the comparative logic of layered privacy mechanisms rather than standardized privacy guarantees.

5.3 Privacy–Utility and Communication Trade-Off

The simulation demonstrates that stronger privacy safeguards are not costless. Predictive accuracy declined from 92.8% in the centralized condition to 90.6% under the strongest modeled federated privacy configuration. The reduction of 2.2 percentage points was modest within this scenario, but actual outcomes could vary substantially depending on privacy budgets, data heterogeneity, number of participants, local sample sizes, optimization algorithms, and task complexity.

Communication demand also increased. Centralized model training was represented with a communication index of 45 because data transfer occurred principally during data collection. Federated learning required repeated model distribution and update exchange, increasing the index to 68. Secure aggregation and differential-privacy operations further increased computational and communication complexity.

These findings illustrate an important design principle: privacy-preserving data analytics cannot be optimized through a single performance metric. A technically acceptable implementation must

balance predictive utility, confidentiality, communication cost, computation, convergence time, attack resistance, and organizational governance.

6. Challenges

6.1 Statistical Heterogeneity and Non-IID Data

One of the most significant challenges in federated learning is that client datasets may differ substantially. A hospital specializing in cardiovascular care will produce a different patient distribution from a general hospital. A financial institution operating in one market may observe transaction behaviors that differ from those of institutions in other regions. Mobile-device users similarly generate highly personalized data.

Such heterogeneity means that local optimization directions may conflict with one another. Li and colleagues showed theoretically that non-IID data can slow the convergence of Federated Averaging and create important trade-offs between communication efficiency and convergence.

Strategies such as client clustering, personalized federated learning, adaptive aggregation, proximal optimization, client reweighting, and local fine-tuning may help address this issue. Nevertheless, increased methodological sophistication can also complicate validation and governance.

6.2 Privacy Attacks and Security Limitations

Federated learning is sometimes described too simplistically as privacy-preserving because original records do not leave clients. This interpretation overlooks attacks targeting gradients, model updates, global parameters, malicious participants, and the aggregation process.

Secure aggregation can reduce server-level visibility into individual updates, while differential privacy can limit information leakage associated with individual contributions. However, privacy mechanisms introduce assumptions and trade-offs. Differential privacy depends on careful selection and accounting of the privacy budget, while cryptographic aggregation depends on correct implementation, participant authentication, and protocol security.

Federated learning must therefore operate within a broader security architecture involving authentication, encryption in transit, access control, model integrity verification, anomaly monitoring, and incident-response mechanisms.

6.3 Communication and System Heterogeneity

Repeated model exchange can become costly when federations contain thousands or millions of participants. Devices may also have different computational power, battery capacity, network bandwidth, software versions, and availability.

McMahan and colleagues identified communication efficiency as a fundamental constraint in federated training, while later research continued to emphasize systems heterogeneity as a major challenge.

Client selection must consequently balance learning quality and operational feasibility. Repeatedly selecting only high-performance devices may improve training speed but create representation problems if slower or less-connected clients contain systematically different data.

7. Discussion

7.1 Federated Learning as a Data-Minimization Architecture

The simulated findings support the interpretation of federated learning as a mechanism for reducing unnecessary data concentration. Centralized analytics requires participating organizations or devices to relinquish direct control over copies of their raw information, whereas federated learning allows model computation to occur closer to the original data source.

This design is consistent with the broader principle of data minimization. Kairouz and colleagues note that federated learning can mitigate several privacy risks and costs associated with centralized machine-learning systems because the training data remain decentralized.

However, federated learning should not be interpreted as equivalent to complete anonymization. The model itself, updates transmitted during training, client participation information, and intermediate statistics may remain privacy relevant. Privacy-preserving federated analytics therefore requires layered protection.

7.2 Importance of Secure Aggregation and Differential Privacy

The simulation indicates that the strongest privacy configuration combines decentralization with protection of model updates and formal privacy mechanisms. Secure aggregation protects the confidentiality of individual update vectors from the coordinating server, whereas differential privacy limits information attributable to individual contributions.

Bonawitz and colleagues demonstrated that secure aggregation can be designed in a manner that is robust to participant dropout while allowing the server to obtain aggregate information. Abadi and colleagues demonstrated that neural networks can be trained under differential-privacy constraints while maintaining usable model quality, although privacy protection affects computational and analytical performance.

The present results therefore reinforce the principle that privacy-preserving analytics should be designed as a layered architecture. Local data retention, secure transmission, protected aggregation, differential privacy, access controls, monitoring, and governance address different components of the risk landscape.

7.3 Implications for Collaborative Data Analytics

Federated learning offers particular value when multiple organizations would benefit from collaborative analytics but cannot establish unrestricted centralized data sharing. Healthcare institutions could collaboratively improve predictive models while maintaining patient records within institutional infrastructure. Financial institutions could cooperate in risk or fraud modeling without directly exposing customer transaction databases. Industrial organizations could collaboratively learn failure patterns while retaining proprietary operational data.

The approach may also support cross-device applications in which large numbers of mobile or edge devices contribute to a common model. In such settings, federated learning can reduce the need to continuously upload complete user datasets.

Nevertheless, organizational governance remains essential. Participating parties must define who controls the global model, who may initiate training, which clients are eligible to participate, how

malicious updates are handled, how privacy budgets are allocated, how model outputs may be reused, and how accountability is distributed among participants.

Federated analytics therefore creates not only a technical federation but also a governance federation. Its long-term success depends on establishing clear institutional rules around access, responsibility, model ownership, security, withdrawal, auditability, and acceptable downstream use.

8. Conclusion

Federated learning provides a significant architectural alternative to centralized machine learning for environments where analytical collaboration must coexist with privacy, confidentiality, or institutional data-control requirements. By allowing participating clients to train shared models without transferring their original training records to a common repository, federated learning can reduce direct raw-data exposure and support data-minimization objectives.

The simulation-based analysis presented in this study indicates that federated learning can maintain comparatively strong analytical utility while substantially reducing modeled privacy exposure. Predictive accuracy declined only moderately across the illustrative architectures, from 92.8% under centralized analytics to 91.9% under baseline federated learning and 90.6% under the combined differential-privacy and secure-aggregation architecture. At the same time, the simulated privacy-exposure index declined from 100 to 14 under the strongest privacy configuration. These values should be understood solely as comparative simulated outcomes rather than universal performance benchmarks.

The analysis further demonstrates that keeping data local is not sufficient to establish comprehensive privacy protection. Individual model updates can themselves become sensitive. Secure aggregation can reduce visibility into these contributions, while differential privacy can introduce stronger formal protections against inference relating to individual participants. These safeguards nevertheless introduce additional communication, computation, governance complexity, and potential reductions in predictive utility.

A robust federated-learning implementation should consequently be designed around a clearly defined threat model and should include secure communication, authenticated participation, aggregation protection, privacy accounting, model monitoring, anomaly detection, and transparent governance. Non-IID data, client dropout, limited network connectivity, poisoning attacks, and system heterogeneity must also be considered during implementation.

The central contribution of federated learning is therefore not the elimination of privacy risk but the restructuring of data analytics so that collaborative learning can occur with substantially less unnecessary movement and concentration of raw information. When combined with secure aggregation, differential privacy, strong systems security, and appropriate organizational governance, federated learning provides a promising foundation for privacy-conscious collaborative analytics across devices, institutions, and organizational boundaries.

References

1. McMahan, B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 54, 1273–1282.

2. Konečný, J., McMahan, H. B., Ramage, D., & Richtárik, P. (2016). Federated optimization: Distributed machine learning for on-device intelligence. *arXiv preprint arXiv:1610.02527*.
3. Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2016). Federated learning: Strategies for improving communication efficiency. *arXiv preprint arXiv:1610.05492*.
4. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 1175–1191. <https://doi.org/10.1145/3133956.3133982>
5. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2016). Practical secure aggregation for federated learning on user-held data. *arXiv preprint arXiv:1611.04482*.
6. Shokri, R., & Shmatikov, V. (2015). Privacy-preserving deep learning. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (pp. 1310–1321). ACM. <https://doi.org/10.1145/2810103.2813687>
7. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (pp. 308–318). ACM. <https://doi.org/10.1145/2976749.2978318>
8. Dwork, C., Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4), 211–407. <https://doi.org/10.1561/04000000042>
9. Dwork, C., & Rothblum, G. N. (2016). Concentrated differential privacy. *arXiv preprint arXiv:1603.01887*.
10. Geyer, R. C., Klein, T., & Nabi, M. (2017). Differentially private federated learning: A client level perspective. *arXiv preprint arXiv:1712.07557*.
11. Truex, S., Baracaldo, N., Anwar, A., Steinke, T., Ludwig, H., Zhang, R., & Zhou, Y. (2018). A hybrid approach to privacy-preserving federated learning. *arXiv preprint arXiv:1812.03224*.
12. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology*, 10(2), Article 12. <https://doi.org/10.1145/3298981>
13. Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., Kiddon, C., Konečný, J., Mazzocchi, S., McMahan, H. B., Van Overveldt, T., Petrov, D., & Roselander, J. (2019). Towards federated learning at scale: System design. *Proceedings of Machine Learning and Systems*, 1, 374–388.
14. Aono, Y., Hayashi, T., Wang, L., Moriai, S., & others. (2017). Privacy-preserving deep learning via additively homomorphic encryption. *IEEE Transactions on Information Forensics and Security*, 13(5), 1153–1165.
15. Phong, L. T., Aono, Y., Hayashi, T., Wang, L., & Moriai, S. (2018). Privacy-preserving deep learning via weight encryption. *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2772–2776.
16. Vaidya, J., & Clifton, C. (2002). Privacy-preserving k-means clustering over vertically partitioned data. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 206–215). ACM.



17. Lindell, Y., & Pinkas, B. (2000). Privacy preserving data mining. *Advances in Cryptology—CRYPTO 2000*, 36–54. Springer.
18. Yang, K., & Barria, J. A. (2006). A distributed privacy-preserving framework for data mining. *IEEE Transactions on Knowledge and Data Engineering*, 18(11), 1517–1531.
19. Pathak, M. A., Rane, S., & Raj, B. (2010). Multiparty differential privacy via aggregation of locally trained classifiers. *Advances in Neural Information Processing Systems*, 23, 1876–1884.
20. Duchi, J. C., Jordan, M. I., & Wainwright, M. J. (2013). Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science* (pp. 429–438). IEEE. <https://doi.org/10.1109/FOCS.2013.53>