Federated Learning and AI Regulation in the European Union: Who is Responsible? – An Interdisciplinary Analysis

Anonymous Authors[1]

1. Introduction

With the introduction of the European Union Artificial Intelligence Act (AI Act) (Council of the European Union, 2021) and other international regulations being on the horizon, e.g., in the United States (The White House, 2023) and Canada (House Of Commons of Canada, 2022), everyone concerned with the development and deployment of AI has to adapt to new game rules. This entails data governance, robustness against adversarial scenarios, and energy considerations (Woisetschläger et al., 2024a). The AI Act puts the *service provider* into the spotlight, who has to assume responsibility for model development and deployment. Especially regarding data governance, the AI Act instantiates extensive rules for high-risk and general-purpose AI applications (GPAI, Art. 52) that cater to data privacy and system security. The majority of generative AI applications falls under the GPAI definition.

Federated Learning (FL) presents a privacy-enhancing and data-protecting machine learning technique (McMahan et al., 2017) that has recently received increased attention for enabling access to data siloes for generative AI applications (Woisetschläger et al., 2024b). In FL, a server operator provides an ML model sent to several clients and then trained on the clients' local data, which collaboratively train a global model via a central server, aggregating their local model updates. Private and secure computing techniques like Differential Privacy or Trusted Execution Environments help improve data privacy and system security (Bonawitz et al., 2017; Andrew et al., 2021). FL's data locality removes the key challenge of monitoring data lineage and simplifies accounting for user consent. Specifically, we study the FL workflow in alignment with related work (Li et al., 2020; Hard et al., 2018; McMahan et al., 2017) to touch up on the following:

Data Acquisition. The server operator can only employ a variety of client sampling strategies (Malinovsky et al., 2023; Wang & Ji, 2022; McMahan et al., 2017) for an FL training round, without the ability to directly investigate client data or process integrity.

Data Storage. Similarly, the clients decide how, where, and when to store data. This has implications on data availability, which directly touches upon the AI Act data governance requirements (Art. 10)[1].

Data Preprocessing. While the server operator can provide instructions on how to preprocess data so that the data is compatible with the ML model, the clients have the freedom to run additional preprocessing steps. Since the server operator has no direct data access, verifying data integrity before training is challenging. For FL applications, there are numerous approaches to improve data integrity (Sánchez Sánchez et al., 2024; Roy Chowdhury et al., 2022).

Model Aggregation. While acting as the FL training orchestrator, the server operator handles the model integrity control mechanism when aggregating model updates. Thus, FL appears to be a well-suited solution to open up data silos and provide additional data. This would significantly benefit the training or fine-tuning of generative models due to their sheer appetite for ever-increasing amounts of data (Zhou et al., 2023).

One can think that the server operator is automatically also the *service provider*. Yet, FL is a cooperative ML training technique where a central entity typically provides the ML model, and clients can decide when and what data is being used for training. As such, we see that the server (model) and the clients (data) control parts of the FL lifecycle, rendering them both legally responsible for their respective parts. Thus, this opens up the question:

> *Who is the service provider at what point in the FL workflow, and how can each party assume adequate responsibility?*

Our abstract studies technical and legal requirements that need to be established so that the FL server operator can assume responsibility as a service provider. This requires future technical work on auditability, verifiability, integrity,

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

[1]In the following, the term Art. refers to articles in the AI Act if not specified otherwise

and privacy. Further, we need to establish regulatory references for the terms and services of FL applications.

2. Technical Solutions Need to Focus on Transferring Responsibility to the Server Operator

For practical FL applications, technical solutions must be built with the server operator considered as service provider.

When establishing an FL system that could potentially entail thousands of clients at a time, managing responsibilities is likely to become a key challenge. Thus, we require solutions that suffice auditability, verifiability, and integrity & privacy.

Auditability & verifiability. There is a natural trade-off between privacy and data audits. The core paradigm of FL is to not share data beyond a client's area of control. Thus, we face a challenge when aiming to audit all steps that happen on a client device. For instance, a work by (Liu et al., 2023) uses Bayesian Nash equilibria and a market mechanism to incentivize truthful client behavior, i.e., submission of useful model updates. While this approach significantly reduces the risk of adversarial attacks, the requirements for auditing in the context of the AI Act are well-defined. Quintessentially, any data that is being captured, processed, and used in a training process must be evaluated for potential bias or adversarial information. To achieve this, numerous works combining FL with blockchain technology explore auditing the data processing steps and the training itself (Nguyen et al., 2021; Ma et al., 2020). What remains open is to develop solutions against data tampering.

Integrity & privacy. Particularly, we have to rethink the obligations of the provider concerning data integrity and protection (Art. 8–10), such that responsibility is transferred to the FL server. To account for the asymmetry of access/control-by-design in FL systems, we must develop data integrity measures that capture the nature of client data at the time of collection, while preprocessing the data, and immediately before starting the training process. Peer-based verification schemes of model updates are a promising direction to identify adversarial clients (Roy Chowdhury et al., 2022). Extending such schemes from client models to client data without infringing privacy would be interesting.

3. Regulatory Implementations Need to Foster Integrity and Verifiability

We need FL server operators to assume full responsibility; clients are technically and legally obligated to comply.

Service Provider. The GDPR (Council of the European Union, 2016) defines the term data controller under the GDPR. Complementary, the AI Act defines the Service Provider of AI systems. For data protection assessment when processing personal information at first, we need to clarify who is the data controller responsible and accountable for each distinct phase of the data processing and must demonstrate compliance with the requirements of the GDPR (Art. 5). The AI Act does not have a differentiated allocation of roles for separate processing phases and focuses on a central "provider" of a (secure) AI system, defined in Art. 3, with the obligations arising from Art. 8.

While both the FL server and clients could be considered providers under the AI Act, since the AI Act (unlike the GDPR) focuses less on responsibilities for individual, definable data processing phases and more on secure system design as a whole, the provider concept will have to be teleologically limited to the FL server. Thus, the server acts as the fully responsible service provider under the AI Act (Art. 8), especially concerning data governance (Art. 10) and General-Purpose AI service (Art. 52).

General Terms and Conditions for AI Systems. While in Art. 4 of GDPR, the controller is the person who, alone or jointly with others, decides on the purposes ("why") and means ("how") of the processing of personal data, the AI Act focuses on the (traditionally single and) central provider of an AI system. However, since clients are autonomously in control of their data while the server is in control of the model, we see an inconsistency between what is controllable by the service provider and what he is responsible for. To close this gap, we need a two-pronged approach – technical and legal ("how" & "why"). Responsibility in FL should depend on the server's physical, technical, and legal ability to influence decentral model training and configuration. FL servers and clients must have binding agreements to determine their respective GDPR and AI Act compliance responsibilities. The server operator as the service provider has to oblige clients to provide sufficient reporting compliant with the AI Act. This can be supported by cryptographic tools that minimize the need for trust among entities (Nguyen et al., 2021).

4. Conclusion

In this brief abstract, we study the FL lifecycle responsibilities under the AI Act. We find client-side responsibility for numerous steps, which practically limits the applicability of FL to open up additional data silos that would benefit the training of foundation models. Yet, there are promising directions that deserve increased attention such that a server operator can become the *service provider* without clients being required to assume extensive liability. With this, we drive the adoption of FL and help decrease data bias by directly relying on user data. Further clarifying the outlined service provider question directly responds to the EU AI Office's call for contributions to help implement the AI Act (Nature, 2024).

References

Andrew, G., Thakkar, O., McMahan, B., and Ramaswamy, S. Differentially private learning with adaptive clipping. *Advances in Neural Information Processing Systems*, 34: 17455–17466, 2021.

Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., and Seth, K. Practical Secure Aggregation for Privacy-Preserving Machine Learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, CCS '17. ACM, October 2017. doi: 10.1145/3133956.3133982. URL http://dx.doi.org/10.1145/3133956.3133982.

Council of the European Union. General data protection regulation (GDPR), apr 2016. URL https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679. Document 32016R0679.

Council of the European Union. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL - LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS, apr 2021. URL https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206. Document 52021PC0206.

Hard, A., Kiddon, C. M., Ramage, D., Beaufays, F., Eichner, H., Rao, K., Mathews, R., and Augenstein, S. Federated learning for mobile keyboard prediction, 2018. URL https://arxiv.org/abs/1811.03604.

House Of Commons of Canada. An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts, 6 2022. URL https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading.

Li, T., Sahu, A. K., Talwalkar, A., and Smith, V. Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 37(3):50–60, May 2020. ISSN 1558-0792. doi: 10.1109/msp.2020.2975749. URL http://dx.doi.org/10.1109/MSP.2020.2975749.

Liu, Y., Lou, R., and Wei, J. Auditing for federated learning: A model elicitation approach. In *The Fifth International Conference on Distributed Artificial Intelligence*, DAI '23. ACM, November 2023. doi: 10.1145/3627676.3627683. URL http://dx.doi.org/10.1145/3627676.3627683.

Ma, C., Li, J., Ding, M., Shi, L., Wang, T., Han, Z., and Poor, H. V. When federated learning meets blockchain: A new distributed learning paradigm, 2020. URL https://arxiv.org/abs/2009.09338.

Malinovsky, G., Horváth, S., Burlachenko, K. P., and Richtárik, P. Federated learning with regularized client participation. In *Federated Learning and Analytics in Practice: Algorithms, Systems, Applications, and Opportunities*, 2023. URL https://openreview.net/forum?id=6CDBpf7kNG.

McMahan, B., Moore, E., et al. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Singh, A. and Zhu, J. (eds.), *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pp. 1273–1282. PMLR, 20–22 Apr 2017. URL https://proceedings.mlr.press/v54/mcmahan17a.html.

Nature. There are holes in Europe's AI Act — and researchers can help to fill them. *Nature*, 625(7994): 216–216, January 2024. ISSN 1476-4687. doi: 10.1038/d41586-024-00029-4. URL http://dx.doi.org/10.1038/d41586-024-00029-4.

Nguyen, D. C., Ding, M., Pham, Q.-V., Pathirana, P. N., Le, L. B., Seneviratne, A., Li, J., Niyato, D., and Poor, H. V. Federated learning meets blockchain in edge computing: Opportunities and challenges. *IEEE Internet of Things Journal*, 8(16):12806–12825, August 2021. ISSN 2372-2541. doi: 10.1109/jiot.2021.3072611. URL http://dx.doi.org/10.1109/JIOT.2021.3072611.

Roy Chowdhury, A., Guo, C., Jha, S., and van der Maaten, L. Eiffel: Ensuring integrity for federated learning. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, CCS '22. ACM, November 2022. doi: 10.1145/3548606.3560611. URL http://dx.doi.org/10.1145/3548606.3560611.

Sánchez Sánchez, P. M., Huertas Celdrán, A., Xie, N., Bovet, G., Martínez Pérez, G., and Stiller, B. Federatedtrust: A solution for trustworthy federated learning. *Future Generation Computer Systems*, 152:83–98, March 2024. ISSN 0167-739X. doi: 10.1016/j.future.2023.10.013. URL http://dx.doi.org/10.1016/j.future.2023.10.013.

The White House. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 10 2023. URL https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

Wang, S. and Ji, M. A unified analysis of federated learning with arbitrary client participation. In Oh, A. H., Agarwal, A., Belgrave, D., and Cho, K. (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=qSs7C7c4G8D.

Woisetschläger, H., Erben, A., Marino, B., Wang, S., Lane, N. D., Mayer, R., and Jacobsen, H.-A. Federated learning priorities under the european union artificial intelligence act, 2024a. URL https://arxiv.org/abs/2402.05968.

Woisetschläger, H., Isenko, A., Wang, S., Mayer, R., and Jacobsen, H.-A. A survey on efficient federated learning methods for foundation model training, 2024b. URL https://arxiv.org/abs/2401.04472.

Zhou, C., Li, Q., Li, C., Yu, J., Liu, Y., Wang, G., Zhang, K., Ji, C., Yan, Q., He, L., Peng, H., Li, J., Wu, J., Liu, Z., Xie, P., Xiong, C., Pei, J., Yu, P. S., and Sun, L. A comprehensive survey on pretrained foundation models: A history from bert to chatgpt, 2023. URL https://arxiv.org/abs/2302.09419.