A Review of the Most Well-Known Aggregation Algorithms for Federated Learning Applied to Large Language Models (LLMS)

Ygor Vieira[1*], Oberdan Rocha[1], Davidson Martins[1]
[1]SENAI CIMATEC University Center; Salvador, Bahia, Brazil

Research on federated learning has grown due to its ability to perform local training on distributed devices, especially in the context of artificial intelligence. However, there are still a few studies focused on the aggregation algorithms used in this type of learning, and even fewer addressing their application in large language models (LLMs). This article reviews the literature on federated learning with an emphasis on aggregation techniques applied to LLM training. A scarcity of specific studies was observed, along with the predominance of three algorithms: FedAvg, FedProx, and SCAFFOLD. Each was analyzed in terms of its strengths and weaknesses, including accuracy under data heterogeneity, convergence speed, and aspects of security and privacy. It is concluded that the future of aggregation algorithms in LLM training involves developing solutions that balance these aspects in an integrated manner.
Keywords: Federated Learning. Aggregation Algorithms. Large Language Models.

Federated Learning (FL) has gained significant popularity in research and real-world applications over the past few years [1]. This popularity stems from the fact that this type of machine learning is distributed; that is, participants train local models with local data, aggregating and sharing only the weights with a global model, unlike the traditional model, where all data had to be collected and stored on a central server for training [2]. This aggregation of weights between models is based on an aggregation algorithm, such as the well-known Federated Averaging (FedAvg) [1,3].

In the context of FL applied to large language models (LLMs), which involve challenges such as models with billions of parameters, data security and privacy, vast amounts of non-independent and identically distributed (non-IID) data, the choice of aggregation methods is crucial, as they directly influence the efficiency of distributed training and the convergence of the global model [3–7].

This article discusses methodologies involving the most widely used aggregation algorithms currently employed in LLM training. Another important aspect verified in this work is how FL associated with LLMs generates positive impacts on their training. The articles were compared, and it was noticeable that among the best-known aggregation algorithms, SCAFFOLD presented the best results, with strengths including the correction of statistical drifts, improved convergence in non-IID scenarios, and stable performance across multiple local iterations. Its weakness, however, was a higher communication cost, doubling the communication overhead compared to FedAvg [8–11].

Materials and Methods

An academic search was carried out with the assistance of the Consensus AI tool [12] (a peer-reviewed scientific article search engine that provides access to articles from major publishers such as Elsevier, IEEE, SciELO, Atena, among others).

The search parameters used were: Publishers such as Elsevier, IEEE, Atena, among others; Peer-reviewed scientific articles; Publication period (2019–2024).

To classify the articles, the number of citations was taken into consideration. Five main aspects were defined to meet the objective of this review: (a) the article needed to be directly related to

Received on 10 May 2025; revised 22 July 2025.
Address for correspondence: Ygor Vieira. SENAI CIMATEC University. Avenida Orlando Gomes, 1845. Zipcode: 41650-010. Salvador, Bahia, Brazil. E-mail: ygor.vieira@fbter.org.br. Original extended abstract presented at SAPCT 2025.

J Bioeng. Tech. Health 2025;8(4):393-395

Federated Learning; (b) the article had to discuss and explore aggregation methods; (c) the article needed to cite aggregation algorithms; and (d) the algorithms had to be related to Large Language Models (LLMs) and the challenges faced by Federated Learning.

Through the Consensus search, 60 articles were identified, tracked, mapped, and extracted. A subsequent review and contextual evaluation of these 60 articles was carried out to determine their relevance. Considering the four aspects defined above, 23 relevant articles were obtained.

Theoretical Framework

Of the 60 articles initially found, 23 were considered relevant, as they met the purpose of addressing three key research questions for this study:

Research Question 1: What are the most widely used Federated Learning algorithms for training LLMs?

Research Question 2: What are the strengths and weaknesses of the most widely used Federated Learning algorithms for training LLMs?

Research Question 3: What is the future of Federated Learning algorithms in the context of LLM training?

After this screening, the 23 selected articles were categorized based on the aggregation algorithms they addressed: FedAvg, FedProx, and SCAFFOLD. This categorization was structured into subsections, each presenting a brief introduction of the selected aggregation algorithm, along with an analysis and discussion of the article's contribution.

For Research Question 1, FedAvg, despite being a more straightforward and basic algorithm, is widely known and popularly used, including in LLM training. FedProx, in turn, remains widely used due to its effectiveness in handling data heterogeneity, even as more robust algorithms are being developed. SCAFFOLD, considering the LLM context, proved to be robust and effective, outperforming FedAvg and FedProx, particularly in domains that require specialized knowledge, such as finance and medicine.

For Research Question 2, subsections were created to present the strengths and weaknesses of FedAvg, FedProx, and SCAFFOLD.

Finally, based on the study carried out and the references analyzed, it was concluded that combining techniques to increase robustness in aggregation algorithms is necessary to mitigate the challenges faced by federated learning applied to LLMs. FedAvg shows sensitivity to non-IID data, resulting in slower convergence or suboptimal solutions, and its performance varies with the increase in the number of participating devices. FedProx incurs a higher computational cost compared to FedAvg, and in some cases, model accuracy may degrade when the number of local iterations is increased. SCAFFOLD has the highest computational cost among them, being approximately double that of FedAvg [9–12].

Furthermore, none of the three algorithms incorporates security techniques to reinforce data privacy in FL. Therefore, merging techniques to overcome these weaknesses, either by optimizing existing algorithms or developing new ones, represents the future of aggregation algorithms in FL.

Conclusion

Given the lack of comprehensive research on aggregation algorithms applied to LLM training, this article aimed to provide an analysis of the most widely used algorithms for this purpose.

The review identified that FedAvg, FedProx, and SCAFFOLD have been widely applied in several cases. However, the application and development of these algorithms still require more in-depth research, specifically for LLM training. This review contributes to the field by consolidating the most recent studies on aggregation algorithms applied to LLMs, offering a broad overview of current trends and areas requiring further attention.

Despite significant advances, important gaps remain, such as enhancing security and data privacy in federated learning aggregation without sacrificing model efficiency, as no aggregation algorithms applied to LLMs have been found to address this issue. Another gap concerns how to handle data heterogeneity while reducing computational resource demands.

Additionally, other aggregation algorithms, such as FedNova, have not yet been implemented or tested in LLM training and should also be addressed in future studies.

In summary, this review highlights the importance of aggregation algorithms in the field of LLMs and lays a solid foundation for future research, which may lead to significant advances.

Acknowledgements

This work was carried out with support from the São Paulo Research Foundation (FAPESP) - process n°. 2020/09770-7.

References

1. Gu X, Sabrina F, Fan Z, Sohail S. A review of privacy enhancement methods for federated learning in healthcare systems. Int J Environ Res Public Health. 2023;20(15):6539.
2. Aledhari M, Razzak R, Parizi R, Saeed F. Federated learning: a survey on enabling technologies, protocols, and applications. IEEE Access [Internet]. 2020 [cited 2025 Feb 10];8:1-1. Available from: https://doi.org/10.1109/ACCESS.2020.3013541.
3. Allan MS. Lab-09 - MO809 - Aprendizado de máquina e reconhecimento de padrões [Internet]. Instituto de Computação, UNICAMP; 2025 [cited 2025 Feb 10]. Available from: https://ic.unicamp.br/~allanms/mo809/labs/Lab-09/.
4. Neto HNC, Mattos DMF, Fernandes NC. Privacidade do usuário em aprendizado colaborativo: federated learning, da teoria à prática. In: Segurança em Redes de Computadores e Sistemas Distribuídos (Cap. 3). Anais do SBSEG 2020; 2020 [cited 2025 Feb 10]. Available from: https://sbseg.sbc.org.br/2020/capitulos/capitulo%203.pdf.
5. The Shift. Futuro do treinamento de LLMs é federado [Internet]. 2024 Mar 14 [cited 2025 Feb 10]. Available from: https://theshift.info/hot/futuro-do-treinamento-de-llms-e-federado/.
6. Santos LEB. Middleware baseado em aprendizado federado para aplicações de smart campus [Internet]. Instituto Federal de Educação, Ciência e Tecnologia da Paraíba; 2022 [cited 2025 Feb 10]. Available from: https://repositorio.ifpb.edu.br/handle/177683/4052
7. Autor desconhecido. Aprendizado federado: uma abordagem de aprendizado de máquina descentralizada para privacidade e segurança de dados [Internet]. ICHI.PRO; [cited 2025 Feb 10]. Available from: https://ichi.pro/pt/aprendizado-federado-uma-abordagem-de-aprendizado-de-maquina-descentralizada-para-privacidade-e-seguranca-de-dados-197608543778676
8. Gao Y. Federated learning: impact of different algorithms and models on prediction results based on fashion-MNIST data set. Appl Comput Eng. 2024;86:204-12.
9. Dai J. Comparative analysis of federated learning algorithms under non-IID data. Appl Comput Eng. 2024;86:91-100.
10. Li Q, Diao Y, Chen Q. Federated learning on non-IID data silos: an experimental study [Internet]. arXiv. 2021 [cited 2025 Feb 15]. Available from: https://arxiv.org/abs/2102.02079
11. Zheng W. Comparative analysis of federated learning algorithms under extreme non-IID conditions for computer vision. Appl Comput Eng. 2024;86:168-76.
12. Consensus. Consensus: AI-powered research search engine [Internet]. [cited 2025 Feb 10]. Available from: https://consensus.app/search/.