Federated Learning Priorities Under the European Union Artificial Intelligence Act

Anonymous Authors[1]

1. Introduction

On May 21, 2024, European Union Member States voted to endorse a landmark regulatory framework – the *EU Artificial Intelligence Act* (AI Act) (Council of the European Union, 2021; European Commission, 2023). This is likely the first of many legal frameworks that will affect how applications are developed, deployed, and maintained (The White House, 2023; House Of Commons of Canada, 2022; California Senate, 2024). In order to fit into this new landscape, all ML-based services must align with these regulatory requirements. Our main focus is on what this means for Federated Learning (FL) (Zhang et al., 2021), an approach to ML that offers better data privacy (Mothukuri et al., 2021) and access to siloed data than centralized ML. FL does this through distributed privacy-preserving learning of models between several clients and a server at scale (McMahan et al., 2017a; Tian et al., 2022), with the training data never leaving the clients and only the models being communicated.

Because FL enables access to a larger and more diverse data basis for training, it aligns with the AI Act requirement of mitigating data bias (Art. 10, Rec. 27).[2] This could lead to FL becoming more widely adopted if done right. Applications that are either characterized as «high risk» or «general-purpose AI» per Art. 52 (which most generative AI models are a part of) require trustworthy, reliable, and highly secure models. For end users, this means model responses are truthful, factual, and safe. Under current circumstances, this is a very challenging objective to achieve. More diverse data could get us significantly closer to this objective and help build more versatile AI applications (Bommasani et al., 2021). However, to achieve the goal of broadening the access to data for pre-training and fine-tuning of generative models, we must focus on technical challenges that, if overcome, can catalyze FL's adoption as a mainstream training paradigm for AI models under the AI Act:

- **Data governance** (Art. 10). For FL applications, data provenance, bias mitigation, and strict privacy are challenging to achieve. Currently, available techniques only address parts of the regulatory requirements. We discuss details in Section 2.
- **Privacy** (Art. 10). As data privacy is a priority within the EU (cf. GDPR), the AI Act expands on the importance of data protection. We find the regulatory definition of privacy significantly deviates from private or secure computing techniques such as (ϵ, δ)-Differential Privacy. We discuss this in Section 3.
- **Energy efficiency** (Art. 69). As part of the AI Act, anyone involved as a *service provider* in the training or offering of AI services must install holistic energy consumption tracking and implement energy-efficient practices. We outline the priorities in Section 4.

Our work bridges the gap between regulatory requirements and the technical challenges that arise with the introduction of the AI Act. In the following, we discuss selected research priorities under the AI Act (highlighted with *italics*) and subsequent research questions.

2. The Data Quality Requirements are Currently not Amenable to FL

We need to find solutions to meet the data governance requirements of the AI Act under Art. 10 without having direct access to the data.

The AI Act requires ML service providers to implement strict data governance principles that tackle data bias, privacy warranting GDPR (Rec. 10), and security (Art. 15). While data security has largely driven FL's use of private computing techniques (Jin et al., 2023; McMahan et al., 2017b; Bonawitz et al., 2017), data bias mitigation and regulatory privacy have played a comparatively minor role in research up to this point. Our preliminary results show that private computing techniques not only exhibit significant scalability challenges but can also increase the energy consumption of an FL system by up to 2 orders of magnitude. This instantiates a privacy-energy trade-off that requires the immediate attention of the FL community. If not addressed, FL will fall short of centralized FL regarding the legal priority for sustainable and energy-efficient computing.

[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.

[2]Throughout the paper, we cite an article from the AI Act document as "Art. XY" or a recital as "Rec.".

Additionally, the AI Act requires service providers to evaluate their models (and training data) for data bias (Art. 7, Art. 10). Since FL does not allow for direct data access on clients, the only viable option to test models for potential data bias is via domain-specific reference datasets and extensive validation on the server side. As FL is a data-parallel paradigm, this means all clients are sitting idle while the server validates the model on its reference data. At the same time, the AI Act requires a service provider who engages in high-risk applications or general-purpose AI (Art. 52) to monitor the total energy consumption in a system. This includes idle clients, which can be up to several thousand devices (Hard et al., 2018), resulting in a significant energy draw, all while doing validation on the server. Thus, if data quality can be indirectly inferred through techniques with heavy energy investment, how do they compare to techniques that require direct data access? As such, to make meaningful progress towards the goals of the AI Act, the FL research community must focus on improving data quality management techniques, including energy consumption considerations, and finding interdisciplinary consent on data processing responsibilities.

3. Closing the Gap Between Technical and Regulatory Privacy

There is a discrepancy between the notions of regulatory and technical privacy as well as a privacy-energy trade-off.

The AI Act requires service providers to strictly implement GDPR (Art. 10.2). This entails three things: (I) consent-based data processing, (II) the right to information, and (III) the right to be forgotten.

FL significantly contributes to consent-based data processing by moving computational workloads into the control space of clients, who can decide what data to provide for the training process. This removes a key liability vector that has been a major cause for GDPR-related fines in the past (CMS Law, 2024). Similarly, as data always remains on clients, the simplified data lineage, as compared to centralized training, can help to facilitate the right to information describing what and where data has been used during training. While implementing the right to be forgotten via model unlearning is technically possible in FL applications (Halimi et al., 2022), there is an ongoing discussion about whether a global model comprised of client model updates represents personal data.

GDPR also requires us to prevent unauthorized data access. This is the part referred to as *technical privacy*, i.e., the data leakage risk. The prevention of unauthorized data access can be facilitated by differential privacy (DP) (McMahan et al., 2017b) or by two cryptographic methods, homomorphic encryption (HEC) (Jin et al., 2023) and secure multi-party computation (SMPC) (Bonawitz et al., 2017). Our preliminary experiments show that the cryptographic methods come with exponential computational and communication algorithmic complexities, which lead to significant overheads, increased energy consumption, and ultimately limited scalability. Conversely, DP comes with very little computational and communication overhead. Yet, the perturbation mechanism requires smaller learning rates, extending the training time by up to 2 orders of magnitude. This intensifies the previously introduced privacy-energy trade-off (cf. Section 2).

4. CO_2-based Optimization to Compete with Centralized Training

FL is currently not achieving competitive energy efficiency compared to centralized training.

While energy reporting is uncommon for ML applications, the AI Act requires service providers of high-risk or generative AI solutions to implement holistic reporting on key performance indicators, including energy consumption. For FL systems, this could entail the monitoring of computational and communication energy consumption. While computational energy can be measured on clients, communication energy is challenging to measure. The best model currently available is the per-bit energy consumption model (Yousefpour et al., 2023; Salh et al., 2023; Kim et al., 2023; Albelaihi et al., 2022). Preliminary experiments on fine-tuning a BERT model for text classification (News20 dataset) with 100 clients and a participation rate of 10% over 2000 training steps yield a total energy consumption of 0.56 kWh, of which 0.3 kWh are computation, 0.17 kWh client idle times, and 0.09 kWh communication. For context, a centralized application for the exact same workload consumes 0.09 kWh or 16% of the total FL energy consumption.

Additionally, it is key to consider the energy-efficient design and ongoing system optimization efforts in data centers for ML applications (Castro, 2024), which we currently do not exhibit in FL research. To comply with the AI Act, we need to foster research towards finding consent on energy monitoring standards and improving the overall energy efficiency of FL applications.

5. Conclusion

With this brief abstract, we set forth the key priorities for the FL research community to become compliant with the AI Act. Our analysis touches upon the three pillars of the new regulation: data governance, privacy, and energy efficiency. The EU AI Office opened a call for contributions from research and practice to help shape the implementation of the AI Act in a joint effort (Nature, 2024). Thus, the time for raising our voices and making a difference in developing technical baselines for regulatory compliance is *now*.

References

Albelaihi, R., Yu, L., Craft, W. D., Sun, X., Wang, C., and Gazda, R. Green Federated Learning via Energy-Aware Client Selection. In *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, pp. 13–18, 2022. doi: 10.1109/GLOBECOM48099.2022.10001569.

Bommasani, R., Hudson, D. A., et al. On the opportunities and risks of foundation models, 2021. URL https://arxiv.org/abs/2108.07258.

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.

California Senate. Safe and Secure Innovation for Frontier Artificial Intelligence Systems Act (SB 1047), February 2024. URL https://leginfo.legislature.ca.gov/faces/billStatusClient.xhtml?bill_id=202320240SB1047.

Castro, D. Rethinking concerns about ai's energy use, Jan 2024. URL https://www2.datainnovation.org/2024-ai-energy-use.pdf.

CMS Law. GDPR Enforcement Tracker, 01 2024. URL https://www.enforcementtracker.com/.

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.

European Commission. A European approach to artificial intelligence, 2023. URL https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence.

Halimi, A., Kadhe, S., Rawat, A., and Baracaldo, N. Federated Unlearning: How to Efficiently Erase a Client in FL?, 2022. URL https://arxiv.org/abs/2207.05521.

Hard, A., Rao, K., Mathews, R., Ramaswamy, S., Beaufays, F., Augenstein, S., Eichner, H., Kiddon, C., and Ramage, D. 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, June 2022. URL https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading.

Jin, W., Yao, Y., Han, S., Joe-Wong, C., Ravi, S., Avestimehr, S., and He, C. FedML-HE: An Efficient Homomorphic-Encryption-Based Privacy-Preserving Federated Learning System, 2023. URL https://arxiv.org/abs/2303.10837.

Kim, M., Saad, W., Mozaffari, M., and Debbah, M. Green, Quantized Federated Learning over Wireless Networks: An Energy-Efficient Design. *IEEE Transactions on Wireless Communications*, pp. 1–1, 2023. doi: 10.1109/TWC.2023.3289177.

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 2017a. URL https://proceedings.mlr.press/v54/mcmahan17a.html.

McMahan, H. B., Ramage, D., Talwar, K., and Zhang, L. Learning Differentially Private Recurrent Language Models, 2017b. URL https://arxiv.org/abs/1710.06963.

Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., and Srivastava, G. A survey on security and privacy of federated learning. *Future Generation Computer Systems*, 115:619–640, 2021.

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.

Salh, A., Ngah, R., Audah, L., Kim, K. S., Abdullah, Q., Al-Moliki, Y. M., Aljaloud, K. A., and Talib, H. N. Energy-Efficient Federated Learning With Resource Allocation for Green IoT Edge Intelligence in B5G. *IEEE Access*, 11:16353–16367, 2023. doi: 10.1109/ACCESS.2023.3244099.

The White House. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, October 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/.

Tian, Y., Wan, Y., et al. FedBERT: When Federated Learning Meets Pre-training. *ACM Transactions on Intelligent Systems and Technology*, 13(4):1–26, August 2022. ISSN 2157-6912. doi: 10.1145/3510033. URL http://dx.doi.org/10.1145/3510033.

Yousefpour, A., Guo, S., Shenoy, A., Ghosh, S., Stock, P., Maeng, K., Krüger, S.-W., Rabbat, M., Wu, C.-J., and Mironov, I. Green Federated Learning, 2023. URL https://arxiv.org/abs/2303.14604.

Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., and Gao, Y. A survey on federated learning. *Knowledge-Based Systems*, 216:106775, March 2021. ISSN 0950-7051. doi: 10.1016/j.knosys.2021.106775. URL http://dx.doi.org/10.1016/j.knosys.2021.106775.