Federated Learning Priorities Under the European Union Artificial Intelligence Act

Herbert Woisetschläger [1] Alexander Erben [1] Bill Marino [2] Shiqiang Wang [3]
Nicholas D. Lane [2 4] Ruben Mayer [5] Hans-Arno Jacobsen [6]

Abstract

The age of AI regulation is upon us, with the *European Union Artificial Intelligence Act* (AI Act) leading the way. Our key inquiry is how this will affect *Federated Learning* (FL), whose starting point of prioritizing data privacy while performing ML fundamentally differs from that of centralized learning. We believe the AI Act and future regulations could be the missing catalyst that pushes FL toward mainstream adoption. However, this can only occur if the FL community reprioritizes its research focus. In our position paper, we perform a first-of-its-kind interdisciplinary analysis (legal and ML) of the impact the AI Act may have on FL and make a series of observations supporting our primary position through quantitative and qualitative analysis. We explore data governance issues and the concern for privacy. We establish new challenges regarding performance and energy efficiency within lifecycle monitoring. Taken together, our analysis suggests there is a sizable opportunity for FL to become a crucial component of AI Act-compliant ML systems and for the new regulation to drive the adoption of FL techniques in general. Most noteworthy are the opportunities to defend against data bias and enhance private and secure computation.

1. Introduction

On December 8$^{\text{th}}$, 2023, the European Union (EU) Commission and Parliament found a political agreement on an unprecedented regulatory framework – the *EU Artificial Intelligence Act* (AI Act) (Council of the European Union, 2021; European Commission, 2023b). This is the first, but likely one of many regulations that will affect how ML applications are developed, deployed, and maintained. In order to comply with this new landscape, ML of all kinds will likely need to undergo significant changes. Our main focus is on what this means for Federated Learning (FL) (Zhang et al., 2021), a fundamentally different approach to ML that offers unique benefits, such as privacy (Mothukuri et al., 2021) and access to siloed data, compared to its more centralized counterpart. FL enables distributed privacy-preserving learning of models between several clients and a server at scale (McMahan et al., 2017a; Tian et al., 2022) while the training data never leaves the clients, and only the models are communicated. We believe that the AI Act and subsequent regulations could serve as the catalyst to pushing FL towards mainstream adoption. However, this will require the FL community to shift some of its research priorities.

In this position paper, we perform a first-of-its-kind interdisciplinary analysis (legal and ML) of the AI Act and FL (Section 2). Based on our methodology that aligns with the priorities set out by the AI Act (Section 3), we make several key observations in support of our primary position (Section 4):

First, FL struggles to cope with the new performance trade-offs highlighted in the AI Act. As a result, there is a need for a reconsideration of FL research priorities to address these issues, particularly in terms of energy efficiency and the computational costs of privacy. While governance has been a focus for FL in the past, the AI Act brings new challenges, such as performance parity with centralized approaches and lifecycle monitoring under privacy-preserving operations.

Second, FL has inherent advantages over centralized approaches with respect to data lineage and the ability to address bias and related concerns through access to siloed data. However, there are remaining technical hurdles for data man-

[1] School of Computation, Information and Technology, Technical University of Munich, Germany [2] Department of Computer Science and Technology, University of Cambridge, United Kingdom [3] IBM T.J. Watson Research Center, United States [4] Flower Labs, Germany [5] Department of Computer Science, University of Bayreuth, Germany [6] Department of Electrical and Computer Engineering, University of Toronto, Canada. Correspondence to: Herbert Woisetschläger <herbert.woisetschlaeger@tum.de>, Alexander Erben <alex.isenko@tum.de>, Bill Marino <wlm27@cam.ac.uk>, Shiqiang Wang <wangshiq@us.ibm.com>, Nicholas D. Lane <ndl32@cam.ac.uk>, Ruben Mayer <ruben.mayer@uni-bayreuth.de>, Hans-Arno Jacobsen <jacobsen@eecg.toronto.edu>.

This paper should not be treated as legal advice.

agement and governance issues. At the same time, these technical hurdles have been solved in centralized learning due to its lack of concern for data movement and its effects on privacy. It is currently unclear how to cope with GDPR at scale and how the right to privacy will be expressed in practice.

Our analysis indicates due to AI regulation that FL has a significant opportunity to become even more widely adopted. If the FL community can redirect their research efforts to address the new priorities highlighted by the AI Act, and combine this with the inherent advantages of FL, it could become the go-to approach for building compliant ML systems. Therefore, we advocate for a large fraction of the energy that will undoubtedly go into revising all forms of ML to align with the societal values encoded in this act to be directed into FL rather than centralized approaches. This will lead to us more quickly having access to suitable methods for deployment in this new landscape, and we expect the act to be a new driver (along with the long-standing issue of pure privacy) towards the adoption of FL techniques in general.

Our contributions:

- **Requirement analysis for FL based on the AI Act**. We examine the impact of the AI Act on FL systems and methods, outlining requirements and linking them to challenges in FL, aiming to align the legal and ML perspectives.
- **Quantitative and qualitative analysis of FL under the AI Act**. We quantify the costs associated with FL, identify the current inefficiencies, and discuss the potential energy implications. Through our experiments, we introduce the privacy-energy trade-off that arises when fine-tuning a large model in a practical FL framework while aiming to be compliant with the AI Act. Further, we provide a qualitative understanding of the potential of FL under the AI Act.
- **Future outlook on novel research priorities for the FL community.** By distilling our results into a list of future research priorities, we aim to provide guidance such that FL can become the go-to choice for applications incorporating governing EU fundamental rights.

2. The EU Artificial Intelligence Act

The AI Act's latest draft as of January 23rd, 2024 is referenced throughout this section (European Parliament and Council, 2024). This first-of-its-kind, comprehensive, legal framework around AI development and application aims "[...] *to promote* [...] *trustworthy artificial intelligence while ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter, including* [...] *environmental protection* [...]" (Rec. 1)[1]. While it is not finalized yet and must be implemented as national law in every EU country, it may set the basis for other non-EU jurisdictions to decide their legislation (The White House, 2023; House Of Commons of Canada, 2022). The penalties for violations of the obligations outlined in the AI Act are currently set at a maximum of €35M or 7% of the company's worldwide annual turnover, whichever is higher (Art. 71.1). As such, the fines range in similar dimensions as those of the General Data Protection Regulation (Regulation (EU) 2016/679) ("GDPR") Art. 83.5.

The AI Act differentiates in its classification of AI applications within two dimensions: risk-based (Art. 6) and general-purpose AI models (GPAI) (Art. 52). We specifically cover the risk-based classification and the associated requirements for high-risk systems (Art. 8). If an application falls under this "high-risk" category, it must follow strict robustness and cybersecurity (Art. 15) and data governance guidelines (Art. 10), including compliance with GDPR. Additionally, high-risk system providers may soon have to follow energy-efficiency standards once those are finalized by EU standardization entities (Art. 40.2). As it happens, most applications that benefit from federated aspects fall under this category by default, such as medical applications (Pfitzner et al., 2021) or management of critical infrastructure (electricity, water, gas, heating, or road traffic) (Wang et al., 2021; El Hanjri et al., 2023; Tun et al., 2021; Liu et al., 2020).

The root cause of most GDPR infringements is data collection and unlawful processing (CMS Law, 2024). The AI Act recognizes this fact and emphasizes the importance of the GDPR in its legal text, naming "*data protection by design and default*" and "[...] *ensuring compliance* [...] *may include* [...] *the use of technology that permits algorithms to be brought to the data* [...] *without the transmission between parties*" (Rec. 45a). This aligns with the Act's broad insistence that "*right to privacy and to protection of personal data* [...] *be guaranteed throughout the entire lifecycle of the AI system*" (Rec. 45a). Since FL specifically addresses these privacy concerns and removes data movement and direct access by definition, we must now understand how we can leverage the introduction of the AI Act to enable its legal compliance. For FL, the following three aspects of the AI Act are relevant to understand.

Data Governance. The biggest hurdle that the AI Act imposes on high-risk FL applications is data governance, which requires strong oversight of the data that is being used for the entire model lifecycle of development, training, and deployment (Art. 10.2). The practices shall include an "*examination in view of possible biases that are likely to*

[1] We explain the difference between an article (Art.) and recital (Rec.) in Appendix A. When not specified otherwise, Rec. and Art. refer to the EU AI Act.

affect the health and safety of persons [...]" and "*appropriate measures to detect, prevent and mitigate possible biases*" (Art. 10.2f,fa). With these requirements, we can foresee a future where data access is necessary to comply with forthcoming rules. However, this data access is a reason why data providers might be hesitant to participate, as it is currently unclear how privacy preservation will be enacted and if they might be liable under GDPR.

Federated Learning provides another outlook on this issue. While the training data is not accessible by design with FL and, thus, cannot easily comply with the requirements under Article 10.2, it can ease the angst of data providers as data is processed on a strict "need-to-know" basis and will never be moved from the source. This can be a more promising path forward to create access to data in a privacy-preserving manner, simply due to the number of participants. Additionally, FL includes an emerging research field that implements different techniques to reach specific privacy guarantees, which we cover in Section 4.1. While data quality and techniques to detect biases are recognized as having a high priority when developing DL applications (Whang et al., 2023), FL has to close this gap with indirect techniques to comply with Article 10. It is up to debate if techniques, e.g., that combat non-IID data in a federated setting (Zhao et al., 2018), provide adequate robustness guarantees or if additional safeguards will be needed.

As high-risk applications typically involve personal data and are required to conform to the GDPR Article 10, we take a closer look at how FL is meeting the key requirements of the GDPR:

Security while processing data. GDPR Art. 32.2 calls for strict security guidelines when processing data: "[...] *the appropriate level of security account shall be taken* [...] *that are presented by processing, in particular* [...] *unauthorized disclosure of, or access to personal data transmitted, stored or otherwise processed*". While minimizing the risk of data leakage without any data movement, FL shares the model updates during training, providing an attack vector. To combat this, threat models and security measures for misuse of data by gradient inversion or membership inference attacks have been explored thoroughly (Zhang et al., 2023; Huang et al., 2021; Geiping et al., 2020). FL is also vulnerable to data poisoning attacks whereby attackers corrupt client-side data in an attempt to sabotage the model, which is being combated by comprehensive benchmarking (Han et al., 2023; Zhao et al., 2023). Nevertheless, research on FL security remains a key task, as new attacks could emerge.

The right to information. While access to data is minimized in FL by only sharing model updates, the GDPR reserves the right for individuals to request all information a service provider has stored (GDPR Art. 15, GDPR Rec. 63 & 64). This also includes how data has been used for learning models, which is already being evaluated as client participation is a key priority in FL systems. Existing studies on personalized FL have established accuracy variance and client update norm as metrics to evaluate the value add a client generates for an FL system (Tan et al., 2023; Chen et al., 2022; Fallah et al., 2020).

The right of clients to revoke their consent at any time. With the AI Act installing GDPR as the adjacent privacy regulation, clients in FL systems may make a request to delete their data or revoke their consent to use it at any time (AI Act Art. 17; Art. 7). This can lead to two consequences: the removal of any user data stored in the FL system and, depending on interpretations, the need to unlearn the client's training progress from the global model. Removing the data is trivial, as the data lineage guarantees provided by FL guard the data from being moved from the clients. There are a few approaches to machine unlearning (Xu et al., 2023), such as the teacher-student framework (Kurmanji et al., 2023) or amnesic unlearning (Graves et al., 2021). However, both techniques need access to the training data or even the entire training progress with client-level model snapshots that are usually unavailable in a federated setting. In the specific case of FL, existing works focus on unlearning entire clients and provide a possibility for GDPR compliance (Halimi et al., 2022) without direct data access.

Energy Efficiency. While we focus on high-risk applications, the AI Act also promotes the environmentally sustainable development of AI systems regardless of the application. A voluntary Code of Conduct (CoC) will be drawn up to create clear objectives and key performance indicators (Art. 69) to help set best practices regarding, among others, energy efficiency. It is still up to discussion which high-risk requirements will be included in this CoC, but it is clear that the position of the AI Act reflects a fundamental value of the EU, namely, sustainability. While state-of-the-art data centers are designed to be energy-efficient and capable of running on mostly regenerative energy (Google, 2023), edge clients used in FL are powered by the average energy mix at their locations (Yousefpour et al., 2023; Ritchie & Rosado, 2020). This is echoed by the current trends, which indicate that specialized edge devices can compete with data-center GPUs on sample efficiency (sample-per-Watt) (Woisetschläger et al., 2023), but only when looked at the raw throughput, and not in time-to-accuracy comparing FL to centralized training (cf. Section 4). As such, we find a natural trade-off between energy efficiency and privacy that has yet to be quantified (cf. Section 4.2). Although we see promising progress toward quantifying how and where energy is being consumed in FL applications (Mehboob et al., 2023; Qiu et al., 2023), there are still many fundamental open challenges. For instance, we need to find consensus on how the energy-cost responsibility is being assigned in FL with devices not owned by the training provider and how it will compare to future energy-consumption baselines.

Robustness and Quality Management. Unsurprisingly, high-risk AI systems should have an "*appropriate level of accuracy, robustness* [...] *and perform consistently*" (Art. 15.1), and this should be guaranteed by a quality management system that takes "[...] *systematic actions to be used for the development, quality control, and quality assurance*" (Art. 17.1c). While it is in the interest of the AI providers to guarantee specific performance goals when deploying, the development of an AI system could be severely prolonged by the need for training to be as energy-efficient as possible. One technique to guarantee model robustness is early stopping to avoid overfitting, which tracks the model performance on a validation dataset (Prechelt, 2002). From the earlier data governance requirements on representative data, the time to validate a model may increase as the validation dataset becomes large (cf. Section 4.3). Combining frequent validation with the need for energy consumption monitoring poses a new optimization problem: Is it more energy-efficient to keep clients idle while a subset validates, or should the next round start in parallel with a chance of overfitting and wasting the energy? As FL stands currently, this shifts the focus towards techniques that increase the validation efficiency per data sample, e.g., as done in dataset distillation (Lei & Tao, 2023). As we anticipate a trade-off between energy efficiency and quality management, this could lead to an increased performance gap between centralized learning and FL.

3. Methodology

Our analysis in Section 2 has highlighted a series of core challenges pertaining to data governance without direct data access, energy efficiency, robustness, and overall quality management. This section presents our evaluation criteria and how they align with the AI Act. We also introduce the methodologies for our qualitative and quantitative analysis.

3.1. Evaluation criteria

Data Governance. Data governance in the AI Act focuses on data bias reduction and strict enforcement of regulatory privacy. Our qualitative analysis focuses on identifying the potential of FL to mitigate data bias. Therefore, we study the effect FL can have on the availability of data such that a broader data basis becomes available for training. A broader and potentially continuously evolving training dataset could improve the generalization capability of a model and better account for minority groups (Torralba & Efros, 2011). For privacy, we look into the technical capabilities of private and secure computing currently available to FL applications. We study whether there is a gap between state-of-the-art technical privacy methods and the regulatory privacy requirements introduced by the AI Act and GDPR.

Energy Efficiency. In centralized DL, we often fine-tune FMs on servers with multiple GPUs and, thus, require very high bandwidth interconnects ($> 200 \text{GB/s}$) between the GPUs either via NVLink or Infiniband (Li et al., 2020a; Appelhans & Walkup, 2017). FL only requires low bandwidth interconnects ($< 1 \text{GB/s}$) since communication happens sparingly compared to multi-GPU centralized learning (Xu & Wang, 2021). This creates major design differences in the training process and an entirely different cost model. In the following, we point out essential components of the cost model for FL.

The AI Act indicates that further guidelines around energy efficiency are forthcoming. When it comes to how those guidelines define and measure energy efficiency, we propose using a holistic methodology that accounts for computation and communication. Based on such conservative methodology, we can develop comprehensive baselines to compare against. The total energy consumption P consists of two major components, computational P_c and communication energy P_t, i.e., $P = P_c + P_t$.

P_c can be measured directly on the clients via the real-time power draw with an on-board energy metering module (Beutel et al., 2020) or deriving the energy consumption based on floating point operations and a client's system specifications (Desislavov et al., 2023). At the same time, P_t is generally more challenging to measure as multiple network hops are involved. Often, the network infrastructure components, such as switches and routers, are owned by multiple parties and are impossible to monitor for a service provider. However, the bit-wide energy consumption model is available to calculate the cost of transmitting data (Vishwanath et al., 2015). The costs are directly tied to the number of parameters of a client update in an FL system (Yousefpour et al., 2023). As such, we can calculate the total energy consumption of communication as

$$P_\text{t} = E_t \cdot \mathcal{B} = \begin{aligned}[t] &(n_\text{as} \cdot E_\text{as} + E_\text{bng} + n_e \cdot E_e \\ &+ n_c \cdot E_c + n_d \cdot E_d) \cdot \mathcal{B}. \end{aligned} \tag{1}$$

From a client to a server, the communication network and its total energy consumption E_t is organized as follows: E_as, E_bng, E_e, E_c, E_d are the per-bit energy consumption of edge ethernet switches, the broadband network gateway (BNG), one or more edge routers n_e, one or more core routers n_c, and one or more data center Ethernet switches n_d, respectively. To get the total energy consumption for communication, we multiply E_t with the size of a model update d in bits b, $\mathcal{B} = d \cdot b$. Usually, a model parameter has a precision of $b = 32$ bits but can vary based on the specific application (Gupta et al., 2015). Jalali et al. (2014) present the per-bit energy consumption for at least one device per network hop that can be used as a guideline. While it is possible to trace what route a network package takes (Butskoy, 2023), it is currently impossible to track

Table 1: The algorithmic costs estimate how well the privacy mechanisms scale. Especially, the server-side communication provides evidence that the cryptographic methods are significantly more expensive than (ϵ, δ)-DP.

Privacy Technique	Pot. AI Act compliant*	Client Computation	Client Communication	Client Space	Server Computation	Server Communication	Server Space	Algorithm
(ϵ, δ)-DP**	✓	$\mathcal{O}(d)$***	$\mathcal{O}(1)$	$\mathcal{O}(d)$	$\mathcal{O}(\lvert K\rvert)$	$\mathcal{O}(\lvert K\rvert)$	$\mathcal{O}(\lvert K\rvert)$	Andrew et al. (2021)
SMPC	✓	$\mathcal{O}(\lvert K\rvert^2 + \lvert K\rvert \times d)$	$\mathcal{O}(\lvert K\rvert + d)$	$\mathcal{O}(\lvert K\rvert + d)$	$\mathcal{O}(\lvert K\rvert^2 \times d)$	$\mathcal{O}(\lvert K\rvert^2 + \lvert K\rvert \times d)$	$\mathcal{O}(\lvert K\rvert^2 + d)$	Bonawitz et al. (2017)
HEC	Limited	$\mathcal{O}(d)$	$\mathcal{O}(d)$	$\mathcal{O}(d)$	$\mathcal{O}(\lvert K\rvert \times d)$	$\mathcal{O}(\lvert K\rvert \times d)$	$\mathcal{O}(d)$	Jin et al. (2023)

* Potential evaluation for future AI Act compliance
** $\mathcal{O}(d)$ for computation originates from clipping a model update. When the FL aggregator is running in a secure enclave, we can also clip updates on the server at cost $\mathcal{O}(|K| \times d)$
*** d is the dimensionality of w

the real energy consumption of a data package sent over the network. It specifically depends on what device has been used at what point in the communication chain. As such, if the AI Act requires us to track the *total energy* consumed by a service, we have to develop solutions to track the networking-related energy consumption. We already see promising progress towards holistically accounting for energy efficiency in FL applications (Mehboob et al., 2023; Qiu et al., 2023; Wiesner et al., 2023).

Robustness and Quality Management. Aside from energy, the AI Act also requires FL service providers to provide a robust model with consistently high performance. Since FL does not allow immediate data access, we must find indirect ways to evaluate the model quality and ensure robustness against over-time-evolving input data. We look into what indirect strategies exist to control model quality and measure the cost of existing solutions. Further, we study existing secure and private computing methods with regard to their applicability in FL applications under the AI Act that holds the FL service provider liable for any robustness or quality management issues.

3.2. Quantitative Analysis

We design experiments to quantify those measurable components of changes we have to introduce to FL systems to comply with the AI Act. We use FL to fine-tune a 110M parameter BERT (Devlin et al., 2018) model to classify emails of the 20 News Group dataset (Lang, 1995). Such a setup can be found in job application pre-screening tools, which are classified as *high-risk applications* under the AI Act. Details on the training pipeline and the exact experimental setup are available in Appendix B.

The AI Act *data governance* regulation requires FL service providers to adhere to GDPR and protect data by design. With the absence of data movement in FL applications, we have already taken a major step toward private-by-design applications. However, existing research demonstrates that there are still open attack vectors (Geiping et al., 2020), and closing them comes at a cost. We aim to understand the trade-off between scaling a system and the cost incurred by introducing private and secure computation methods (Section 4.1).

The forthcoming introduction of the AI Act *energy efficiency* directives may require us to implement FL applications with sustainable and energy-saving techniques in mind. However, the additional duties to account for data governance, robustness, and quality management require us to frequently analyze the FL model, track the energy consumption of the whole system, and ensure privacy throughout the entire application. As this introduces a computational overhead, we aim to understand exactly where potentials for improved energy efficiency can be found and how to address them (Section 4.2).

The *robustness and quality management* requirements introduce the necessity of closely monitoring the FL model while training. This is to ensure consistently high performance. Close monitoring naturally increases the requirement to communicate and validate the FL model. This incurs additional costs. We evaluate the question of how expensive robustness and quality management are in FL applications and how they could be mitigated (Section 4.3).

3.3. Qualitative Analysis

In our qualitative analysis, we focus specifically on the characteristics of FL that are not empirically measurable. To do so, we take legislators' perspective and look at the qualitative potential of FL. We aim to identify the potential of FL to serve the fundamental rights of privacy and data bias prevention. Our objective is to evaluate whether FL has the significant potential to become *the* most adopted privacy-preserving ML technique for high-risk applications under the AI Act.

Overall, our analysis aims to add to the understanding of the future potential of FL under the AI Act and derive research priorities to help with the broad adoption of FL.

4. Analysis

Our analysis combines quantitative analysis considering data governance, energy efficiency, as well as robustness, and quality management. We expand on our empirical results with a qualitative analysis to identify the characteristics of FL under the AI Act that cannot be easily measured. The **key insight** are highlighted.

4.1. Data Governance

Secure Multi-Party Computation (SMPC), Homomorphic Encryption (HEC), and (ϵ, δ)-Differential Privacy ((ϵ, δ)-DP) all provide technical measures to improve data privacy in FL. SMPC and HEC are cryptographic methods that rely on key exchange between clients (Bonawitz et al., 2017; Jin et al., 2023). The client model update encryption removes the ability to track a client's individual contribution toward a global model. At the same time, aggregation remains possible as SMPC and HEC keep arithmetic properties.

A clear strategy to employing the *right* private and secure computation technique in an AI Act-compliant FL system is required. We find all methods to come at significant costs (Figure 1 and Table 1). While the cryptographic methods keep the original shape of the model updates in an encrypted form, they require extensive communication and, in the worst case, point-to-point communication between clients. This creates practical challenges when scaling an FL system (Jin et al., 2023). However, this is where (ϵ, δ)-DP excels (McMahan et al., 2017b; Andrew et al., 2021). Instead of requiring the clients K to establish a joint secure computation regime, (ϵ, δ)-DP introduces privacy by model parameter perturbation. In detail, we perturb and clip each model weight $w^k_{t+1} \in \mathbb{R}^d$ with dimension d of a client $k \in K$ with random noise ξ sampled from a Gaussian distribution $\mathcal{N}(0, \sigma^2_\Delta)$. The variance σ^2_Δ depends on the number of clients per aggregation round and how many clients have exceeded the clipping threshold in the previous training round. The quantity z scales the noise that is actually applied to local client update w^k_{t+1} and ultimately determines the degree of privacy we achieve under a constrained privacy budget ϵ and a data leakage risk δ,

$$w_{t+1} = \frac{1}{|K|} \sum_{k=1}^{|K|} \left(w^k_{t+1} + z \cdot \xi \right). \tag{2}$$

As can be seen, the perturbation mechanism benefits from increasing the number of clients in an aggregation round. Thus, (ϵ, δ)-DP is particularly useful for scaled systems, while the cryptographic methods can be useful in smaller systems. The optimal strategy for choosing the right privacy technique is yet to be found.

It is unclear whether (ϵ, δ)-DP can be compliant with regulatory privacy as enacted by GDPR. While we know that an $\epsilon \leq 1$ provides strong guarantees for privacy, the guarantee always depends on δ (Dwork & Roth, 2013). While setting δ is trivial in centralized learning, as we know the dataset size before training, it is challenging in FL. We cannot be certain about how many clients will eventually participate in the training process and how many data points each client contributes. As such, we require heuristics to set δ appropriately for training in a dynamically evolving FL system. An effort to evaluate (ϵ, δ)-DP for regulatory compliance is planned in the United States (The White House, 2023); the same should be done in the EU, potentially as a joint effort.

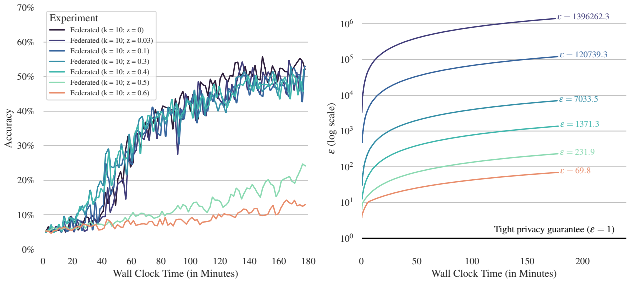

Figure 1: To achieve high privacy guarantees in small systems, we require high z that come at significant model performance and efficiency costs. Training stability also diminishes with increasing z. ϵ is calculated based on $\delta = \frac{1}{16{,}000}$.

4.2. Energy Efficiency

Optimizing for energy efficiency must become a priority in FL research. In our experiment with the BERT pipeline, we find FL to be 5× less energy efficient than centralized training when looking at the computational and communication effort to reach 50% accuracy (TTA50) (Figure 2). Communication accounts for 26% of the total energy draw. With parameter-efficient fine-tuning (PEFT), we save on computational resources and reduce communication costs significantly.[2] As such, efficient methods for computation in large models can also reduce communication. First works have shown significant potential for PEFT methods to address energy efficiency but still come at extensive warm-up costs (Babakniya et al., 2023) that we need to mitigate in the future. As the EU has introduced the Emission Trading System-2 (ETS-2), CO_2 emissions are capped by the number of total emission certificates available (Abrell et al., 2023). This immediately impacts the electricity price, which will surge as the ETS-2 Market Mechanism comes into effect in 2027. The price for CO_2 certificates is expected to rise by as much as 6×, from €45 in 2024 to €300 with a market-made price. Overall, this increases the need for energy efficiency improvements.

The AI Act introduces a privacy-energy efficiency tradeoff. As pointed out in Section 4.1, we do not know about the *right* choice of private and secure computation for any given FL application as it depends on the number of clients in the system, the number of clients per training round, and the amount of input data available on each client. The cryp-

[2] It is important to note if we were to use full-model fine-tuning, the power consumption for computing would amount to 0.48 kWh and communication to 196 kWh to reach TTA50 (2100× more than PEFT). We communicate 110M parameters over 1,057 rounds.

tographic methods introduce significant computational and communication overhead, while (ϵ, δ)-DP does not. However, for small-scale FL systems (< 100 clients per aggregation round), the z has to be comparably larger than in large-scale systems (McMahan et al., 2017b). This significantly reduces the model performance and slows the training process (Figure 1). As such, we face a privacy-energy trade-off in current-state FL systems, regardless of the private and secure computation technique. We must address this challenge in light of the AI Act and its call for more energy efficiency.

4.3. Robustness & Quality Management

We pay significantly for robustness guarantees. Frequent validation in FL under the control of the service provider (Rec. 45), i.e., the server, is a necessity to track model performance, understand a model's robustness against data heterogeneity (Li et al., 2021), and domain shifts (Huang et al., 2023). However, the energy consumption of idle clients while waiting for a model to validate and be ready for the next aggregation round has not been part of the power equation thus far. With the AI Act, a service provider may have to account for the *total energy* consumed during training (Art. 40, Rec. 85a). Thus, we must account for these idle times as well. As seen in Figure 2, these idle times consume 31% of all power. To address this challenge, we could regulate the validation process. Similar to what has been done for fair FL methods, we can make validation depend on the loss volatility (Li et al., 2020b;c) and validate as follows:

1. *Only validate the final model*. The fastest way to train is to only validate the final model. However, this approach induces the risk of creating a model with no utility and wasting all energy consumed. Also, legal compliance is in doubt since sparse monitoring contradicts the AI Act requirements (Art. 17).
2. *Validate after every i^{th} aggregation round*. While a frequent validation strategy reduces the risk of overfitting a model, it creates significant idle time. Trading off the validation frequency for energy efficiency could be a promising approach to achieving full compliance with the AI Act.
3. *Validate asynchronously*. We may validate models while starting the next aggregation round to avoid any idle energy consumption. This bears the risk of producing an overfitted model but can save energy after all. A careful trade-off can help create an energy-efficient system while producing robust models.

The applicability of HEC under the AI Act is potentially limited. Since HEC denies server-side model evaluation by design (Jin et al., 2023), we must rely on client-side validation techniques. This is only feasible in applications with trustworthy clients and where validation datasets can be distributed to clients. Promising directions for trustworthy computing are secure enclaves and trusted execution environments (Sabt et al., 2015). In the case of client-side validation under the AI Act, the FL service provider still remains liable for a consistent and high-quality model.

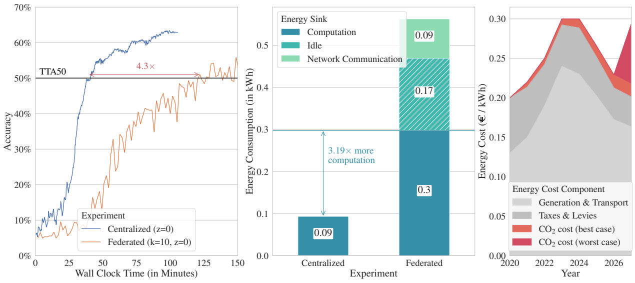

Figure 2: Baseline Experiments. We identify major causes of energy efficiencies in FL systems. The projected energy costs in the EU, especially CO$_2$ pricing, require us to focus on improving the energy efficiency of FL.

4.4. Qualitative Analysis

Access to siloed data. Creating data sets is complex and can, at best, be based on the entire internet (Schuhmann et al., 2022; Gao et al., 2020). Storing and transmitting such huge amounts of data can quickly become costly. Additionally, data quality is just as important as the data itself (Whang et al., 2023). We assume that a lot of high-quality, simultaneously personally identifiable data is naturally not publicly accessible. Despite the EU's plan to make anonymized data available worldwide (Rec. 45), collecting such data poses significant challenges, as we outlined in this section. FL can provide us access to this data, potentially greatly improving the high-risk application's functionality.

With broader data access, we generate more representative models and data. The AI Act emphasizes the importance of examining and mitigating potential biases in the data used for training. This is particularly important if these biases affect fundamental rights (Art. 10.2f). To achieve this, the datasets must be curated and prepared for training after they are centrally aggregated. If a concept shift alters the basic assumptions about the data (Lu et al., 2019), the dataset must be adjusted anew. FL offers a potential solution to this problem. As FL operates on the clients close to the data source, it means that, by definition, we have access to the latest and most representative data. Given that we train for many rounds and randomly sample clients for aggregation out of an evolving client base, we automatically create a representative global model over time since the model evolves along with the client base. As such, it can be easier to comply with the AI Act requirements by design.

FL provides simple data lineage. Since the training data never leaves the clients in FL, it is less complex to track the data lineage, meaning where the data originated, where it has gone, and its usage. On the one hand, as data is easy to trace, the GDPR requirement to know how the data is being used (cf. Section 2) is easy to answer and easier to ensure. On the other hand, every time data is sent, it is open to man-in-the-middle attacks (Conti et al., 2016), and when it is stored in multiple locations, all data hosts are vulnerable to unauthorized access. Additionally, every time a human is in the loop regarding data management, there is a potential risk of error (Evans et al., 2016), which can lead to data leakage to a third party. This lack of vulnerability in FL systems removes most of the potential penalties under the GDPR, which are closing in at €4.5B over the last five years by January 2024 (CMS Law, 2024). This fact alone can encourage data providers to make data available to FL applications as the risk on their side is significantly lower than before.

5. Future Research Priorities

The challenges highlighted in our analysis indicate that FL can strongly align with the needs of the AI Act if the core challenges are being addressed soon. To do so, we outline the future research priorities that we see as a necessary redirection for the FL community to make FL a legally compliant and commercially viable solution. The **research priorities** are highlighted.

The data quality requirements are currently not amenable to FL. We need to find solutions to meet the data quality requirements of the AI Act under Art. 10 without having direct access to the data. First, if data quality can be indirectly inferred through techniques with heavy energy investment, how do they compare to direct techniques that require direct data access? Second, do techniques that combat non-IID data provide enough robustness guarantees to qualify for compliance, and if not, what is missing? Third, the AI Act mentions that data processing methods at the source are desirable (cf. Section 2), but it is not made clear who is responsible for the data if multiple parties are involved. To make meaningful progress towards the goals of the AI Act, it is imperative that the FL research community focuses on improving data quality techniques and ensures that Art. 10, under legal guidance, can be effectively implemented in real-world systems.

CO_2-based optimization to compete with centralized training. FL is currently not achieving competitive energy efficiency compared to centralized training (cf. Section 4.2). Even if using DP will be considered partially compliant regarding data governance, it results in extensive energy costs, just as training and quality monitoring do. Therefore, we require new techniques to address these costs concurrently. While there is ongoing research focused on energy efficiency in specific use cases (Yousefpour et al., 2023; Salh et al., 2023; Kim et al., 2023; Albelaihi et al., 2022), there is a need for a designated effort to bridge the gap to centralized baselines, which might be running on fully renewable energy or be more energy efficient by default due to locality.

Expression of privacy in the context of the EU AI Act. FL is private by design and should fit EU AI Act compliance well. Unfortunately, we found significant shortfalls in energy efficiency and data governance compared to centralized training (Section 4). If the FL research community does not act now, centralized training may be seen as the best approach for high-risk applications. This could pose a problem for individuals if privacy is not considered a key component from the outset. If centralized training is deemed the best approach due to better energy efficiency and easier data governance compliance, it is unclear how the right to privacy will be expressed in practice. It is crucial that the interpretation of the law, such as with the GDPR and subsequent cookie banners (The European Commission, 2023), does not result in the end-user bearing the entire burden while operators take no responsibility.

Privacy-preserving techniques alignment within the AI Act. We evaluated SMPC, HEC, and (ϵ, δ)-DP within their current applicability to the energy and data governance aspects and found them to be lacking in multiple ways (cf. Section 4.1). From a technical point of view, we need to work on improving these techniques to be more energy-efficient. However, researchers should advocate for concrete privacy goals to help align legal and arithmetic privacy.

Technical framework for regulatory compliance and representative AI Act baselines. We require a framework that specifically caters to FL, as it has distinct differences from centralized DL in terms of model lifecycle and data access. This framework is necessary so that not everyone is faced with complying with the AI Act from the outset, but to propose best practices to provide a solid basis (in conjunction with the standardization organizations in Art. 40). Through this framework, the development of comparable baselines is necessary to set the standard on privacy-by-design deep learning in high-risk applications. Specifically, this framework should strive to standardize edge hardware comparisons, clarify who is responsible for customer energy costs, and establish clear targets for training and deployment.

6. Conclusion

In this position paper, we analyze the AI Act and its impact on FL. We outline how we need to redirect research priorities with regard to achieving regulatory compliance, the energy-privacy trade-off introduced by the AI Act, and the need for new optimization dimensions in FL. Depending on forthcoming energy efficiency requirements, it may also require us to think about holistic monitoring systems while staying energy efficient. It is also important to address challenges that have been solved in centralized learning such that FL can keep up. With this, we, as the FL research community, can send a clear signal to legislation and the broad public that we have a strong interest in making FL *the* distributed privacy-preserving DL technology of the future by incorporating societal priorities into our research. We can do so by answering the open call by the EU Commission to support the newly established EU AI Office to close the gap between regulatory framing and technical implementation (Nature, 2024).

Impact Statement

This paper presents work whose goal is to suggest future research directions that will help ensure that FL, with its worthwhile goal of preserving privacy, aligns with other societal values espoused by the EU AI Act, such as keeping AI systems robust, unbiased, energy efficient, transparent, ethical, and secure, especially for high-risk use cases. This paper transparently addresses the challenges that FL may encounter as regards the data governance, energy efficiency, and robustness provisions of the Act and the associated trade-offs that AI providers must be aware of and responsibly navigate when complying with the Act and the societal ideals it encapsulates.

Acknowledgements

We would like to thank Hadrien Pouget at the Carnegie Endowment for International Peace, Yulu Pi at the Leverhulme Center for the Future of Intelligence, Bill Shen at the University of Cambridge, and The Future Society for their help in understanding the legal aspects of the AI Act and the White House Executive Order on AI.

This work is partially funded by the Bavarian Ministry of Economic Affairs, Regional Development and Energy (Grant: DIK0446/01) and the German Research Foundation (Grant: 392214008).

We welcome feedback on our work and are open to including aspects that we might have missed in future versions of this paper.

References

Abrell, J., Bilici, S., et al. Optimal allocation of the EU carbon budget: A multi-model assessment, may 2023. URL https://ariadneprojekt.de/media/2022/06/Ariadne-Analysis_Carbon-Budget-multi-model-assessment_June2022.pdf. Document 32023L0959.

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.

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.

Appelhans, D. and Walkup, B. Leveraging NVLINK and asynchronous data transfer to scale beyond the memory capacity of GPUs. In *Proceedings of the 8th Workshop on Latest Advances in Scalable Algorithms for Large-Scale Systems*, SC '17. ACM, November 2017. doi: 10.1145/3148226.3148232. URL http://dx.doi.org/10.1145/3148226.3148232.

Babakniya, S., Elkordy, A. R., Ezzeldin, Y. H., Liu, Q., Song, K.-B., El-Khamy, M., and Avestimehr, S. SLoRA: Federated parameter efficient fine-tuning of language models, 2023. URL https://arxiv.org/abs/2308.06522.

Beutel, D. J., Topal, T., Mathur, A., Qiu, X., Fernandez-Marques, J., Gao, Y., Sani, L., Li, K. H., Parcollet, T., de Gusmão, P. P. B., and Lane, N. D. Flower: A Friendly Federated Learning Research Framework, 2020. URL https://arxiv.org/abs/2007.14390.

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.

Butskoy, D. Linux Traceroute, 12 2023. URL http://traceroute.sourceforge.net/.

Chen, D., Gao, D., Kuang, W., Li, Y., and Ding, B. pFL-Bench: A Comprehensive Benchmark for Personalized Federated Learning. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2022. URL https://openreview.net/forum?id=2ptbv_JjYKA.

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

Conti, M., Dragoni, N., and Lesyk, V. A Survey of Man In The Middle Attacks. *IEEE Communications Surveys & Tutorials*, 18(3):2027–2051, 2016. doi: 10.1109/COMST.2016.2548426.

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.

Desislavov, R., Martínez-Plumed, F., and Hernández-Orallo, J. Trends in AI inference energy consumption: Beyond the performance-vs-parameter laws of deep learning. *Sustainable Computing: Informatics and Systems*, 38:100857, April 2023. ISSN 2210-5379. doi: 10.1016/j.suscom.2023.100857. URL http://dx.doi.org/10.1016/j.suscom.2023.100857.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018. URL https://arxiv.org/abs/1810.04805.

Dwork, C. and Roth, A. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2013. ISSN 1551-3068. doi: 10.1561/0400000042. URL http://dx.doi.org/10.1561/0400000042.

El Hanjri, M., Kabbaj, H., Kobbane, A., and Abouaomar, A. Federated learning for water consumption forecasting in smart cities. In *ICC 2023-IEEE International Conference On Communications*, pp. 1798–1803. IEEE, 2023.

European Commission. Market analysis - Electricity market - recent developments, 07 2023a. URL https://energy.ec.europa.eu/data-and-analysis/market-analysis_en.

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

European Environment Agency. Greenhouse gas emission intensity of electricity generation, Oct 2023. URL https://www.eea.europa.eu/data-and-maps/daviz/co2-emission-intensity-14#tab-chart_7.

European Parliament and Council. Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts, 01 2024. URL https://www.linkedin.com/posts/dr-laura-caroli-0a96a8a_ai-act-consolidated-version-activity-7155181240751374336-B3Ym/.

Eurostat. Electricity prices for household consumers, Oct 2023. URL https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Electricity_price_statistics.

Evans, M., Maglaras, L. A., He, Y., and Janicke, H. Human behaviour as an aspect of cybersecurity assurance. *Security and Communication Networks*, 9(17):4667–4679, 2016.

Fallah, A., Mokhtari, A., and Ozdaglar, A. Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 3557–3568. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/24389bfe4fe2eba8bf9aa9203a44cdad-Paper.pdf.

Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., Phang, J., He, H., Thite, A., Nabeshima, N., et al. The pile: An 800gb dataset of diverse text for language modeling. *arXiv preprint arXiv:2101.00027*, 2020.

Geiping, J., Bauermeister, H., Dröge, H., and Moeller, M. Inverting Gradients - How easy is it to break privacy in federated learning? In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 16937–16947. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/c4ede56bbd98819ae6112b20ac6bf145-Paper.pdf.

Google. 24/7 Carbon-Free Energy by 2030, 2023. URL https://www.google.com/about/datacenters/cleanenergy/.

Graves, L., Nagisetty, V., and Ganesh, V. Amnesiac machine learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 11516–11524, 2021.

Gupta, S., Agrawal, A., Gopalakrishnan, K., and Narayanan, P. Deep Learning with Limited Numerical Precision. In Bach, F. R. and Blei, D. M. (eds.), *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pp. 1737–1746. JMLR.org, 2015. URL http://proceedings.mlr.press/v37/gupta15.html.

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.

Han, S., Buyukates, B., Hu, Z., Jin, H., Jin, W., Sun, L., Wang, X., Wu, W., Xie, C., Yao, Y., Zhang, K., Zhang, Q., Zhang, Y., Avestimehr, S., and He, C. FedMLSecurity: A Benchmark for Attacks and Defenses in Federated Learning and Federated LLMs, 2023.

He, C., Li, S., So, J., Zeng, X., Zhang, M., Wang, H., Wang, X., Vepakomma, P., Singh, A., Qiu, H., Zhu, X., Wang, J., Shen, L., Zhao, P., Kang, Y., Liu, Y., Raskar, R., Yang, Q., Annavaram, M., and Avestimehr, S. FedML: A Research Library and Benchmark for Federated Machine Learning, 2020. URL https://arxiv.org/abs/2007.13518.

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.

Huang, W., Ye, M., Shi, Z., Li, H., and Du, B. Rethinking Federated Learning With Domain Shift: A Prototype View. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16312–16322, June 2023.

Huang, Y., Gupta, S., Song, Z., Li, K., and Arora, S. Evaluating Gradient Inversion Attacks and Defenses in Federated Learning. In Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=0CDKgyYaxC8.

Jalali, F., Ayre, R., Vishwanath, A., Hinton, K., Alpcan, T., and Tucker, R. Energy Consumption of Content Distribution from Nano Data Centers versus Centralized Data Centers. *ACM SIGMETRICS Performance Evaluation Review*, 42(3):49–54, December 2014. ISSN 0163-5999. doi: 10.1145/2695533.2695555. URL http://dx.doi.org/10.1145/2695533.2695555.

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.

Klimas, T. and Vaiciukaite, J. The law of recitals in European Community legislation. *ILSA J. Int'l & Comp. L.*, 15:61, 2008.

Kurmanji, M., Triantafillou, P., and Triantafillou, E. Towards Unbounded Machine Unlearning. *arXiv preprint arXiv:2302.09880*, 2023.

Lang, K. NewsWeeder: Learning to Filter Netnews. In Prieditis, A. and Russell, S. (eds.), *Machine Learning Proceedings 1995*, pp. 331–339. Morgan Kaufmann, San Francisco (CA), 1995. ISBN 978-1-55860-377-6. doi: https://doi.org/10.1016/B978-1-55860-377-6.50048-7. URL https://www.sciencedirect.com/science/article/pii/B9781558603776500487.

Lei, S. and Tao, D. A comprehensive survey to dataset distillation. *arXiv preprint arXiv:2301.05603*, 2023.

Li, A., Song, S. L., Chen, J., Li, J., Liu, X., Tallent, N. R., and Barker, K. J. Evaluating Modern GPU Interconnect: PCIe, NVLink, NV-SLI, NVSwitch and GPUDirect. *IEEE Transactions on Parallel and Distributed Systems*, 31(1):94–110, January 2020a. ISSN 2161-9883. doi: 10.1109/tpds.2019.2928289. URL http://dx.doi.org/10.1109/TPDS.2019.2928289.

Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., and Smith, V. Federated Optimization in Heterogeneous Networks. In Dhillon, I., Papailiopoulos, D., and Sze, V. (eds.), *Proceedings of Machine Learning and Systems*, volume 2, pp. 429–450, 2020b. URL https://proceedings.mlsys.org/paper_files/paper/2020/file/1f5fe83998a09396ebe6477d9475ba0c-Paper.pdf.

Li, T., Sanjabi, M., Beirami, A., and Smith, V. Fair Resource Allocation in Federated Learning. In *International Conference on Learning Representations*, 2020c. URL https://openreview.net/forum?id=ByexElSYDr.

Li, X., JIANG, M., Zhang, X., Kamp, M., and Dou, Q. FedBN: Federated Learning on Non-IID Features via Local Batch Normalization. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=6YEQUn0QICG.

Liu, Y., James, J., Kang, J., Niyato, D., and Zhang, S. Privacy-preserving traffic flow prediction: A federated learning approach. *IEEE Internet of Things Journal*, 7(8):7751–7763, 2020.

Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., and Zhang, G. Learning under Concept Drift: A Review. *IEEE Transactions on Knowledge and Data Engineering*, 31(12):2346–2363, 2019. doi: 10.1109/TKDE.2018.2876857.

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.

Mehboob, T., Bashir, N., Iglesias, J. O., Zink, M., and Irwin, D. Cefl: Carbon-efficient federated learning, 2023. URL https://arxiv.org/abs/2310.17972.

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.

Pfitzner, B., Steckhan, N., and Arnrich, B. Federated learning in a medical context: a systematic literature review. *ACM Transactions on Internet Technology (TOIT)*, 21(2):1–31, 2021.

Prechelt, L. Early stopping-but when? In *Neural Networks: Tricks of the trade*, pp. 55–69. Springer, 2002.

Qiu, X., Parcollet, T., Fernandez-Marques, J., Gusmao, P. P. B., Gao, Y., Beutel, D. J., Topal, T., Mathur, A., and Lane, N. D. A first look into the carbon footprint of federated learning. *Journal of Machine Learning Research*, 24(129):1–23, 2023. URL http://jmlr.org/papers/v24/21-0445.html.

Reddi, S., Charles, Z., Zaheer, M., Garrett, Z., Rush, K., Konečný, J., Kumar, S., and McMahan, H. B. Adaptive Federated Optimization, 2020. URL https://arxiv.org/abs/2003.00295.

Ritchie, H. and Rosado, P. Energy Mix. *Our World in Data*, 2020. https://ourworldindata.org/energy-mix.

Sabt, M., Achemlal, M., and Bouabdallah, A. Trusted Execution Environment: What It is, and What It is Not. In *2015 IEEE Trustcom/BigDataSE/ISPA*. IEEE, August 2015. doi: 10.1109/trustcom.2015.357. URL http://dx.doi.org/10.1109/Trustcom.2015.357.

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.

Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C., Wightman, R., Cherti, M., Coombes, T., Katta, A., Mullis, C., Wortsman, M., et al. Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems*, 35:25278–25294, 2022.

Tan, A. Z., Yu, H., Cui, L., and Yang, Q. Towards Personalized Federated Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 34(12):9587–9603, December 2023. ISSN 2162-2388. doi: 10.1109/tnnls.2022.3160699. URL http://dx.doi.org/10.1109/TNNLS.2022.3160699.

The European Commission. Cookie Pledge, 2023. URL https://commission.europa.eu/live-work-travel-eu/consumer-rights-and-complaints/enforcement-consumer-protection/cookie-pledge_en.

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/.

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.

Torralba, A. and Efros, A. A. Unbiased look at dataset bias. In *CVPR 2011*, pp. 1521–1528. IEEE, 2011.

Tun, Y. L., Thar, K., Thwal, C. M., and Hong, C. S. Federated learning based energy demand prediction with clustered aggregation. In *2021 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 164–167. IEEE, 2021.

Vishwanath, A., Jalali, F., Hinton, K., Alpcan, T., Ayre, R. W. A., and Tucker, R. S. Energy Consumption Comparison of Interactive Cloud-Based and Local Applications. *IEEE Journal on Selected Areas in Communications*, 33(4):616–626, April 2015. ISSN 0733-8716. doi: 10.1109/jsac.2015.2393431. URL http://dx.doi.org/10.1109/JSAC.2015.2393431.

Wang, Y., Bennani, I. L., Liu, X., Sun, M., and Zhou, Y. Electricity consumer characteristics identification: A federated learning approach. *IEEE Transactions on Smart Grid*, 12(4):3637–3647, 2021.

Whang, S. E., Roh, Y., Song, H., and Lee, J.-G. Data collection and quality challenges in deep learning: A data-centric ai perspective. *The VLDB Journal*, 32(4):791–813, 2023.

Wiesner, P., Khalili, R., Grinwald, D., Agrawal, P., Thamsen, L., and Kao, O. Fedzero: Leveraging renewable excess energy in federated learning. 2023. doi: 10.48550/ARXIV.2305.15092. URL https://arxiv.org/abs/2305.15092.

Woisetschläger, H., Isenko, A., et al. FLEdge: Benchmarking Federated Machine Learning Applications in Edge Computing Systems. *arXiv preprint arXiv:2306.05172*, 2023.

Xu, H., Zhu, T., Zhang, L., Zhou, W., and Yu, P. S. Machine Unlearning: A Survey. *ACM Comput. Surv.*, 56(1), aug 2023. ISSN 0360-0300. doi: 10.1145/3603620. URL https://doi.org/10.1145/3603620.

Xu, J. and Wang, H. Client Selection and Bandwidth Allocation in Wireless Federated Learning Networks: A Long-Term Perspective. *IEEE Transactions on Wireless Communications*, 20(2):1188–1200, February 2021. ISSN 1558-2248. doi: 10.1109/twc.2020.3031503. URL http://dx.doi.org/10.1109/TWC.2020.3031503.

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.

Zhang, L., Li, L., Li, X., Cai, B., Gao, Y., Dou, R., and Chen, L. Efficient Membership Inference Attacks against Federated Learning via Bias Differences. In *Proceedings of the 26th International Symposium on Research in Attacks, Intrusions and Defenses*, RAID 2023. ACM, October 2023. doi: 10.1145/3607199.3607204. URL http://dx.doi.org/10.1145/3607199.3607204.

Zhao, H., Du, W., Li, F., Li, P., and Liu, G. FedPrompt: Communication-Efficient and Privacy-Preserving Prompt Tuning in Federated Learning. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2023.

Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., and Chandra, V. Federated learning with non-iid data. *arXiv preprint arXiv:1806.00582*, 2018.

Appendix

A. Details on the AI Act

In this appendix section, we provide additional background on the legal aspects of our work.

A.1. Article vs. Recitals in the AI Act

In our main paper, we argue with Articles and Recitals. Understanding the difference between both is vital. The following explanations are based on Klimas & Vaiciukaite (2008).

Article. An article formulates the actual binding law and defines requirements that need to be implemented in technical solutions. This is ultimately what decides on violations. However, some parts can appear ambiguous and leave room for interpretation. This is where Recitals come into play.

Recitals. They provide interpretation to the Articles and help in guiding what needs to be done to ensure full compliance by reciting elements of the Articles and putting them into context. As such, Recitals provide procedural details on how to implement a law in practice. While they form the basis for a common understanding of the AI Act, they are not legally binding.

A.2. The latest AI Act version

By the time of writing this paper late 2023 and early 2024, the official Journal of the European Union hosts the original draft of the AI Act, which was released on Apr. 21st, 2021. In January 2024, EU policymakers and journalists released the pre-final version of the AI Act based on the high public demand. Our work is based on this latest version since it contains the final regulation as it will eventually come into effect. It is available here: `https://www.linkedin.com/posts/dr-laura-caroli-0a96a8a_ai-act-consolidated-version-activity-7155181240751374336-B3Ym/` and `https://drive.google.com/file/d/1xfN5T8VChK8fSh3wUiYtRVOKIi9oIcAF/view`.

B. Additional Experimental Details

Table 2: Training hyperparameters per training regime.

Training	Data	Tot. Samples	Client							Server				
regime	Dist.	Seen	MB Size	Optimizer	LR	WD	Mom.	Damp.	Loc. Iter.	K	k	Strategy	LR	Mom.
Centralized	IID	80K	20	SGD	0.01	0.001	0.9	0.9	5	–	–	–	–	–
Federated	non-IID	80K	2	SGD	0.01	0.001	0.0	0.0	2	100	10	FedAvgM	1.0	0.9

Here, we provide additional details about our experimental results. For our empirical evaluations, we fine-tune the 110M parameter BERT transformer (Devlin et al., 2018) over the 20 News Group Dataset (Lang, 1995) such that we can reliably classify emails into one of 20 categories. For example, such a classification application can be used in a company's human resource processes to screen job applications. Under the AI Act, such a system is considered a high-risk application.

B.1. Dataset

In our empirical analysis, we use a state-of-the-art text classification task in FL research by means of the 20 Newsgroup Dataset (Lang, 1995), which consists of 18,000 email bodies that each belong to one of 20 classes. The dataset has a total of $18,000$ samples, of which we use $16,000$ for training, $1,000$ for validation, and $1,000$ for testing. As our work aims to quantify the cost of FL and associated private computing methods in realistic systems in line with the EU AI Act requirements (Council of the European Union, 2021), we chose to sample 100 non-IID client subsets via a Latent Dirichlet Allocation (LDA) with $\alpha = 1.0$, which is widely used in FL research (Babakniya et al., 2023; He et al., 2020; Reddi et al., 2020). The data distribution is visualized in Figure 3.

B.2. Model

We fine-tune the BERT model (Devlin et al., 2018) with 110M parameters by using the parameter-efficient fine-tuning technique Low-Rank Adapters (LoRA). We use a LoRA configuration that has been well explored in FL settings (Babakniya

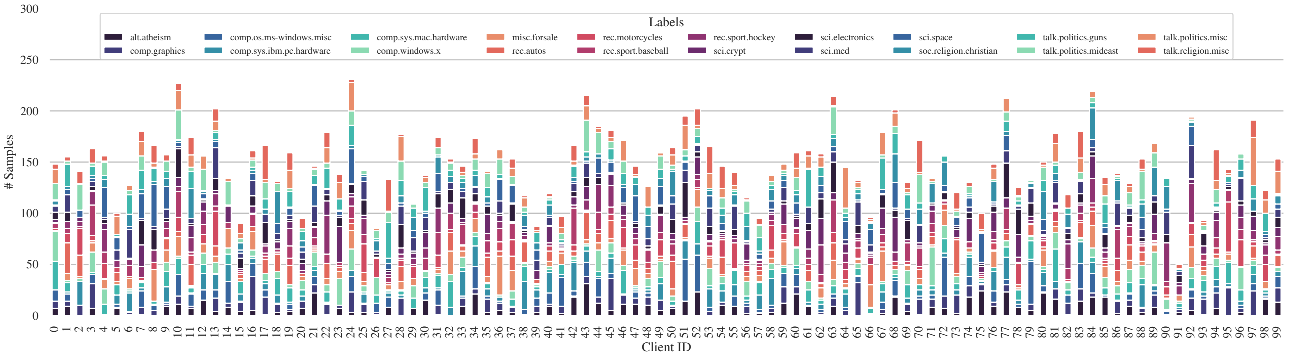

Figure 3: Visualization of client subsets for all of our experiments.

et al., 2023), which results in 52K trainable parameters (0.05% of total model parameters). This reduces the computational intensity of the task at hand and minimizes the communication load for the FL setup, as we must only communicate the trainable parameters. The BERT model is used to classify the emails into the 20 distinct categories in the dataset, which resembles a realistic task as it is frequently found in job application pre-screening applications, where the email bodies (input data) often contain sensitive and personal data.

FL configuration. We use the Federated Averaging (FedAvg) algorithm to facilitate all FL experiments (McMahan et al., 2017a) and train for 2000 aggregation rounds. We choose a participation rate of 10% for each aggregation round, i.e., $k = 10$ out of $K = 100$.

(ϵ, δ)-DP configuration. We employ sample-level (ϵ, δ)-DP for centralized learning, and for FL, we use user-level (ϵ, δ)-DP. Both methods provide the same privacy guarantees (Dwork & Roth, 2013). The parameterization for both is identical with $z = [0.0, 0.03, 0.1, 0.3, 0.4, 0.5, 0.6]$ and $\delta = \frac{1}{16{,}000}$, setting the data leakage risk to the inverse of the number of total training samples (Andrew et al., 2021; McMahan et al., 2017b). For the experiment with $z = [0.5; 0.6]$, we had to change the Learning Rate from 0.01 to 0.001.

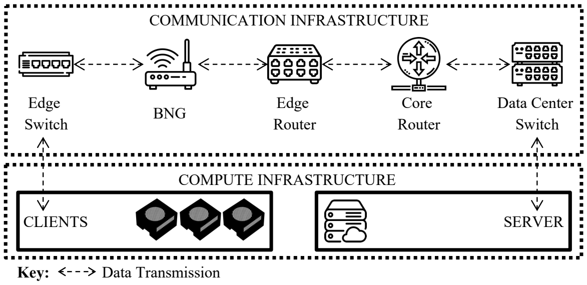

Figure 4: FL system design depicting the network topology for an aggregation round in FL between clients and the aggregation server. Every communication point consumes energy per transmitted bit, which must be accounted for.

Energy monitoring. We monitor our dedicated clients - NVIDIA Jetson AGX Orin - with 2Hz and measure their total energy consumption while participating in our FL setup. We also use a single Orin device for the centralized experiments for a fair comparison. For our cost estimations, we use the average price per kWh in the EU, 0.29 $\frac{€}{\text{kWh}}$ (Eurostat, 2023). The EU Commission produces quarterly reports on the electricity price trends (European Commission, 2023a). Directly proportional to the power consumption, we emit 252 $\frac{gCO_2e}{\text{kWh}}$ (European Environment Agency, 2023). Regarding communication energy, we assume the average communication route from a private household to a data center with $n_{as} = 1$, $n_e = 3$, $n_c = 5$, and $n_d = 2$ (cf. Equation (1)) (Jalali et al., 2014). For the energy consumption per transmitted bit per network hop, we adopt the values from Vishwanath et al. (2015); Jalali et al. (2014) (Table 3).

Table 3: Energy consumption per bit network communication for our holistic energy monitoring approach. Values are adopted from Vishwanath et al. (2015); Jalali et al. (2014).

Network Location	Device Name	Upload Cost (nJ/bit)	Download Cost (nJ/bit)
Edge Switch	Fast Ethernet Gateway	352	352
BNG	ADSL2+ Gateway (100 Mbit/s)	14809	2160
Edge Router	–	37	37
Core Router	–	12.6	12.6
Data Center Switch	Ethernet Switch	19.6	19.6

B.3. Hardware

We evaluate the training pipeline on a state-of-the-art embedded computing cluster with NVIDIA Jetson AGX Orin 64 GB devices (Orin), where each device has 12 ARMv8 CPU cores, an integrated GPU with 2048 CUDA cores, and 64 Tensor cores. The CPU and GPU share 64 GB of unified memory. The network interconnect is 10 GBit/s per client. We monitor the system metrics with a sampling rate of 2 Hz, including energy consumption in Watt (W). We use a data center server as an FL server. The server has 112 CPU cores, 384 GB of memory, an NVIDIA A40 GPU, and a 40 GBit/s network interface.

C. Algorithmic Cost Analysis for Private and Secure Computing Techniques in FL

In this section, we outline how we identified the algorithmic costs of state-of-the-art secure and private computing techniques. We omit the algorithmic costs of FedAvg and focus only on the privacy overhead. We discuss (ϵ, δ)-DP as introduced by Andrew et al. (2021), SMPC as introduced by Bonawitz et al. (2017), and HEC as introduced by Jin et al. (2023).

C.1. (ϵ, δ)-Differential Privacy

The following algorithm (Algorithm 1) is taken verbatim from Andrew et al. (2021). For the client, the computational complexity $O(d)$ originates from adding ξ to each parameter of a model update as well as by computing Δ. The communication complexity is $O(1)$ as we need to communicate the standard deviation to parameterize ξ as well as the clipping threshold. The space complexity $O(d)$ originates from storing θ.

The server computational complexity $O(|K|)$ originates from computing $\tilde{b}^t$ and the communication complexity $O(|K|)$ as we only communicate constants between clients and the server. The space complexity $O(|K|)$ comes from storing b_i.

Algorithm 1 DPFedAvg-M with adaptive clipping

```
function Train($m$, $\gamma$, $\eta_c$, $\eta_s$, $\eta_C$, $z$, $\sigma_b$, $\beta$)
  Initialize model $\theta^0$, clipping bound $C^0$
  $z_\Delta \leftarrow \left(z^{-2} - (2\sigma_b)^{-2}\right)^{-\frac{1}{2}}$
  for each round $t = 0, 1, 2, \ldots$ do
    $\mathcal{Q}^t \leftarrow$ (sample $m$ users uniformly)
    for each user $i \in \mathcal{Q}^t$ in parallel do
      $(\Delta_i^t, b_i^t) \leftarrow \text{FedAvg}(i, \theta^t, \eta_c, C^t)$
    end for
    $\sigma_\Delta \leftarrow z_\Delta C^t$
    $\tilde{\Delta}^t = \frac{1}{m}\left(\sum_{i \in \mathcal{Q}^t} \Delta_i^t + \mathcal{N}(0, I\sigma_\Delta^2)\right)$
    $\bar{\Delta}^t = \beta\bar{\Delta}^{t-1} + \tilde{\Delta}^t$
    $\theta^{t+1} \leftarrow \theta^t + \eta_s \bar{\Delta}^t$
    $\tilde{b}^t = \frac{1}{m}\left(\sum_{i \in \mathcal{Q}^t} b_i^t + \mathcal{N}(O, \sigma_b^2)\right)$
    $C^{t+1} \leftarrow C^t \cdot \exp\left(-\eta_C(\tilde{b}^t - \gamma)\right)$
  end for
end function

function FedAvg($i$, $\theta^0$, $\eta$, $C$)
  $\theta \leftarrow \theta^0$
  $\mathcal{G} \leftarrow$ (user $i$'s local data split into batches)
  for batch $g \in \mathcal{G}$ do
    $\theta \leftarrow \theta - \eta\nabla\ell(\theta; g)$
  end for
  $\Delta \leftarrow \theta - \theta^0$
  $b \leftarrow \mathbb{I}_{||\Delta|| \leq C}$
  $\Delta' \leftarrow \Delta \cdot \min\left(1, \frac{C}{||\Delta||}\right)$
  return $(\Delta', b)$
end function
```

C.2. Secure Multi-Party Computation

The SecAgg algorithmic costs (Table 4) are taken from Bonawitz et al. (2017) Table 1. The naming convention has been adapted to our paper.

Table 4: SecAgg costs

computation	
User	$O(\lvert K\rvert^2 + d \cdot \lvert K\rvert)$
Server	$O(d \cdot \lvert K\rvert^2)$
communication	
User	$O(\lvert K\rvert + d)$
Server	$O(\lvert K\rvert^2 + d \cdot \lvert K\rvert)$
storage	
User	$O(\lvert K\rvert + d)$
Server	$O(\lvert K\rvert^2 + d)$

C.3. Homomorphic Encryption

The following algorithm (Algorithm 2) is taken verbatim from Jin et al. (2023). For the client, computational complexity $O(d)$ originates from encrypting and decrypting the model. The communication complexity $O(d)$ comes from communicating the aggregation mask once. The space complexity $O(d)$ is created by storing the aggregation mask.

The server computational complexity $O(|K| \times d)$ originates from the server-side model aggregation while the communication complexity $O(|K| \times d)$ comes from sending the encryption mask once. Storing the encryption mask on the server results in space complexity $O(d)$.

Algorithm 2 HE-Based Federated Aggregation

- $\llbracket \mathbf{W} \rrbracket$: the fully encrypted model | $[\mathbf{W}]$: the partially encrypted model;
- p: the ratio of parameters for selective encryption;
- b: (optional) differential privacy parameter.

```
// Key Authority Generate Key
$(pk, sk) \leftarrow HE.KeyGen(\lambda)$;
// Local Sensitivity Map Calculation
for each client $i \in [N]$ do in parallel
    $\mathbf{W}_i \leftarrow Init(\mathbf{W})$;
    $\mathbf{S}_i \leftarrow Sensitivity(\mathbf{W}, \mathcal{D}_i)$;
    $\llbracket \mathbf{S}_i \rrbracket \leftarrow Enc(pk, \mathbf{S}_i)$;
    Send $\llbracket \mathbf{S}_i \rrbracket$ to server;
end
// Server Encryption Mask Aggregation
$\llbracket \mathbf{M} \rrbracket \leftarrow Select(\sum_{i=1}^{N} \alpha_i \llbracket \mathbf{S}_i \rrbracket, p)$;
// Training
for $t = 1, 2, \ldots, T$ do
    for each client $i \in [N]$ do in parallel
        if $t = 1$ then
            Receive $\llbracket \mathbf{M} \rrbracket$ from server;
            $\mathbf{M} \leftarrow HE.Dec(sk, \llbracket \mathbf{M} \rrbracket)$;
        end
        if $t > 1$ then
            Receive $[\mathbf{W}_{\text{glob}}]$ from server;
            $\mathbf{W}_i \leftarrow HE.Dec(sk, \mathbf{M} \odot [\mathbf{W}_{\text{glob}}]) + (\mathbf{1} - \mathbf{M}) \odot [\mathbf{W}_{\text{glob}}]$;
        end
        $\mathbf{W}_i \leftarrow Train(\mathbf{W}_i, \mathcal{D}_i)$;
        // Additional Differential Privacy
        if Add DP then
            $\mathbf{W}_i \leftarrow \mathbf{W}_i + Noise(b)$;
        end
        $[\mathbf{W}_i] \leftarrow HE.Enc(pk, \mathbf{M} \odot \mathbf{W}_i) + (\mathbf{1} - \mathbf{M}) \odot \mathbf{W}_i$;
        Send $[\mathbf{W}_i]$ to server $\mathcal{S}$;
    end
    // Server Model Aggregation
    $[\mathbf{W}_{\text{glob}}] \leftarrow \sum_{i=1}^{N} \alpha_i \llbracket \mathbf{M} \odot \mathbf{W}_i \rrbracket + \sum_{i=1}^{N} \alpha_i ((\mathbf{1} - \mathbf{M}) \odot \mathbf{W}_i)$;
end
```