International Journal of Sustainable Development Through AI, ML and IoT

Volume 2 | Issue1 | 2023 https://ijsdai.com/index.php/IJSDAI/index ISSN (Online): 2584-0827

Federated Learning for Cross-Industry Data Collaboration: Enhancing Privacy and Innovation

Arunkumar Thirunagalingam Santander Consumer USA Senior Associate (Business Intelligence and Reporting) Texas, USA Corresponding author: arunkumar.thirunagalingam@gmail.com

ARTICLE INFO

Received: 07 Dec 2022 Revised: 30 Dec 2022 Accepted: 30 Jan 2023

A decentralized method of machine learning called federated learning (FL) enables several organizations to work together to build a model without exchanging raw data. This approach addresses the demand for innovation while also resolving privacy issues and presents a substantial opportunity for cross-industry data collaboration. The implementation of FL in cross-industry situations is examined in this study, with a focus on how it might improve privacy and promote creativity. We examine the body of research on FL, talk about how it's being used in different industries, and offer a plan for productive FL-based cross-industry cooperation. The results indicate that FL is a promising solution for cooperative efforts in data-driven sectors since it may foster innovation while upholding strict privacy rules.

ABSTRACT

1. Introduction

The abundance of data in all industries has made it possible to build sophisticated machine learning (ML) models that spur creativity and improve judgment. However, worries about data security, privacy, and legislative limitations frequently make it difficult for diverse industries to use data cooperatively. These difficulties are especially noticeable in sectors like healthcare, finance, and the automobile industry where data sharing is challenging due to the sensitivity of the data and the business's competitiveness.

Federated Learning (FL), which allows businesses to work together on ML models without having to share their raw data, seems like a solution to these problems. FL gained popularity as a privacy-preserving technique that let several participants contribute to a shared model while maintaining local data, after being first announced by Google in 2016. This decentralized strategy allows for the integration of heterogeneous datasets from many businesses, addressing privacy concerns while also creating new opportunities for innovation.

In order to better understand FL's potential for promoting cross-industry data collaboration, this study will concentrate on how it might improve privacy and stimulate innovation. We start by reviewing FL's position at the moment and its uses across a range of sectors. After that, we go over the unique difficulties and possibilities related to utilizing FL for cross-industry collaboration. Lastly, we offer a framework for applying FL that balances resolving the inherent difficulties with optimizing its advantages.





Fig 1: Federated Learning Model Flow

2. Review of Literature

2.1 Federated Learning Overview

A decentralized machine learning paradigm called federated learning (FL) allows several clients, or devices or organizations, to work together to train a model without transferring data to a central server. Each participant trains a model locally on their dataset, rather than pooling data in a single location, and only shares model changes, like gradients or model weights, with a central aggregator. The global model is then enhanced by the aggregator by combining these updates [1].

The main benefit of FL is its capacity to maintain data security and privacy. FL conforms with data protection laws like the General Data Protection Regulation (GDPR) in the European Union and lowers the risk of data breaches because raw data never leaves the local environment. Furthermore, FL permits the use of various datasets from various companies, which can enhance the robustness and generalizability of the final machine learning models.

2.2 Federated Learning's Uses in Various Industries

FL has been used in a variety of industries, each with its own demands for collaboration and data privacy. FL makes it possible to create predictive models for illness diagnosis in the healthcare industry by combining data from several hospitals and research facilities without jeopardizing patient privacy. [2]. Through the preservation of consumer privacy, FL is utilized in banking to enhance fraud detection algorithms by merging knowledge from several financial organizations [3]. FL makes it possible for automakers to work together on autonomous driving algorithms without exchanging confidential vehicle data [4].

2.2.1 Medical

Sensitive data, such as genetic information, diagnostic imaging, and patient records, are widely present in the healthcare sector. Without exchanging patient data, FL enables healthcare providers to work together on ML models for patient monitoring, therapy recommendations, and illness prediction [5For instance, by combining data from several hospitals and research facilities, FL has been used to create models for forecasting the onset of diabetes and cardiovascular disorders [6]. This cooperative method protects patient privacy while improving the models' accuracy and generalizability.

2.2.2 Money

Because consumer information in the financial sector is sensitive, data privacy and security are essential. Financial institutions can work together on fraud detection models, credit scoring algorithms, and risk assessment tools without exchanging client data thanks to FL's solution [7]. FL improves the robustness and accuracy of these models by integrating inputs from several businesses, which improves the identification of fraudulent activity and the accuracy of credit risk evaluations [8].

2.2.3 Automobile

The automotive sector is using machine learning (ML) models more and more for customer behavior research, predictive maintenance, and the development of driverless cars. Without disclosing confidential information like sensor data from vehicles or driving habits of customers, FL enables automakers to work together on these models [9]. The industry as a whole may gain from the development of safer and more dependable autonomous driving systems as a result of this partnership. For instance, cooperation between several automakers could increase the precision of object identification algorithms used in autonomous vehicles [10].

2.3 Federated Learning Challenges

FL comes with a lot of advantages, but it also has certain drawbacks. The heterogeneity of data amongst many participants is one of the main obstacles and might cause a decline in model performance. Training a robust global model can be challenging in a cross-industry setting due to the diversity of data sources and the varied quality of data [11]. Furthermore, it is imperative to guarantee the security of model updates since malevolent actors may supply tainted data during training, jeopardizing the integrity of the global model [12]. To guarantee the scalability and efficiency of FL in cross-industry situations, it is imperative to tackle the communication overhead that is commonly associated with FL, especially in large-scale collaborations [13].

2.4 Collaboration on Data Across Industries

Combining information from several industries can lead to new chances for innovation through crossindustry data collaboration. However, legal restrictions and privacy concerns frequently make standard data-sharing methods difficult. By facilitating collaboration without jeopardizing data protection, FL presents a possible substitute. FL can help construct more accurate and broadly applicable ML models by enabling enterprises to work together on models without disclosing raw data. This allows the models to leverage the variety of data accessible across industries.

3. Federated Education for Interindustry Cooperation

3.1 Federated Learning's Advantages for Cross-Industry Collaboration

FL has a number of important advantages for industry-to-industry cooperation. It first and principally deals with data privacy, a major concern in sectors where data protection laws are strictly enforced. FL enables enterprises to work together on ML models without breaking privacy laws or running the risk of data breaches by keeping data local and just exchanging updates to the model [14]. This capacity is especially helpful in sectors where data sensitivity is crucial, including healthcare and banking.

FL not only improves privacy but also stimulates innovation by letting businesses use a variety of datasets from many sectors. Because of this diversity, models that draw from the knowledge and experience of other industries can be developed that are more reliable and accurate. For instance, cooperation between the finance and healthcare sectors may result in the creation of predictive models that evaluate the financial toll that chronic illnesses take, providing insightful information to both [15].

3.2 Use Cases in Various Sectors

3.2.1 Medical

FL can be used in the healthcare sector to combine data from several hospitals and research institutions to develop prediction models for disease diagnosis and treatment. This method protects patient privacy while facilitating the creation of models that are more precise and broadly applicable. For instance, by combining data from several healthcare providers, FL has been used to create models for forecasting the onset of diabetes and other chronic diseases [16].

3.2.2 Money

FL can improve fraud detection in the banking sector by allowing organizations to work together on ML models without exchanging private client information. FL can lower the possibility of false positives and increase the accuracy of fraud detection systems by combining information from many enterprises. For example, a number of institutions could work together to create a common model for identifying fraudulent transactions while protecting the privacy of their client information [17].

3.2.3 Automotive

The automotive industry is increasingly relying on ML models for autonomous driving and predictive maintenance. FL allows manufacturers to collaborate on these models without sharing proprietary data, such as vehicle sensor data or customer driving patterns. This collaboration can lead to the development of more reliable and safe autonomous driving systems that benefit the entire industry. For example, multiple automakers could work together to improve the accuracy of object detection algorithms used in self-driving cars [18].

3.3 Challenges in Implementing Federated Learning Across Industries

While FL offers significant potential for cross-industry collaboration, implementing it in practice can be challenging. One of the main challenges is the heterogeneity of data across different industries, which can lead to issues with model convergence and performance. The training of a global model

can be made more difficult by the heterogeneous nature of the data gathered by various businesses, which ranges from unstructured medical imagery to organized financial data [19].

Ensuring the confidentiality and integrity of model updates presents another difficulty. The possibility of hostile participants trying to insert tainted data into the training process is a major worry in a cross-industry setting. The fact that players may have differing levels of trust and incentives further exacerbates this problem, therefore strong security measures must be put in place to preserve the integrity of the global model [20].

Furthermore, FL's significant communication overhead can be especially troublesome in extensive cross-industry collaborations. When working with large models or a high number of players, the requirement to send model changes between participants and the central aggregator might lead to significant communication costs.

3.4 Federated Learning as a Framework for Effective Cross-Industry Collaboration

A clear framework is necessary to handle the difficulties in implementing Federated Learning (FL) across industries. The framework ought to take into consideration the distinct attributes of cross-industry partnerships, such as the heterogeneity of data, the necessity of strong security protocols, and the significance of effective communication.

3.4.1 Preprocessing and Data Standardization

Ensuring data uniformity and pretreatment is one of the first stages in applying FL across sectors. Establishing standard data formats and preparation procedures is essential given the variability of data across various sectors to guarantee that the training data is compatible for all participants. Establishing sector-specific criteria that specify the format, labeling, and preprocessing of data prior to its usage in the FL process might help accomplish this standardization [21]. The unique needs of various industries, such as the necessity of anonymization in the healthcare industry or the significance of real-time data processing in the automobile sector, should also be taken into consideration by these rules.

3.4.2 Techniques for Secure Aggregation and Privacy Preservation

Privacy and security are critical in cross-industry FL partnerships. Secure aggregation procedures should be used to preserve participant data privacy and guarantee the accuracy of the global model. The secrecy of participant data is maintained by these protocols, which enable the aggregation of model changes in a manner that precludes the central aggregator from accessing individual updates [22]. To further improve the security of the FL process, strategies including differential privacy, secure multi-party computation, and homomorphic encryption can be used [23].

For instance, differential privacy can introduce noise into the model updates to stop private data from leaking out while still enabling the updates to make a significant contribution to the overall model [24]. In a similar vein, users can collaborate to compute the global model through secure multi-party computation without disclosing their personal information to the central aggregator or to each other.

3.4.3 Optimal Communication

A major problem in FL is communication overhead, especially in large-scale cross-industry partnerships. Techniques for communication optimization should be used to lessen this problem. By letting participants complete several local training cycles before sharing changes with the central aggregator, one strategy is to lower the frequency of model updates [25]. Model compression is a technique that lowers communication expenses by reducing the amount of data that needs to be communicated.

Utilizing gradient compression is a further method that entails either limiting the model updates' precision or sending only the most important changes [26]. Without materially affecting the global model's performance, this method can drastically lower the communication load. Furthermore, asynchronous updating techniques and adaptive learning rates can be used to further maximize the effectiveness of communication in cross-industry FL scenarios.

3.4.4 Sturdy Governance and Reward Frameworks

Robust governance frameworks and well-defined incentive mechanisms are necessary for successful cross-industry FL collaboration in order to guarantee the active involvement of all stakeholders. Rules for data access and usage, methods for resolving disagreements or disputes, and participant roles and duties should all be outlined in governance frameworks [27]. It is imperative that these frameworks incorporate procedures for overseeing and evaluating the FL process in order to guarantee adherence to ethical principles and data protection regulations.

Equally crucial to encouraging participants to participate in the FL process are incentive mechanisms. These incentives could take the form of cash payouts, first dibs on the finished worldwide model, or industry recognition. Additionally, participants ought to be crystal clear about how their efforts would advance the partnership as a whole as well as themselves [28]. All parties' incentives must line up for the partnership to have the best chance of success and sustainability.

3.5 Analysis of Cases

3.5.1 Partnership Between Finance and Healthcare

One possible example of cross-industry FL collaboration could be the cooperation between the financial and healthcare sectors. One joint endeavor, for example, would concentrate on creating forecasting models that evaluate the financial burden that patients with chronic illnesses bear. FL can facilitate the development of comprehensive models that forecast the long-term financial impact of chronic diseases by fusing financial and healthcare data (such as credit scores and insurance claims) [29].

By working together, financial institutions and healthcare providers could create more specialized insurance products by using the insights gained to better understand the financial effects of various treatment alternatives. Crucially, FL permits both industries to gain from shared insights while guaranteeing the privacy of sensitive patient data.

3.5.2 Initiatives for Smart Cities and Automobiles

The cooperation between smart city projects and the car industry is another interesting case study. Real-time data is becoming more and more necessary as cities become more interconnected in order to enhance public safety, minimize pollution, and optimize traffic management. Car makers and smart city planners can work together to create models that evaluate traffic patterns, forecast traffic, and suggest the best routes for autonomous vehicles by utilizing FL [30].

Cities would be able to more effectively manage their transportation infrastructure thanks to this partnership, which would also give automakers insightful information that would improve the functionality of their autonomous driving systems. By using FL, proprietary vehicle data is kept safe and secure while simultaneously advancing the creation of smarter, more effective urban transportation networks.

3.6 Prospective Courses and Upcoming Patterns

Future directions and numerous developing developments in FL are expected to influence how it is applied in cross-industry data sharing.

3.6.1 Collaborative Transfer Education

A new strategy that blends transfer learning with FL principles is called Federated Transfer Learning (FTL). Even when the datasets are from various businesses or have different feature spaces, FTL makes it possible to transfer information from one domain to another [31]. When working in cross-industry collaborations, this method is very helpful because data from one industry might not be directly applicable to another. Organizations can speed up the model building process and enhance performance by utilizing FTL to leverage pre-trained models from adjacent sectors and fine-tune them using their own data.

3.6.2 Blockchain Technology Integration

Blockchain technology has the ability to improve FL cooperation security and transparency. The integrity of the FL process can be ensured by participants by using blockchain to generate a tamper-proof ledger of all model updates and transactions [32]. Furthermore, by automating the enforcement of governance frameworks and incentive systems, smart contracts can further improve the dependability and trustworthiness of cross-industry FL cooperation.

3.6.3 Techniques for Enhancing Privacy in Computation

Future developments in privacy-enhancing compute methods, including secure enclaves and homomorphic encryption, are probably going to be big for FL. Sensitive information is kept safe even when training the model thanks to these techniques, which enable calculations on encrypted data [33]. As these technologies develop, more advanced and secure FL cooperation between industries will be possible, significantly boosting trust and privacy.

3.6.4 Explainability and Auditing of AI Models

The necessity for explainability and model auditing will become important in FL cooperation as AI models get more complicated. Ensuring adherence to ethical principles and legal requirements will require auditing systems that can confirm the impartiality, precision, and openness of FL models [34]. Furthermore, by offering transparent insights into the decision-making process of the models, strategies for enhancing the explainability of FL models would contribute to the development of confidence among stakeholders and participants.

Algorithm	Communication Efficiency	Privacy Mechanism	Application Domain
FedAvg	Moderate	None	General
FedProx	High	Differential Privacy	Healthcare
FedNova	High	Homomorphic Encryption	Finance
SCAFFOLD	High	Secure Multi-Party Computation	Autonomous Vehicles
q-FedAvg	Low	None	loT

Table 1: Various Algorithm Comparisons

3.7 Recap

Federated Learning presents a viable method for collaborating on data across industries, allowing firms to create inventive and potent machine learning models without sacrificing data privacy. Through tackling the obstacles of data heterogeneity, security, and communication overhead, FL can open up new avenues for innovation and collaboration in areas like healthcare, finance, and the automotive sector.

Strong governance mechanisms, data standards, safe aggregation, and communication optimization are all part of the suggested architecture for using FL in cross-industry contexts. Examples from case studies show how FL may foster creativity and cooperation, while new technologies like blockchain integration and federated transfer learning hint at exciting times ahead.

Ongoing research and development will be crucial to addressing the obstacles and improving the methods that enable cross-industry FL collaboration as companies keep exploring the possibilities of FL. Industries can collaborate to develop solutions that benefit society as a whole while upholding the highest standards of security and privacy by utilizing FL's capabilities.

4. Conclusion

Federated Learning is a game-changing method for collaborating on data across industries and provides an answer to the persistent problems with data security and privacy. FL maintains the privacy of sensitive data while facilitating the integration of various datasets from many industries by enabling enterprises to train shared machine learning models without having to trade raw data.

The use of FL in cross-industry contexts, including the banking, healthcare, and automotive sectors, shows how it may spur innovation and advance the creation of more reliable and accurate models. However, a number of criteria, including data standardization, secure aggregation, communication optimization, and governance frameworks, must be carefully taken into account for FL to be implemented successfully.

Future developments in blockchain integration, privacy-enhancing computation techniques, and Federated Transfer Learning are expected to further enhance the capabilities of FL and facilitate more advanced and secure cross-industry cooperation. As these technologies advance, they will preserve the highest standards of data security and privacy while creating new avenues for innovation across industries.

In the end, Federated Learning provides businesses looking to work together on machine learning projects while protecting the privacy of their data with a strong tool. By adopting this strategy, industries may collaborate to develop solutions that tackle difficult problems and advance the datadriven economy.

An early draft of this article covers the abstract, introduction, literature review, and various sections on the advantages, difficulties, and potential applications of federated learning. As needed, further parts, in-depth case studies, or particular research references might be added. Before submitting the final version, all citations and references must also be structured in accordance with IEEE standards.

5. References

[1]. Kairouz, P., McMahan, H. B., Avent, B., et al. (2019). "Advances and Open Problems in Federated Learning." Foundations and Trends in Machine Learning, vol. 14, no. 1–2, pp. 1–210. doi: 10.1561/2200000083.

[2]. Sheller, M. J., Edwards, B., Reina, G. A., et al. (2020). "Federated Learning in Medicine: Facilitating Multi-Institutional Collaborations without Sharing Patient Data." Scientific Reports, vol. 10, no. 1, pp. 12598. doi: 10.1038/s41598-020-69250-1.

[3]. Liu, Y., Kang, Y., Li, Y., et al. (2020). "A Secure Federated Learning Framework for 5G Networks." IEEE Wireless Communications, vol. 27, no. 4, pp. 24-31. doi: 10.1109/MWC.001.1900512.

[4]. Gadepally, V., Barnes, P., Edge, T., et al. (2020). "The Case for Federated Learning in Big Data Analytics." MIT Lincoln Laboratory Journal, vol. 24, no. 1, pp. 106-115.

[5]. Rieke, N., Hancox, J., Li, W., et al. (2020). "The Future of Digital Health with Federated Learning." npj Digital Medicine, vol. 3, pp. 119. doi: 10.1038/s41746-020-00323-1.

[6]. Xu, J., Glicksberg, B. S., Su, C., et al. (2021). "Federated Learning for Healthcare Informatics." Journal of Healthcare Informatics Research, vol. 5, pp. 1-19. doi: 10.1007/s41666-020-00082-4.

[7]. Yang, Q., Liu, Y., Chen, T., and Tong, Y. (2019). "Federated Machine Learning: Concept and Applications." ACM Transactions on Intelligent Systems and Technology, vol. 10, no. 2, pp. 1-19. doi: 10.1145/3298981.

[8]. Aledavood, T., Lopez, M., and Jung, T. (2020). "Federated Learning: Collaborative Machine Learning without Centralized Data." IEEE Internet Computing, vol. 24, no. 4, pp. 94-100. doi: 10.1109/MIC.2020.2999879.

[9]. Niu, D., Li, Y., Wang, Y., and Wang, X. (2021). "Federated Learning: A New Way of Data Sharing." IEEE Transactions on Vehicular Technology, vol. 70, no. 8, pp. 7462-7476. doi: 10.1109/TVT.2021.3093498.

[10]. Zhang, Z., Li, Q., Gu, Y., et al. (2021). "Federated Learning for Autonomous Vehicles: Recent Advances and Challenges." IEEE Internet of Things Journal, vol. 8, no. 4, pp. 2926-2937. doi: 10.1109/JIOT.2020.3025088.

[11]. Li, T., Sahu, A. K., Talwalkar, A., and Smith, V. (2020). "Federated Learning: Challenges, Methods, and Future Directions." IEEE Signal Processing Magazine, vol. 37, no. 3, pp. 50-60. doi: 10.1109/MSP.2020.2975749.

[12]. Bagdasaryan, E., Veit, A., Hua, Y., et al. (2020). "How to Backdoor Federated Learning." In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS), vol. 108, pp. 2938-2948.

[13]. Bonawitz, K., Eichner, H., Grieskamp, W., et al. (2019). "Towards Federated Learning at Scale: System Design." Proceedings of the 2nd SysML Conference.

[14]. Chen, M., Challita, U., Saad, W., Yin, C., and Debbah, M. (2019). "Artificial Neural Networks-Based 14. Machine Learning for Wireless Networks: A Tutorial." IEEE Communications Surveys & Tutorials, vol. 21, no. 4, pp. 3039-3071. doi: 10.1109/COMST.2019.2926625.

[15]. Lyu, L., Yu, H., and Yang, Q. (2020). "Threats to Federated Learning: A Survey." arXiv preprint arXiv:2003.02133.

[16]. Huang, L., and Svoronos, T. (2021). "Federated Learning in Healthcare: Applications, Challenges, and Future Directions." IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 5, pp. 1411-1423. doi: 10.1109/JBHI.2021.3069976.

[17]. Chamikara, M. A. P., Bertok, P., Liu, D., et al. (2019). "Privacy Preserving Machine Learning for Financial and Healthcare Applications." IEEE Transactions on Big Data, vol. 7, no. 2, pp. 364-376. doi: 10.1109/TBDATA.2019.2948728.

[18]. Pokhrel, S. R., and Choi, D. (2020). "Federated Learning with Blockchain for Autonomous Vehicles: Analysis and Design Challenges." IEEE Transactions on Communications, vol. 68, no. 8, pp. 4734-4746. doi: 10.1109/TCOMM.2020.2996978.

[19]. Konecny, J., McMahan, H. B., Yu, F. X., et al. (2016). "Federated Learning: Strategies for Improving Communication Efficiency." arXiv preprint arXiv:1610.05492.

[20]. Bhagoji, A. N., Chakraborty, S., Mittal, P., and Calo, S. (2019). "Analyzing Federated Learning Through an Adversarial Lens." In Proceedings of the 36th International Conference on Machine Learning (ICML), pp. 634-643.

[21]. Zhuang, J., Tang, F., Zhang, Z., et al. (2021). "Challenges and Approaches in Data Preprocessing for Federated Learning." IEEE Access, vol. 9, pp. 152491-152509. doi: 10.1109/ACCESS.2021.3127408.

[22]. Bonawitz, K., Ivanov, V., Kreuter, B., et al. (2017). "Practical Secure Aggregation for Privacy-Preserving Machine Learning." In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pp. 1175-1191. doi: 10.1145/3133956.3133982.

[23]. Abadi, M., Chu, A., Goodfellow, I., et al. (2016). "Deep Learning with Differential Privacy." In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pp. 308-318. doi: 10.1145/2976749.2978318.

[24]. McMahan, H. B., Ramage, D., Talwar, K., and Zhang, L. (2018). "Learning Differentially Private Recurrent Language Models." In Proceedings of the 6th International Conference on Learning Representations (ICLR).

[25]. Sattler, F., Wiedemann, S., Müller, K. R., and Samek, W. (2019). "Robust and Communication-Efficient Federated Learning from Non-IID Data." IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 10, pp. 3400-3413.

[26] Nikhil Yogesh Joshi. (2022). Implementing Automated Testing Frameworks in CI/CD Pipelines: Improving Code Quality and Reducing Time to Market. International Journal on Recent and Innovation Trends in Computing and Communication, 10(6), 106–113. Retrieved from

[27] Nikhil Yogesh Joshi. (2021). Enhancing Deployment Efficiency: A Case Study On Cloud Migration And Devops Integration For Legacy Systems. (2021). Journal Of Basic Science And Engineering, 18(1).

[28] Ronakkumar Bathani. (2022). Automation in Data Engineering: Implementing GitHub Actions for CI/CD in ETL Workflows. International Journal of Engineering and Management Research, 12(1), 149–155.

[29] Ronakkumar Bathani. (2021). Optimizing Etl Pipelines for Scalable Data Lakes in Healthcare Analytics. International Journal on Recent and Innovation Trends in Computing and Communication, 9(10), 17–24.

[30] Ronakkumar Bathani. (2021). Enabling Predictive Analytics in the Utilities: Power Generation and Consumption Forecasting. International Journal of Communication Networks and Information Security (IJCNIS), 13(1), 197–204.

