Advanced Federated Learning Strategies: A Multi-Model Approach for Distributed and Secure Environments

João Fonseca[1][0000-1111-2222-3333] and Second Author[2][1111-2222-3333-4444]

[1] Princeton University, Princeton NJ 08544, USA
[2] Springer Heidelberg, Tiergartenstr. 17, 69121 Heidelberg, Germany
lncs@springer.com

Abstract. Federated Learning (FL) emerges as a solution for training artificial intelligence (AI) models in distributed environments while preserving data privacy and security. This work presents a comprehensive review of the literature on FL, explainability techniques, FL frameworks, and strategies for integrating federated learning with explainability. Key gaps are identified, such as the lack of native support for explainability in widely used frameworks and challenges in applying techniques like Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) in federated systems. Innovative approaches are proposed, combining the use of inherently explainable models, the local application of explainability techniques on devices, and the aggregation of these explanations on the central server. Future experiments will assess the efficiency of these strategies, contributing to the adoption of ethical and transparent practices in critical sectors such as healthcare and the Internet of Things (IoT).

Keywords: Federated Learning, Explainability.

1 Introduction

The development of AI-based solutions has benefited various fields, including healthcare [1], communications [2], industry, finance [3], IoT networks [4], and urban sustainability [5]. However, the increasing use of sensitive data in training machine learning models raises challenges related to privacy, security, and explainability [1, 6, 7]. Federated Learning (FL) has emerged as a methodology that enables the collaborative training of models in distributed environments without centralizing data [8, 9], paving the way for the ethical and efficient use of AI in critical sectors such as healthcare, communications, and IoT, while ensuring data security and regulatory compliance [1, 6]. Beyond protecting sensitive data, FL allows for the integration of multiple models in a collaborative environment, making it crucial that these models are not only accurate but also transparent and interpretable. Explainability techniques are essential for increasing trust in AI systems, particularly in healthcare, where understanding the reasoning behind predictions is fundamental [10].

Federated Learning (FL) preserves data locally on originating devices, ensuring privacy and security in the training of machine learning models [7, 9], making it particularly relevant in sectors handling sensitive data, such as healthcare [1]. FL significantly reduces data transfer, improving latency and decreasing bandwidth consumption, which makes it well-suited for IoT devices and environments with limited connectivity [5, 8]. However, its implementation faces challenges such as data heterogeneity, security against adversarial attacks, and the need for model explainability in critical sectors [1, 6]. Model transparency is essential for the acceptance of AI systems in practical applications and for compliance with regulations such as the General Data Protection Regulation (GDPR) in the European Union [6]. In this context, integrating explainability techniques into FL presents an opportunity to combine privacy, security, and model interpretability, fostering ethical, effective, and reliable solutions across various sectors.

The inherent challenges of Federated Learning (FL) implementation highlight the need to develop strategies that enhance its applicability in real-world scenarios. The main goal is to optimize both the accuracy and transparency of predictions generated in distributed environments. The combination of classification models and explainability tools allows not only for more accurate predictions but also for interpretable decision-making processes, increasing model transparency [1, 3].

By operating in a federated context, where data remains locally stored on user devices, the proposed approach reinforces data privacy protection while simultaneously enhancing trust in AI-generated outcomes in sensitive sectors, such as healthcare [1], industry, and secure communications [3]. However, it is crucial to address security and system robustness concerns, particularly against adversarial attacks, to ensure the correctness and reliability of the training process [6, 8].

Finally, the validation of the proposed framework in virtualized and containerized environments demonstrates its scalability and efficiency, underscoring its potential to ensure a more ethical, transparent, and secure adoption of FL in real-world applications [11].

The primary objective of this study is to develop advanced strategies for Federated Learning (FL) that leverage multiple models while integrating data privacy, system security, and explainability in distributed environments. Additionally, the study emphasizes explainability as a fundamental element for increasing trust in AI systems. This will be achieved using explainability tools that identify key variables influencing model decisions and provide clear, interpretable insights into AI-generated outcomes.

2 Literature Review

2.1 Federated Learning

Federated Learning (FL) is a distributed machine learning paradigm that enables collaborative model training across multiple devices or servers without centralizing data. Introduced by Google in 2016, its initial application was in the Google Keyboard to

Comentado [FS1]: Aqui devemos detalhar estratégias de agregação de federated learning. Acho que esse isso seria importante. (ref: https://www.sciencedirect.com/science/article/pii/S2405844024141680)

Depois é também de relevo, referir que embora questões como privacidade encontram-se frequentemente associadas a FL a parte do XAI não.

Por fim, podemos ainda referir que time series pode acrescentar um desafio adicional

facilitate collaborative learning on Android smartphones [12]. FL preserves data privacy by keeping information locally stored, reducing the risk of breaches during transmission [13] and enhancing security against cyberattacks [3].

By bringing the code to the data, FL addresses concerns related to privacy, data ownership, and localization [7] while optimizing computational efficiency through distributed model training and updates. FL has the potential to revolutionize various fields, including healthcare, transportation, finance, and smart homes [12]. However, its implementation faces challenges such as scalability and data heterogeneity [14], requiring advances like the Federated Averaging (FedAvg) algorithm to optimize performance [3].

Indeed, model aggregation strategies play a crucial role in the collaborative training process in FL. Beyond FedAvg, several advanced aggregation strategies have been proposed specifically to address challenges related to data heterogeneity and communication efficiency. Relevant examples include algorithms like FedProx, FedAdam, FedYogi, FedAdagrad, and Scaffold [15]. FedProx, for instance, introduces a proximal term to handle heterogeneity, while FedAdam and FedYogi incorporate adaptive gradient methods to stabilize the learning process across diverse datasets [15].

In addition to technical aspects related to privacy and security, another fundamental yet often neglected factor is explainability (XAI). The integration of XAI techniques remains underexplored in the context of FL, yet it is crucial for providing transparency and trust in critical applications such as healthcare and finance.

Finally, it is important to consider that applying FL to time series data introduces additional complexities. In this scenario, specific challenges arise, such as temporal dependencies between data from different federated nodes, synchronization difficulties, and varying temporal patterns among clients. Exploring and overcoming these issues is crucial for fully leveraging the advantages of FL in dynamic and sequential data environments.

Thus, a thorough understanding of the fundamentals and challenges of FL is crucial for its successful practical implementation, ensuring significant benefits in terms of security, compliance, and operational efficiency for businesses and end-users.

2.2 Explainable AI

Comentado [FS2]: Falta ligar com FL..
Podemos falar de diferentes formas de ligar XAI. Aqui ainda falaremos de forma abstrata, na parte do trabalho detalhamos como construímos a nossa versão

algumas refs interessantes:
https://arxiv.org/pdf/2411.05874
https://www.sciencedirect.com/science/article/pii/S2352711023002017

The increasing complexity of machine learning models and the need for transparency have driven the relevance of Explainable AI (XAI). Although the terms "explainability" and "interpretability" are often used interchangeably, they represent distinct concepts: Interpretability refers to a human's ability to understand the reasoning behind a model's decision. Explainability describes how the model arrives at its output, revealing the underlying mechanisms. A model can be explainable without being directly interpretable [10]. XAI aims to make AI models transparent, mitigating the black-box problem, increasing trust, promoting responsible adoption, and aiding in bias detection. Ribeiro et al. [16] provide a systematic review of the most commonly used explainability methods in the literature, among which the following stand out:

- **LIME** (*Local Interpretable Model-Agnostic Explanations*), developed by Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin in 2016 [17].

This model-agnostic technique interprets predictions locally by perturbing input data. LIME has been widely used in medical applications, offering flexible and intuitive explanations. However, it faces computational limitations when applied to high-dimensional data.

- **SHAP** (*SHapley Additive exPlanations*), proposed by Scott Lundberg and Su-In Lee in 2017 [18]. SHAP is widely used for both local and global explanations based on cooperative game theory, assigning a Shapley value to each feature to assess its contribution to the model's predictions.

For effective explanations, models must adapt to end-user needs, using, for example, medical terminology or simplified representations for patients. There is often a trade-off between performance and explainability: Complex models (e.g., neural networks) offer higher accuracy but lower interpretability. Simpler models (e.g., decision trees) provide greater interpretability but at the cost of predictive performance. The choice of model depends on the application domain and the importance of explainability [19]. Key challenges in XAI include: Assessing explanation quality, Subjectivity in interpretation, Ensuring explainability in highly complex models. Ultimately, XAI is fundamental for the ethical and responsible use of AI, ensuring fairness, transparency, and accountability in decision-making processes.

The integration of XAI with Federated Learning (FL) has led to significant advancements in areas such as healthcare, finance, and IoT. Several conceptual pathways can be pursued, each addressing unique aspects of interpretability, transparency, and privacy preservation:

- **Federated Training of Interpretable-by-design Models**: Certain models, such as decision trees, linear regression, rule-based systems (RBSs), and linguistic fuzzy models, inherently provide transparency due to their structure, making them ideal for federated learning scenarios where interpretability is a priority [19–21]. Federating these models allows for balancing accuracy, interpretability, and privacy, exemplified by frameworks like OpenFL-XAI [22].
- **Local Explainability Techniques**: After training on local devices, techniques such as SHAP and LIME can be applied to provide detailed explanations tailored to local data characteristics, enhancing trust and transparency at the node level [2, 17, 18, 23].
- **Aggregation of Local Explanations**: Locally generated explanations can be aggregated centrally to create global insights. This approach generates generalized explanations while potentially sacrificing specific insights from individual nodes [2, 20, 23]. It is particularly useful in regulated sectors for monitoring, auditing, and understanding global model trends.
- **Hybrid Strategy for Local and Global Explanations**: This approach combines local explanations on devices with new explanations derived from the aggregated federated model at the central server, providing granular insights into local data and a broader evaluation of global model behavior. Such a strategy enhances transparency for local users and system auditors, making it advantageous in complex scenarios requiring detailed and consolidated analyses [2, 20, 23].

- **Challenges with Time Series in Federated XAI**: Dealing with time-series data introduces complexities such as synchronization of federated nodes and differing temporal patterns across clients. Specialized methods are needed to provide meaningful explanations in federated environments handling dynamic data [23].

The integration of FL and XAI thus presents unique opportunities but also poses significant methodological challenges. Abstractly exploring these interactions allows researchers and practitioners to design more robust, secure, and transparent AI systems suitable for critical applications.

2.3 Federated Learning Frameworks

Following the discussion on Federated Learning, a major challenge arises in selecting a framework that meets the specific needs of different application scenarios. The choice depends on multiple criteria, including scalability, privacy, interoperability, and support for explainability techniques, among others. This study evaluates the following FL frameworks: Flower, FedML, PySyft, TensorFlow Federated (TFF). The analysis considers key features and limitations, based on criteria such as functionality, ease of use, simulation capabilities, and interoperability, as discussed by Riedel et al. [24] and Elshair et al. [25]. The findings are summarized in Figure 1, which is a direct adaptation from Riedel et al. [24]. The figure was developed using detailed assessment criteria, including scalability, privacy, and interoperability, reflecting weighted scores assigned to each framework. The evaluation follows a quantitative methodology, where each criterion is assigned a specific weight, ensuring a balanced and adaptable analysis for distributed environments. For instance, factors such as support for heterogeneous devices and advanced algorithm availability were given high importance in the evaluation process.

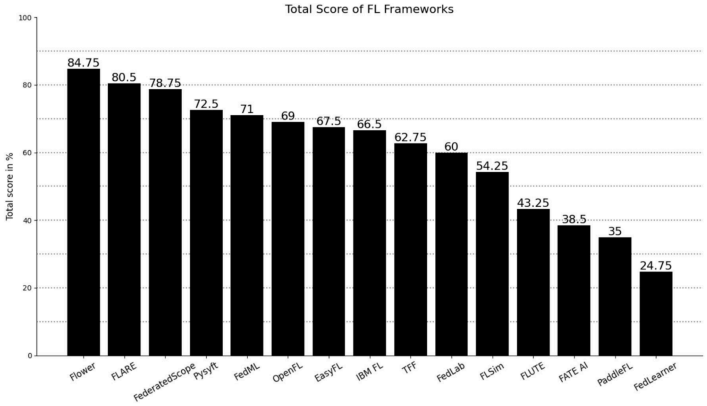

Figure 1- Total Scores (in Percentage) of Compared FL Frameworks [24]

The results highlight that Flower leads with a score of 84.75%, primarily due to its flexibility and support for heterogeneous environments [11]. Flower supports FL of inherently interpretable models, including rule-based systems, and has been effectively used for time-series applications, such as human activity recognition tasks utilizing DeepConvLSTM models on heterogeneous devices [11]. PySyft, while strong in privacy protection - offering techniques such as Differential Privacy - faces scalability limitations and lacks explicit documented support for explainability or specialized handling of time-series data in the reviewed literature [25]. TensorFlow Federated (TFF) proves to be efficient for research, but its integration is restricted to the TensorFlow ecosystem, with no explicit mention of native XAI or dedicated time-series support in the reviewed resources [11]. FedML excels in large-scale simulations, yet it lacks seamless integration with edge devices, and similar to PySyft and TFF, it does not offer documented native support for XAI or specialized functionality for time-series data according to available literature [11].

Notably, none of the analysed frameworks provide comprehensive native support for explainability, representing a significant gap [11]. This shortcoming necessitates external libraries or the adoption of inherently explainable models to enhance interpretability in federated learning applications. OpenFL-XAI, an extension of Intel's OpenFL, explicitly supports the federated learning of inherently interpretable models, such as fuzzy rule-based systems, indicating a direction other frameworks could potentially adopt for enhanced XAI support [22].

As a result, the choice of the optimal framework depends on the specific use case, requiring a careful balance between privacy, scalability, explainability integration, and specialized data handling such as time-series data.

Comentado [FS3]: Além disto devemos falar sobre se as frameworks estão preparadas para XAI, e para dados em séries temporais.
Estes são os casos de uso que mais nos interessam.

3 Strategies for Implementing Federated Learning and Explainability

The integration of Federated Learning (FL) with explainability techniques has led to significant advancements in areas such as healthcare, finance, and IoT. This study highlights widely discussed approaches in the literature, including inherently explainable models, local explainability techniques, and hybrid strategies that combine both local and global explanations.

Inherently explainable models, such as decision trees, linear regressions, or fuzzy rule-based systems, offer direct transparency in predictions, making them ideal for scenarios where interpretability is a priority [19–21].

After training on local devices, techniques such as SHAP and LIME can be applied to provide detailed explanations tailored to the specific characteristics of local data, enhancing trust and transparency [2, 17, 18].

At the central server, the locally generated explanations can be aggregated to create a global and interpretable view of the federated model. This approach is particularly useful in regulated sectors, where consolidating local explanations facilitates monitoring, auditing, and understanding global model trends [2, 20].

On the other hand, the hybrid strategy combines the application of local explanations on devices with the creation of new explanations based on the aggregated federated model at the central server. This allows for both granular insights into local data and a broader evaluation of global model behavior. Such an approach enhances transparency for both local users and system auditors, making it particularly advantageous in complex scenarios that require detailed and consolidated analyses [2, 20].

The approaches described will be experimentally explored in this study to determine which method is most efficient in terms of accuracy, interpretability, and real-world applicability. Figure 2 presents conceptual diagrams illustrating the implementation of FL with XAI, as described above.

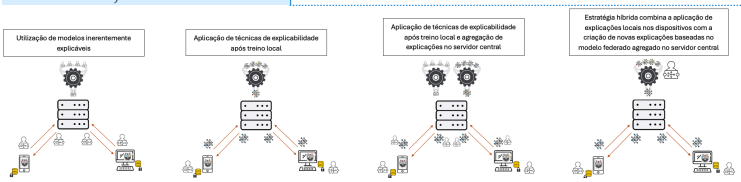

Figure 2 - Conceptual Schemes for the Implementation of FL with XAI

Comentado [JF4]: Com as alterações efetuadas no ponto 2.2, penso que este deixa de fazer sentido

4 Implementation and Evaluation of Federated Learning Strategies for Predictive Modeling

Comentado [FS5]: Falta uma arquitetura do sistema. Seria importante apresentar uma que modular, que permitisse experimentares vários conceitos e modelos…

A visualização local pode ser um componentes que podemos aplicar ao modelo simples e/ou a modelos XAI…

Depois devemos apresentar o exemplos da tua aplicação, não só de RMSE, mas também da parte de XAI…

Devemos escolher algumas imagens para os gráficos Shap e Lime.

Building on the discussion in the previous section, Strategies for Implementing Federated Learning and Explainability, two distinct approaches were explored for predicting real estate prices in California, both implemented using the Flower federated learning framework.

The analysis conducted so far has helped to understand the benefits and challenges of implementing interpretable models in a federated learning context. However, several questions remain open, particularly regarding how to improve model stability, reduce discrepancies between clients, and optimize the interpretability of predictions without compromising performance. In this section, we describe two distinct strategies — a Decision Tree-based model and a Linear Regression model with integrated explainability — and reflect on their effectiveness based on empirical evaluation. The analysis conducted helped to identify strengths, limitations, and points for improvement.

4.1 Decision Tree-Based Approach

The first implemented approach used a DecisionTreeRegressor from scikit-learn to predict real estate prices, leveraging its inherent interpretability and ease of analyzing the decision structure. The model was configured with a maximum depth of 10 and a fixed random state of 42 for reproducibility. Training was conducted over 5 federated rounds using two clients with independently preprocessed versions of the same dataset.

Each client computed the importance of features locally using the feature_importances_ attribute, which were then sent to the server for aggregation. These aggregated

importances were redistributed to all clients, influencing the reconstruction of the decision trees in the next round. The model thus evolved iteratively, prioritizing the most relevant features while maintaining privacy, as raw data never left the clients.

The system maintained detailed logs per round, tracking RMSE, tree structure, and feature importance. Performance varied slightly between clients, with RMSE values of 66,234.26 (Client 1) and 61,409.84 (Client 2). Structural similarity scores ranged from 0.95 to 0.97, indicating consistent convergence across rounds. The interpretability of this approach was further enhanced by visualization of tree structures and analysis of feature relevance.

While this setup demonstrated effective collaboration and explainability, discrepancies in performance due to data heterogeneity revealed the need for improved aggregation strategies and possibly more advanced ensemble models.

4.2 Linear Regression with Explainability Approach

The second approach employed a linear regression model implemented using PyTorch's nn.Linear module, optimized with Adam and using Mean Squared Error (MSE) as the loss function. The training process was structured and reproducible, relying on Docker containers to isolate environments and a centralized Flower server to coordinate model aggregation using FedAvg.

The training began with a client randomly selected to provide the initial model parameters. Across five rounds, the server distributed these parameters to the clients, who then trained locally, computed RMSE and loss, and generated explanations using SHAP and LIME. Updated weights were sent back to the server for aggregation.

The model demonstrated stable convergence, with RMSE values decreasing gradually from 235,232.42 in the first round to 235,231.31 in the final round, showing consistent improvement across all rounds. The progression was steady, with values of 235,232.14, 235,231.88, and 235,231.59 in rounds two, three, and four respectively, indicating a stable learning process.

To ensure comparability across rounds, SHAP explanations were generated using fixed random seeds, while LIME used a fixed instance selected at the start of training. This approach enabled precise tracking of how feature attributions evolved over time and clear identification of the most impactful features—median income and proximity to the ocean.

This approach succeeded in embedding interpretability into the federated pipeline but faced challenges regarding explanation robustness across clients and overall predictive accuracy in the presence of data heterogeneity.

4.3 Reflection and Future Directions

Both approaches showed relevant strengths: the decision tree model ensured direct interpretability and structural consistency, while the linear regression approach integrated post-hoc explainability techniques (SHAP and LIME) and demonstrated stable convergence. Table 1 summarizes their key differences and shared characteristics. More detailed configurations are described in Sections 4.1 and 4.2.

Feature	Decision Tree-Based Approach	Linear Regression with Explainability
Model	DecisionTreeRegressor (max_depth: 10, random_state: 42)	PyTorch nn.Linear with Adam Optimizer
Clients	2 (same dataset, independent preprocessing)	2 (partitioned dataset, containerized setup)
Data partitioning	Random split with seed 42 (80% train, 20% test, 10% validation)	Random split with seed 42 (80% train, 20% test, 10% validation)
Batch size	32	32
Rounds	5	5
Training strategy	Feature importance shared; trees rebuilt each round	FedAvg aggregation of model weights
Explainability	Tree structure + feature importance	SHAP (local, fixed seed), LIME (local, fixed instance)
Monitoring	Tree structure logs, JSON metrics, similarity index	Loss, RMSE, visualizations, reports
Rmse (client 1)	66,234.26	235,231.31
Rmse (client 2)	61,409.84	235,231.31
Structure convergence	Structural similarity: 0.95–0.97	RMSE stabilized from round 3 onward
Privacy	Data stays local, only feature importance shared	Data stays local, only model weights shared
Training time per round	~1.16 seconds	~6.73 seconds
Loss function	Not applicable (feature-based split)	Mean Squared Error (MSE)

Table 1: Summary of Key Differences Between Both Strategies

Future work should include testing hybrid models combining decision trees and linear regression, incorporating additional aggregation strategies, expanding to more clients and data sources, introducing differential privacy and encryption techniques, and quantifying the convergence of explanations — for example, assessing the stability of SHAP and LIME across rounds.

Both strategies confirmed the feasibility of interpretable federated learning and the value of explainability for trustworthy AI systems. However, reducing data heterogeneity impact and reinforcing robustness of interpretability remain open challenges.

ESTOU AQUI
Para já, estou a colocar aqui o texto do artigo do ESTG Masters traduzido para inglês pelo ChatGPT (falta colocar as referências corretamente)
Mais tarde coloco a imagem com os esquemas em inglês

Acknowledgments. A third level heading in 9-point font size at the end of the paper is used for general acknowledgments, for example: This study was funded by X (grant number Y).

Disclosure of Interests. It is now necessary to declare any competing interests or to specifically state that the authors have no competing interests. Please place the statement with a third level heading in 9-point font size beneath the (optional) acknowledgments, for example: The authors have no competing interests to declare that are relevant to the content of this article. Or: Author A has received research grants from Company W. Author B has received a speaker honorarium from Company X and owns stock in Company Y. Author C is a member of committee Z. [7]

References

1. Zhao, L., Xie, H., Zhong, L., Wang, Y.: Explainable federated learning scheme for secure healthcare data sharing. Health Inf Sci Syst. 12, 49 (2024). https://doi.org/10.1007/s13755-024-00306-6.
2. Kalakoti, R., Bahsi, H., Nõmm, S.: Explainable Federated Learning for Botnet Detection in IoT Networks. In: 2024 IEEE International Conference on Cyber Security and Resilience (CSR). pp. 01–08. IEEE, London, United Kingdom (2024). https://doi.org/10.1109/CSR61664.2024.10679348.
3. Dasari, S., Kaluri, R.: 2P3FL: A Novel Approach for Privacy Preserving in Financial Sectors Using Flower Federated Learning. CMES-Comp. Model. Eng. Sci. 140, 2035–2051 (2024). https://doi.org/10.32604/cmes.2024.049152.
4. Zahri, S., Bennouri, H., Chehri, A., Abdelmoniem, A.M.: Federated Learning for IoT Networks: Enhancing Efficiency and Privacy. In: 2023 IEEE 9th World Forum on Internet of Things (WF-IoT). pp. 1–6. IEEE, Aveiro, Portugal (2023). https://doi.org/10.1109/WF-IoT58464.2023.10539528.
5. Rahman, M.A., Hossain, M.S., Showail, A.J., Alrajeh, N.A., Alhamid, M.F.: A secure, private, and explainable IoHT framework to support sustainable health monitoring in a smart city. Sustainable Cities and Society. 72, 103083 (2021). https://doi.org/10.1016/j.scs.2021.103083.
6. Abdulrahman, S., Tout, H., Ould-Slimane, H., Mourad, A., Talhi, C., Guizani, M.: A Survey on Federated Learning: The Journey From Centralized to Distributed On-Site Learning and Beyond. IEEE Internet Things J. 8, 5476–5497 (2021). https://doi.org/10.1109/JIOT.2020.3030072.
7. Liu, J., Huang, J., Zhou, Y., Li, X., Ji, S., Xiong, H., Dou, D.: From Distributed Machine Learning to Federated Learning: A Survey. Knowl Inf Syst. 64, 885–917 (2022). https://doi.org/10.1007/s10115-022-01664-x.

8. Moulahi, W., Jdey, I., Moulahi, T., Alawida, M., Alabdulatif, A.: A blockchain-based federated learning mechanism for privacy preservation of healthcare IoT data. Computers in Biology and Medicine. 167, 107630 (2023). https://doi.org/10.1016/j.compbiomed.2023.107630.
9. Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., Gao, Y.: A survey on federated learning. Knowledge-Based Systems. 216, 106775 (2021). https://doi.org/10.1016/j.knosys.2021.106775.
10. Doshi-Velez, F., Kim, B.: Towards A Rigorous Science of Interpretable Machine Learning, http://arxiv.org/abs/1702.08608, (2017). https://doi.org/10.48550/arXiv.1702.08608.
11. Beutel, D.J., Topal, T., Mathur, A., Qiu, X., Fernandez-Marques, J., Gao, Y., Sani, L., Li, K.H., Parcollet, T., Gusmão, P.P.B. de, Lane, N.D.: Flower: A Friendly Federated Learning Research Framework, http://arxiv.org/abs/2007.14390, (2022). https://doi.org/10.48550/arXiv.2007.14390.
12. Mammen, P.M.: Federated Learning: Opportunities and Challenges, http://arxiv.org/abs/2101.05428, (2021).
13. Hasumi, M., Azumi, T.: Federated Learning Platform on Embedded Many-core Processor with Flower. In: 2024 IEEE 3rd Real-Time and Intelligent Edge Computing Workshop (RAGE). pp. 1–6. IEEE, Hong Kong, Hong Kong (2024). https://doi.org/10.1109/RAGE62451.2024.00015.
14. Borja, T., Alamillo, D., Anhari, A., Demirkol, I.: Scalable Federated Learning Simulations Using Virtual Client Engine in Flower. In: 2023 31st Signal Processing and Communications Applications Conference (SIU). pp. 1–4. IEEE, Istanbul, Turkiye (2023). https://doi.org/10.1109/SIU59756.2023.10223791.
15. Yurdem, B., Kuzlu, M., Gullu, M.K., Catak, F.O., Tabassum, M.: Federated learning: Overview, strategies, applications, tools and future directions. Heliyon. 10, e38137 (2024). https://doi.org/10.1016/j.heliyon.2024.e38137.
16. Ribeiro, J., Santos, R., Analide, C., Silva, F.: Implementing Federated Learning and Explainability Techniques in Regression Models to Increase Transparency and Reliability. SIC. 33, 15–24 (2024). https://doi.org/10.24846/v33i4y202402.
17. Ribeiro, M.T., Singh, S., Guestrin, C.: "Why Should I Trust You?": Explaining the Predictions of Any Classifier, http://arxiv.org/abs/1602.04938, (2016). https://doi.org/10.48550/arXiv.1602.04938.
18. Lundberg, S., Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, http://arxiv.org/abs/1705.07874, (2017). https://doi.org/10.48550/arXiv.1705.07874.
19. Molnar, C.: Interpretable Machine Learning. Lulu.com (2020).
20. Barcena, J.L.C., Ducange, P., Marcelloni, F., Renda, A.: Increasing trust in AI through privacy preservation and model explainability: Federated Learning of Fuzzy Regression Trees. Inf. Fusion. 113, 102598 (2025). https://doi.org/10.1016/j.inffus.2024.102598.
21. Rudin, C.: Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead, http://arxiv.org/abs/1811.10154, (2019). https://doi.org/10.48550/arXiv.1811.10154.
22. Daole, M., Schiavo, A., Corcuera Bárcena, J.L., Ducange, P., Marcelloni, F., Renda, A.: OpenFL-XAI: Federated learning of explainable artificial intelligence models in Python. SoftwareX. 23, 101505 (2023). https://doi.org/10.1016/j.softx.2023.101505.
23. Lopez-Ramos, L.M., Leiser, F., Rastogi, A., Hicks, S., Strümke, I., Madai, V.I., Budig, T., Sunyaev, A., Hilbert, A.: Interplay between Federated Learning and Explainable Artificial

Intelligence: a Scoping Review, http://arxiv.org/abs/2411.05874, (2024). https://doi.org/10.48550/arXiv.2411.05874.
24. Riedel, P., Schick, L., von Schwerin, R., Reichert, M., Schaudt, D., Hafner, A.: Comparative analysis of open-source federated learning frameworks - a literature-based survey and review. Int. J. Mach. Learn. Cybern. 15, 5257–5278 (2024). https://doi.org/10.1007/s13042-024-02234-z.
25. Elshair, I.M., Khanzada, T.J.S., Shahid, M.F., Siddiqui, S.: Evaluating Federated Learning Simulators: A Comparative Analysis of Horizontal and Vertical Approaches. Sensors. 24, 5149 (2024). https://doi.org/10.3390/s24165149.