Federated Learning is Needed to Overcome Key Challenges Arising from the European Union AI Act

Anonymous Author(s)
Affiliation
Address
email

Abstract

The European Union AI Act (AI Act) introduces comprehensive requirements for AI systems regarding data governance, safety and security, and energy efficiency and sustainability, among others. High-risk AI applications, such as AI systems for medical data processing, face particularly stringent compliance requirements. We argue that *Federated Learning (FL) is needed to overcome key challenges arising from the AI Act*, especially with regard to data governance. Through careful analysis of the AI Act from a technical perspective, we show that the distributed architecture of FL inherently addresses regulatory requirements around data privacy, consent-based processing, and computational resource allocation. We critically examine the current shortcomings of FL in the context of the AI Act and map out research priorities that are needed to on the path towards full regulatory compliance.

1 Introduction

The rapid advancement of AI has prompted increased regulatory scrutiny, particularly in the European Union (EU). The EU AI Act (hereafter referred to as AI Act) represents a landmark piece of legislation establishing comprehensive requirements for AI systems, emphasizing high-risk applications and general-purpose AI models [15]. This regulatory framework aims to ensure that AI systems are developed and deployed with transparency, accountability, and societal values in mind.

The AI Act is the first comprehensive legislation among many others that are already underway in other regions like Canada [41] or China [11]. While the US Presidential Executive Order on Trustworthy AI has been rescinded [74], there are several bills planned on the US state-level, e.g., in New York State [64] or Texas [73]. These planned bills and regulatory frameworks adopt requirements closely related to the AI Act.

Thus, we focus on the AI Act as it currently is the only comprehensive AI legislation that has been passed into law. The AI Act introduces stringent requirements across three key dimensions of data governance, safety, and security, as well as energy efficiency and sustainability. In this context, the AI Act distinguishes between low/medium-risk and high-risk applications, where the latter is subject to extensive monitoring and documentation requirements. An example of a high-risk application is a system that handles financial transactions or even a job application screening tool, i.e., any system that can have a fundamental impact on an individual's life. Furthermore, the AI Act also defines a notion of general-purpose AI (GPAI) systems, which host capable models that can serve multiple tasks at a time.

Traditional centralized learning approaches, which dominate today's AI development, face substantial challenges under this new regulatory regime. These approaches typically rely on large-scale data collection through web crawling [75] and centralized data storage [34], raising significant concerns regarding data privacy, copyright compliance, and regulatory adherence [67]. Additionally, centralized systems create resource bottlenecks due to their high energy demand in concentrated locations.

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

In the EU, power grids are operating at capacity, and initiatives to significantly grow the electricity transmission capacities often take several years [21]. This makes the installation of new large-scale AI data centers within the EU in the short- and mid-term difficult, requiring solutions to bridge the resource gap.

We take the position that **Federated Learning (FL) is needed to overcome key challenges arising from the AI Act**, especially for high-risk applications.

Such high-risk applications are typically characterized by extensive process integrity and safety requirements. The localized learning architecture of FL ensures that training data stays on the owners' premises, making FL particularly attractive for scenarios where sensitive personal data or process information is handled [39]. Normally, it takes time to obtain data processing clearance for applications that touch upon trade secrets or export-controlled goods. One such example is automotive semiconductor chip production processes. FL enables learning on such data without ever transferring the raw data outside a production facility but rather leverages on-premise resources to learn a model with relevant information. Such high-risk systems are typically found in already highly regulated environments and typically involve formal agreements between clients and with the server operator. Since there is often only a limited number of clients that contribute relevant data for high-risk applications, it is highly likely that systems will be built with a cross-silo architecture. On a more general note, the owner can decide what data can be used in training, remaining in control of their data at all times. This positively impacts copyright compliance since learning is only conducted on consent-provided data [67].

At the same time, localized learning benefits the reduction of data-induced biases in models. End users can contribute their data to the training process, improving the service quality not only for themselves but also for peers who have similar requirements and interests. The greater flexibility of FL compared to centralized learning when it comes to personalized models boasts the utility of models for downstream tasks and helps users increase the effectiveness of models beyond what is possible with off-the-shelf models [86]. This contributes to meeting the data governance requirements of the AI Act.

Since EU power grids are operating at their limits, especially concerning transmission capacities [21], there is a stark need for methods that can balance loads based on resource availability and train models across wide-area networks, potentially with significantly lower communication bandwidths than what is available in data centers. This improves the overall resource efficiency as capacities are used where they are available, and hardware is utilized more effectively, which aligns with the AI Act's energy monitoring and efficiency requirements. Despite these critical benefits of FL that enable the development of AI models for sensitive processes, there is a set of open challenges that currently hinder the broad adoption of FL. The FL research community needs to address training efficiency, improve the regulatory compliance of secure computing techniques, and further enhance federated monitoring of training and deployment processes.

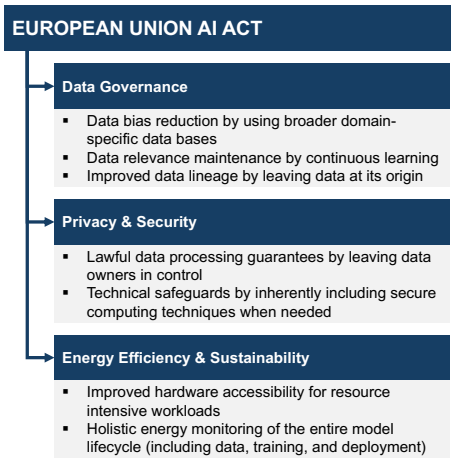

Figure 1: Overview of FL properties that inherently improve regulatory compliance with the AI Act.

In Section 2, we begin by introducing the AI Act cornerstones. We then discuss data governance specifics in Section 3, followed by privacy and security in Section 4. Section 5 introduces energy efficiency considerations. For a broader perspective, we offer an alternative position based on open FL challenges in Section 6. We advocate for future research priorities in Section 7 and conclude in Section 8.

2 The EU AI Act

The AI Act comes into effect in multiple stages, with the rules on prohibited applications (e.g., AI for biometrics) applying from February 2025. In August 2025, the provisions on GPAI systems will take

effect. The rules for high-risk applications will enter into force in August 2026. The final part of the AI Act, Art.[1] 6(1), provisions for systems where AI is used for safety features will come into effect in 2027. The detailed timeline is available in Art. 113.

2.1 Regulatory Risks for Technical Innovation

The inherent tension between regulatory frameworks and technical implementation presents a significant challenge in emerging technologies. While regulations deliberately employ broad, abstract requirements to prevent circumvention and maintain adaptability across different contexts, this approach often creates a fundamental disconnect with technical research priorities, which require concrete, measurable specifications for effective development and validation.

The insufficient alignment between technical priorities and legal requirements has led to a significant implementation gap that demands attention and rapid action. In the first step, it is necessary to understand the current compliance shortcomings from a technical perspective, identify techniques that have the smallest gaps, and map out a path towards full compliance. Second, progressing frameworks like the AI Act must be a joint effort between technical communities and regulators to balance control and the ability to innovate.

For instance, current systems often face competing demands between privacy protection and energy efficiency (see Section 6.2), where strengthening one aspect necessarily compromises the other. This complex interplay of requirements necessitates a pragmatic framework that can guide technical teams in making informed decisions while maintaining regulatory compliance. Depending on the field of application and the computational resources required for a model, the AI Act sets forth different compliance criteria.

2.2 System Assessment Criteria

The AI Act focuses on the real-world impact of AI models on the daily lives of people inside the EU via a risk-based assessment and a more specific, yet still broadly defined, notion of GPAI.

(I) GPAI. Currently, an AI model is classified as GPAI once it meets one of the following three criteria (Art. 51): (i) high impact capabilities, (ii) by decision of the EU regulator, or (iii) whenever a model surpasses a total training compute budget of 10^{25} floating point operations (FLOP). Yet, the regulation has yet to precisely describe what systemic risk is and *how* to measure impact capability. The legal text only contains a general notion to measure the risk by means of "appropriate technical tools and methodologies, including indicators and benchmarks" (Art. 51). It is unclear whether current benchmarks address any of the high-level AI Act requirements. As there is a lack of specific legal criteria for GPAI, this paper focuses on the analysis of domain- and task-specific high-risk applications.

(II) Risk-based assessment. We expect FL to excel in high-risk applications due to the inherent need for strict data governance and process integrity. The AI Act distinguishes AI systems into risk categories that require service providers to adhere to different compliance criteria. High-risk systems are characterized by having the potential to have a major impact on essential services (Art. 6; e.g., CV-sorting software for job application websites or AI-supported medical products). These systems have the largest number of compliance items regarding transparency, data governance, results traceability, data security, privacy, and possibly energy reporting (Art. 40). Furthermore, the AI Act requires high-risk application providers to register their systems in the *EU Database for High-Risk AI Systems* (Art. 71). For a full picture, limited risk systems are subject to transparency regulation only, i.e., AI-generated results must be labeled as such. Minimal and no-risk systems are not subject to regulation but can voluntarily adopt the AI Act requirements.

The following analysis in Sections 3 to 5 discusses several AI Act requirements and how FL can help achieve full compliance, especially for *high-risk* applications. See Figure 1 for a summary of key benefits of FL under the AI Act.

[1] "Art." and "Rec." are the abbreviations for Articles and Recitals in the AI Act, respectively.

3 Improved Data Accessibility Addresses the Key Data Governance Concerns

The AI Act establishes comprehensive requirements for training, validation, and testing datasets as outlined in Art. 10 and further elaborated in Rec. 44-47. The regulation mandates that these datasets meet specific quality criteria: they must be relevant to the intended purpose, representative of the use case, demonstrably free of errors, and complete in their coverage of relevant scenarios. A distinctive feature of these requirements is the explicit consideration of geographical, behavioral, and functional settings specific to the system's intended purpose, ensuring that the AI system performs consistently across diverse operational contexts. This naturally favors paradigms that can tap into diverse data sources. The AI Act emphasizes anti-discrimination safeguards, requiring providers to thoroughly examine datasets for potential biases that could lead to discriminatory outcomes. This examination process must be systematic and documented, specifically identifying and mitigating potential sources of unfair bias that could affect protected characteristics or vulnerable populations. Taken together, the data collection requirements and safeguards constitute the definition of high-quality data.

3.1 FL Simplifies Access to Domain Data for High-Risk AI Applications

As data owners remain in control over their data with FL, the barriers to making (subsets of) data available for training domain-specific models can be reduced. For instance, the high-tech and biotech industries, with their sensitive workloads, have many untapped valuable data sources for domain-specific foundation models [22, 36]. Data from these domains is often considered high-risk as misuse could create significant threats, e.g., in the context of material or drug discovery. With FL, foundation models for drug or material discovery can be trained on a greater variety of data points. The importance of data variety has proven invaluable for material and biosciences as it significantly improves the model quality [28, 55, 72, 77].

Taken together, the sensitivity of workloads, the increasing demand for data, and the fragmentation of data across institutions can be a limiting factor for advancing material sciences and drug discovery initiatives. FL enables data owners to retain control over their data and decide what data to share for training. This provides a strict data lineage as required under the AI Act and a much broader data basis for high-risk, domain-specific models. FL offers several optimization and aggregation techniques that, paired with a more diverse data basis, address quality and relevance bias [42], as required by Art. 11.

3.2 Training at the User Base Directly Addresses the AI Act Data Relevancy Requirement

FL inherently enables continuous learning [84], keeping models up-to-date with the latest data. This improves overall data relevancy [8, 53] and model performance notably [27, 83], as required by the AI Act under Art. 10.

In practice, the AI Act requires organizations developing high-risk AI systems to establish clear criteria for determining data relevance, document their assessment processes for data quality and relevance, maintain records of how they ensure ongoing data appropriateness, and regularly review and update datasets to maintain relevance. Frequent progress leading to quickly evolving databases and changing user requirements necessitates quick and intuitive update processes. For AI applications, this typically involves training data updates, re-training of models, and re-deployment. FL can mitigate the overhead that comes along with these additional steps.

3.3 FL Simplifies Data Lineage as Required by the AI Act

At the same time, FL offers an intuitive approach to simplifying the data lineage when training AI models compared to centralized learning. High-risk applications require thorough documentation of data processing steps, i.e., which data points have been processed at which location and what process steps have been applied. FL can address the first part of the data lineage requirements since data is always left at the client and never moved, which notably simplifies the record keeping process.[2] Regarding processing steps, trusted execution environments offer a viable solution to ensuring process integrity even in distributed environments, rendering FL on par with centralized learning. Similarly, communicating sensitive raw data creates a risk for man-in-the-middle attacks [13]; centralized data storage and humans in the loop open vulnerabilities for unauthorized data access and processing [26]. In fact, data breaches are the main reason for substantial data-related fines in the EU [12]. FL substantially reduces the risk of unlawful data processing in this context.

[2]We note that, contrary to cross-silo FL, cross-device settings can create additional complexities for data lineage when client contributions have to be tracked across a large number of clients ($n > 1000$).

4 FL Enacts Privacy & Security by Design

As privacy violations are the main cause of data-related fines, it is imperative to understand the notions of regulatory and technical privacy. From a legal perspective, privacy is defined as lawful and consent-based data processing, i.e., private data must only be processed for a specific intent with the prior consent of the data owner [14]. More specifically, the AI Act establishes privacy guardrails by installing extensive documentation, monitoring, and reporting requirements that help prevent unauthorized data processing (Art. 10). Technical privacy, typically enacted by secure computing techniques (e.g., differential privacy), specifically addresses data breach risks and is considered a technical safeguard supporting regulatory privacy (Art. 15). That said, the AI Act complements GDPR and its key requirements [14].

4.1 FL Reinforces Lawful Data Processing in High-Risk Applications

The design of FL applications removes the requirement to collect data in a central location for training and, therefore, enables data owners to share insights about their data without exposing the actual raw data that can be of a sensitive nature. Keeping data owners in control also puts a stronger focus on consent-based data processing, further improving regulatory compliance as required by the AI Act under Art. 10.

From a regulatory perspective, this also creates stricter boundaries of how data is being processed. Since FL applications typically have a specific use case, they use training data only in this context. Also, clients can disconnect (opt-out) from an FL application at any time, a very important characteristic needed for compliance with privacy regulation. This removes regulatory exposure for one of the major origins of significant fines: consent-based processing [12].

On a more general note, under the AI Act, access control mechanisms require strict authentication protocols and a clear allocation of responsibilities for data handling while ensuring appropriate human oversight of the system (Art. 29). The AI Act further strengthens data safety through logging requirements (Art. 12), mandating the automatic recording of critical events, including data modifications, access attempts, system changes affecting data integrity, and security anomalies. Art. 62 and 64 establish a comprehensive incident management framework, requiring providers to report serious security incidents, maintain robust incident response procedures, and implement corrective measures based on a systematic analysis of security breaches. Thus, the emphasis of FL on keeping data at its origin constitutes an inherent advantage over centralized learning since there are fewer data leakage risks in the data flow.

4.2 Secure Computing Is an Inherent Component of FL When Needed

Private computing methods that have been integrated with FL are a technical safeguard against data breaches that may occur whenever a model or client model updates are leaked to unauthorized third parties, as required by Art. 10 and 15. In FL environments, model updates transmitted by clients are inherently susceptible to gradient inversion and membership inference attacks, potentially exposing sensitive training data [43]. However, extensive research efforts have successfully addressed these vulnerabilities through two primary approaches: perturbation-based techniques implementing (ϵ, δ)-Differential Privacy ((ϵ, δ)-DP) guarantees [20, 57] and cryptographic methods such as homomorphic encryption (HE) [38] and secure multi-party computation (SMPC) [7]. While these technical solutions may not directly address privacy in a regulatory context, they establish a robust technical foundation that substantially reduces potential attack vectors and mitigates the risk of financial liability under Art. 101, which outlines penalties for non-compliance with the AI Act. It is important to note that anonymization techniques are particularly effective in large-scale databases, which conceal sensitive information at a much lower cost than perturbation or cryptographic methods [5, 31, 33]. Overall, this can provide a stronger compliance basis compared to centralizing learning as data is encrypted, anonymized, or made private directly at the client side and processed in a distributed fashion.

4.3 Cross-Silo FL for High-Risk AI Applications Removes Many Attack Vectors Compared to Cross-Device FL

High-risk FL systems under the AI Act will most likely be designed as cross-silo applications, given the already highly regulated nature of high-risk domains. Thus, there will be a small number of clients (typically < 100), each with a considerable amount of data. This setup allows for contractual agreements between all participating clients and the server operator, which can increase the trust

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. Further details on each technique are available in Appendix D. Key: N = number of model parameters, K = total number of clients in system.

Privacy Technique	Pot. AI Act compliant*	Client Computation	Client Communication	Client Space	Server Computation	Server Communication	Server Space	Algorithm
(ϵ, δ)-DP	✓	$\mathcal{O}(N)$**	$\mathcal{O}(1)$	$\mathcal{O}(N)$	$\mathcal{O}(\lvert K\rvert)$	$\mathcal{O}(\lvert K\rvert)$	$\mathcal{O}(\lvert K\rvert)$	Andrew et al. [1]
SMPC	✓	$\mathcal{O}(\lvert K\rvert^2 + \lvert K\rvert \times N)$	$\mathcal{O}(\lvert K\rvert + N)$	$\mathcal{O}(\lvert K\rvert + N)$	$\mathcal{O}(\lvert K\rvert^2 \times N)$	$\mathcal{O}(\lvert K\rvert^2 + \lvert K\rvert \times N)$	$\mathcal{O}(\lvert K\rvert^2 + N)$	Bonawitz et al. [7]
HE	Limited	$\mathcal{O}(N)$	$\mathcal{O}(N)$	$\mathcal{O}(N)$	$\mathcal{O}(\lvert K\rvert \times N)$	$\mathcal{O}(\lvert K\rvert \times N)$	$\mathcal{O}(N)$	Jin et al. [46]

* Potential evaluation for future AI Act compliance ** $\mathcal{O}(N)$ 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 $\mathcal{O}(\lvert K\rvert \times N)$

between parties and formalize contribution requirements per client. Such contribution requirements can include a minimum amount of training data points per client. Given the regulated nature, there are already standards in place the act as data quality gates and help standardize the data format across organizations. Also, contractc can be used to negotiate minimum infrastructure requirements. As such, tools like Trusted Execution Environments (TEEs) can be used across all clients [60]. Furthermore, key security concerns such as Byzantine attacks (e.g., data and model poisoning) [54, 82], membership inference attacks client model updates [62], or freerider attacks [30] can be mitigated largely through formal agreements.

5 Energy & Sustainability Considerations Are Central Components of Regulation

Aside from data governance-related requirements, the AI Act integrates fundamental environmental and sustainability considerations. Rec. 69 and 76 provide the foundational context, emphasizing lifecycle environmental assessment and acknowledging the growing ecological footprint of AI. These principles are operationalized through several key articles: Art. 17 mandates systematic documentation of resource consumption within quality management systems. It requires detailed environmental impact assessments for high-risk AI systems, and Art. 61, in connection with Rec. 142, promotes voluntary sustainability initiatives through codes of practice. Together, these provisions create a regulatory architecture that combines mandatory environmental reporting with voluntary industry initiatives, reflecting the EU's broader commitment to technological advancement within ecological constraints.

Environmental provisions of the AI Act represent an important step toward sustainable AI development, though their practical effectiveness will depend on implementation and enforcement mechanisms. Generally, this creates an inevitable trade-off between the need for privacy (Art. 10) and energy efficiency (see Section 6.2). The trade-off is particularly relevant in light of the EU power grid condition. The grids of member states operate at capacity, limiting the deployment of large-scale computational facilities, particularly data centers required for advanced machine learning operations.

5.1 FL Can Improve the Accessibility to Distributed Training Hardware

With FL, we improve the accessibility of scattered resources by leveraging hardware within the trusted perimeters of data owners and decentralizing energy requirements. A frequent use case for FL is medical data processing. For example, Germany, the EU's largest economy, has 1,872 hospitals with 17.6M patient admissions in 2023 [29]. Each hospitalization produces a separate record. When assuming the average length on a single record is similar to those found in the MIMIC-IV dataset [47], German hospitals have generated roughly 39.9B tokens in 2023 that can be used for model training. Typically, domain-specific language models are based on pre-trained models such as Llama 3.1 8B [28, 34] and fine-tuned on domain knowledge. Facilitating the fine-tuning process on 2023 German patient records requires approx. 57 GPU days on Nvidia H100 GPUs (see Appendix A for a full calculation), an amount of compute that either requires the use of cloud services or a sophisticated and fault-tolerant distributed processing architecture. For sensitive workloads, cloud outsourcing is typically challenging as numerous regulatory clearance processes are involved when moving data [50]. In fact, the European Health Data Space (EHDS), a new regulation adopted by the EU Council in January 2025, aims to provide controlled access to patient data of EU residents in a secure, anonymous, and intuitive way [16]. Taken together, the EHDS and AI Act draw a path to federated data processing for high-risk systems in the field of medical data[3]. The same is true

[3]We note that there are several practical AI deployment challenges mainly originating from budgeting constraints, bureaucracy, and resource availability that have to be solved before AI can be used in hospitals at scale.

for other domains where applications are likely falling into the high-risk category, such as financial transactions or biotech.

5.2 FL Can Improve Energy Consumption Transparency

In FL systems, data processing occurs at the point of origin, providing direct insights into the actual computational and communication costs associated with data collection and processing. This transparency becomes particularly relevant when considering the AI Act's emphasis on environmental impact documentation and resource efficiency. The stark contrast in energy requirements becomes evident when comparing different data collection scenarios: collecting data from remote sensing devices via wireless wide-area networks incurs energy costs that can be orders of magnitude higher than data collection within centralized data centers. This granular visibility of energy consumption patterns aligns with the AI Act's requirements for the systematic recording of energy metrics and environmental impact assessment, potentially facilitating compliance with regulatory frameworks while enabling more accurate optimization of system-wide energy efficiency. To fully capture energy consumption in FL systems, we can use the client energy consumption, the per-bit communication cost model [45, 76] and pair it with trusted computing techniques [61] on the client side such that the data lineage also covers the energy footprint of a data point. While monitoring of distributed systems is more complex compared to centralized systems, the inherent transparency of resource utilization in FL systems not only aids in regulatory compliance but also provides a starting point for energy-based optimization of AI systems.

6 Alternative View: Centralized Learning Is Overall More Efficient and Offers Fewer Attack Vectors

Despite the beneficial properties of FL in light of the AI Act, there are several practical challenges that currently limit the wider adoption of FL. Centralized learning is currently the primary approach for training ML models, mainly because the operator (e.g., Meta or DeepSeek) remains in full control of the entire training pipeline [17, 34]. This provides the operator with fine-grain control over the data that is being used for training. Similarly, it is easier to integrate training optimizations (e.g., via hardware-software co-design) when all training hardware is owned by the operator [32].

In the following, we use theoretical analysis and experimental results to outline a set of key challenges where centralized learning is favorable over FL. Our experimental results are based on a federated training pipeline as it is frequently used in FL literature [3, 40, 69] but with dedicated hardware. We train a BERT-based transformer model [19] for email classification on the 20 News Group Dataset [51] that is already sufficient to outline the current shortcomings of FL. From a practical perspective, such a training pipeline can also be employed for job application screening applications that fall under the AI Act definition of high-risk systems. Experimental details are available in Appendix B.

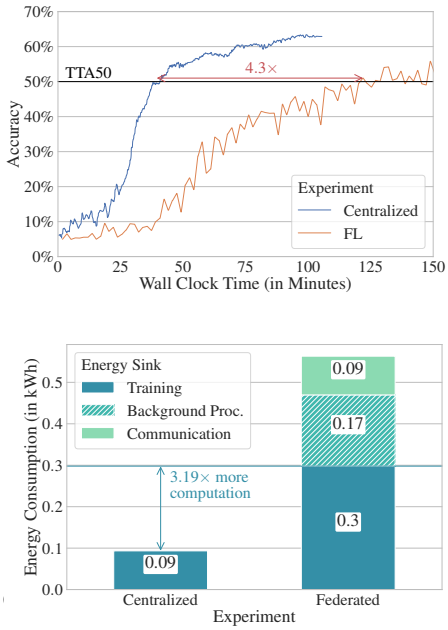

Figure 2: Energy-efficiency experiments. Quantification of energy sinks in FL applications compared to centralized learning. We use a target accuracy of 50% (TTA50) and train over 4000 mini-batches in each scenario.

6.1 Energy Efficiency in FL Systems Is Worse than in Centralized Learning

FL exhibits an inevitable tradeoff between learning and communication efficiency. More local training steps in between client-server communication rounds yield better communication efficiency but can also lead to divergent models. Such a divergence significantly reduces training efficiency. Such a trade-off is challenging to control and optimize [10]. In Figure 2, we showcase the impact of

federated communication over a wide-area network. Even in such small-scale experiments with communication after every two local training steps, the FL performance disadvantage compared to centralized learning is notable. Note, communication frequency in real-world systems is often significantly lower than in our experiment, leading to stronger non-i.i.d. effects, longer training times, and overall higher energy consumption [44, 65].

6.2 Private Computing Techniques Create Significant Overheads

Our theoretical analysis shows that secure computing techniques such as (ϵ, δ) Differential Privacy ((ϵ, δ)-DP), Secure Multi-Party Communication (SMPC), and Homomorphic Encryption (HE) come at notable cost overheads, creating a trade-off between data security and energy efficiency (Table 1). While (ϵ, δ)-DP shows favorable properties in scaled systems, smaller-sized applications require high perturbation levels to ensure no individual data points can be revealed, which can be impractical. Generally, this leads to significantly increased computational requirements as more training steps are required to reach the same model performance as without (ϵ, δ)-DP. Strong (ϵ, δ)-DP guarantees $(\epsilon < 1)$ can lead to prohibitively long training times as the utility of training samples degrades with increased noise. This creates two challenges for FL applications. First, in systems with clients that participate infrequently or only once, the learning effect from individual clients can be minimal. This may remove some of the benefits FL can have by opening data siloes for higher data variety and deny building more representative models. Second, the energy efficiency of FL systems decreases. Given the already worse efficiency compared to centralized learning, this increases the performance gap to potentially impractical levels. Similarly, Secure Multi-Party Communication and Homomorphic Encryption can also introduce notable cost overheads through extensive communication and computation requirements that depend on the number of model parameters and the number of FL clients. When used with billion-parameter-scale models, both cryptographic methods can reduce the utility of an FL application significantly. In addition, HE also denies the FL server operator from inspecting the model, which can violate data governance requirements of the AI Act, as Art. 72 requires model/service quality monitoring.

6.3 The Right to Be Forgotten Requires Sharing Gradients with Specific Information

Compared to centralized training regimes, the distributed and aggregate nature of federated learning makes it difficult to isolate and remove the influence of specific training data since model updates are intertwined across multiple clients and training rounds. Work towards solving federated unlearning under consideration of communication efficiency has shown promising progress [37]. Yet, a key concern remains. If a server stores the global model over multiple FL rounds and a client requests to be forgotten, the client submits an update that has been updated so that the client data is removed exactly as the client had requested. However, applying such updates that remove certain data points or patterns bears the risk of gradient inversion, a major security vulnerability of FL that is mainly treated with secure computing. At scale, the only viable option to tackle gradient inversion is (ϵ, δ)-DP, and it is unclear whether unlearning is effective in combination with perturbation-based methods. This constitutes a privacy risk. More generally, there is notable progress in pattern unlearning for centralized systems [66, 71]. Yet, unlearning implicitly learned and more complex concepts is still overall challenging and, in wide areas, unsolved [87].

7 Future Federated Learning Research Agenda

Taking the benefits and challenges of FL together, we find that especially data governance calls for methods to increase the data variety when training high-risk applications under the AI Act and for moving training closer to the end user, i.e., becoming more adaptive towards changing environments. In this context, the design of FL offers notable benefits over centralized learning. Yet, our analysis and the alternative view reveals fundamental trade-offs that currently hinder the broader adoption of FL. To address these shortcomings and enhance regulatory compliance of FL with the AI Act and other emerging global regulations, we formulate a non-exhaustive set of research priorities for the FL community (Figure 3).

Data governance. A central component that bears significant risks for legal fines is the right to be forgotten. Generally, but especially in federated settings, unlearning techniques need further attention, particularly when it comes to implicit concepts, e.g., a hidden relationship between two individuals discovered from message threads. Such concepts typically have a deep interconnect that

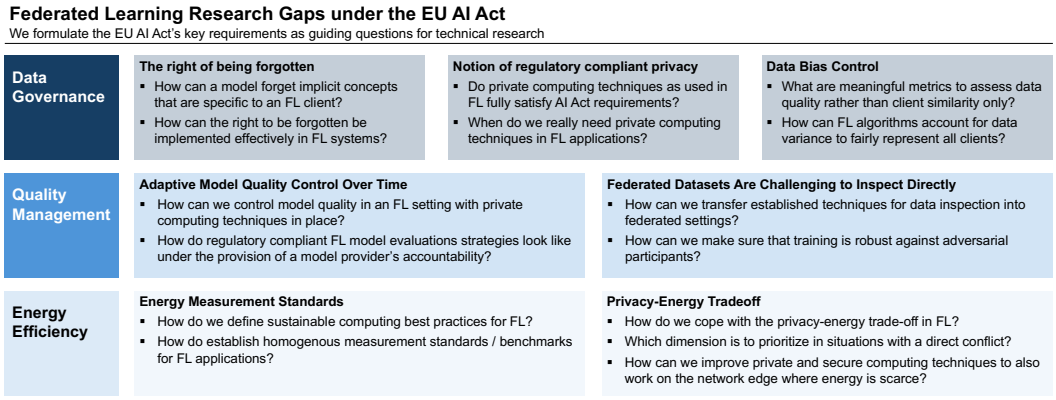

Figure 3: For each analysis section in this position paper (Sections 3 to 5), we formulate a set of open technical research questions based on the AI Act requirements. The FL community must address these questions to enhance regulatory compliance.

is challenging to separate [70]. Additionally, there is a notable gap between the perception of privacy in technical communities and the regulatory definition. While user consent constitutes privacy-conforming data processing, technical privacy can establish strong safeguards against unauthorized data access. However, in the case of (ϵ, δ)-DP, it is unclear what privacy budget is sufficient to comply with the AI Act. Lastly, FL improves data accessibility, but at the same time, data audits become more challenging since accounting for a broad range of bias sources can be difficult at scale [59]. It is also an open question of how to capture the various biases and balance them.

Quality management. Ensuring data quality and integrity across decentralized participants requires robust monitoring processes. Providing meaningful human oversight as required by the AI Act, thorough testing and validation, and detailed technical documentation necessitates coordination and standardized practices among entities. While there are some works providing guidance on how AI service providers can steer the FL training process and take ownership [80], there are still many open research questions, such as the technical implementation of federated data inspection. Often, data inspection can be realized when obtaining user consent and applying anonymization to remove personal data. Further, without a strong understanding of the data basis, it is challenging to generate explanations for federated learning model outputs [4]. Designing systems with data protection safeguards while maintaining model performance is complex and involves numerous trade-offs. For instance, data security is a key concern to providing high service quality, and, to date, it is unclear what a regulatory-compliant secure computing strategy could look like, especially considering the privacy-energy trade-off.

Energy efficiency. While we see that learning efficiency on clients is on par with data center clients [79], the effectiveness of FL systems needs to improve overall, especially when it comes to the combination of federated and private computing techniques. However, we also need a common understanding of contributes to even better energy transparency, i.e., what aspects are relevant, only the training compute or also the way we collect and preprocess data? With their per-bit energy cost model, Jalali et al. [45] have proposed a good starting point, but its practical applicability with many stakeholders involved (e.g., ISP or data center operators) is limited.

8 Conclusions

In this position paper, we evaluate the key requirements of the AI Act from a technical perspective and discuss how FL can help build legally compliant high-risk AI applications. Despite the notable benefits, we also highlight the current shortcomings of FL in light of AI regulation, outlining relevant research challenges going forward. Now is the time to act by assessing operational systems and re-considering the fundamental design of future high-risk AI applications. Furthermore, the EU AI Office has released a call to co-create regulatory and technical implementations of the AI Act [63]. This creates a great opportunity for researchers, lawmakers, and both technical and legal practitioners to work together to shape the future of AI.

References

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

[2] D. Appelhans and B. Walkup. 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, Nov. 2017. doi: 10.1145/3148226.3148232. URL http://dx.doi.org/10.1145/3148226.3148232.

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

[4] J. L. C. Bárcena, M. Daole, P. Ducange, F. Marcelloni, A. Renda, F. Ruffini, and A. Schiavo. Fed-xai: Federated learning of explainable artificial intelligence models. In *CEUR Workshop Proceedings*, 2022. URL https://ceur-ws.org/Vol-3277/paper8.pdf.

[5] R. Bayardo and R. Agrawal. Data privacy through optimal k-anonymization. In *21st International Conference on Data Engineering (ICDE'05)*, pages 217–228, 2005. doi: 10.1109/ICDE.2005.42.

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

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

[8] K. A. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. M. Kiddon, J. Konečný, S. Mazzocchi, B. McMahan, T. V. Overveldt, D. Petrou, D. Ramage, and J. Roselander. Towards federated learning at scale: System design. In *SysML 2019*, 2019. URL https://arxiv.org/abs/1902.01046.

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

[10] S. Caldas, S. M. K. Duddu, P. Wu, T. Li, J. Konečný, H. B. McMahan, V. Smith, and A. Talwalkar. Leaf: A benchmark for federated settings, 2018. URL https://arxiv.org/abs/1812.01097.

[11] China Law Translate. Interim measures for the management of generative artificial intelligence services, jul 2023.

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

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

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

[15] Council of the European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)Text with EEA relevance., jul 2024. URL https://eur-lex.europa.eu/eli/reg/2024/1689/oj. Document 32024R1689.

[16] Council of the European Union. European health data space: Council adopts new regulation improving cross-border access to eu health data, 2025. URL https://www.consilium.europa.eu/en/press/press-releases/2025/01/21/european-health-data-space-council-adopts-new-regulation-improving-cross-border-access-to-eu-health-data/.

[17] DeepSeek-AI. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, 2025. URL https://arxiv.org/abs/2501.12948.

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

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

[20] C. Dwork and A. Roth. 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.

[21] EU Agency for the Cooperation of Energy Regulations. Transmission capacities for cross-zonal trade of electricity and congestion management in the eu, jul 2024. URL https://www.acer.europa.eu/monitoring/MMR/crosszonal_electricity_trade_capacities_2024.

[22] European Centre for the Development of Vocational Training. Employment growth in high-tech economy, 2022. URL https://www.cedefop.europa.eu/en/tools/skills-intelligence/employment-growth-high-tech-economy.

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

[24] 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.

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

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

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

[28] Y. Fang, X. Liang, N. Zhang, K. Liu, R. Huang, Z. Chen, X. Fan, and H. Chen. Mol-instructions: A large-scale biomolecular instruction dataset for large language models. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=Tlsdsb6l9n.

[29] Federal Statistical Office of Germany. Medical facilities, hospital beds and movement of patient, 2024. URL https://www.destatis.de/EN/Themes/Society-Environment/Health/Hospitals/Tables/gd-hospitals-years.html.

[30] Y. Fraboni, R. Vidal, and M. Lorenzi. Free-rider attacks on model aggregation in federated learning. In A. Banerjee and K. Fukumizu, editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 1846–1854. PMLR, 13–15 Apr 2021. URL https://proceedings.mlr.press/v130/fraboni21a.html.

[31] A. Gadotti, L. Rocher, F. Houssiau, A.-M. Creţu, and Y.-A. de Montjoye. Anonymization: The imperfect science of using data while preserving privacy. *Science Advances*, 10(29):eadn7053, 2024. doi: 10.1126/sciadv.adn7053. URL https://www.science.org/doi/abs/10.1126/sciadv.adn7053.

[32] Gemma Team, T. Mesnard, C. Hardin, R. Dadashi, S. Bhupatiraju, S. Pathak, L. Sifre, M. Rivière, M. S. Kale, J. Love, P. Tafti, L. Hussenot, P. G. Sessa, A. Chowdhery, A. Roberts, A. Barua, A. Botev, A. Castro-Ros, A. Slone, A. Héliou, A. Tacchetti, A. Bulanova, A. Paterson, B. Tsai, B. Shahriari, C. L. Lan, C. A. Choquette-Choo, C. Crepy, D. Cer, D. Ippolito, D. Reid, E. Buchatskaya, E. Ni, E. Noland, G. Yan, G. Tucker, G.-C. Muraru, G. Rozhdestvenskiy, H. Michalewski, I. Tenney, I. Grishchenko, J. Austin, J. Keeling, J. Labanowski, J.-B. Lespiau, J. Stanway, J. Brennan, J. Chen, J. Ferret, J. Chiu, J. Mao-Jones, K. Lee, K. Yu, K. Millican, L. L. Sjoesund, L. Lee, L. Dixon, M. Reid, M. Mikuła, M. Wirth, M. Sharman, N. Chinaev, N. Thain, O. Bachem, O. Chang, O. Wahltinez, P. Bailey, P. Michel, P. Yotov, R. Chaabouni, R. Comanescu, R. Jana, R. Anil, R. McIlroy, R. Liu, R. Mullins, S. L. Smith, S. Borgeaud, S. Girgin, S. Douglas, S. Pandya, S. Shakeri, S. De, T. Klimenko, T. Hennigan, V. Feinberg, W. Stokowiec, Y.-h. Chen, Z. Ahmed, Z. Gong, T. Warkentin, L. Peran, M. Giang, C. Farabet, O. Vinyals, J. Dean, K. Kavukcuoglu, D. Hassabis, Z. Ghahramani, D. Eck, J. Barral, F. Pereira, E. Collins, A. Joulin, N. Fiedel, E. Senter, A. Andreev, and K. Kenealy. Gemma: Open models based on gemini research and technology, 2024. URL https://arxiv.org/abs/2403.08295.

[33] G. Ghinita, P. Karras, P. Kalnis, and N. Mamoulis. Fast data anonymization with low information loss. In *Proceedings of the 33rd International Conference on Very Large Data Bases*, VLDB '07, page 758–769. VLDB Endowment, 2007. ISBN 9781595936493.

[34] A. Grattafiori, A. Dubey, et al. The llama 3 herd of models, 2024. URL https://arxiv.org/abs/2407.21783.

[35] S. Gupta, A. Agrawal, K. Gopalakrishnan, and P. Narayanan. Deep Learning with Limited Numerical Precision. In F. R. Bach and D. M. Blei, editors, *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*, pages 1737–1746. JMLR.org, 2015. URL http://proceedings.mlr.press/v37/gupta15.html.

[36] A. Haaf, S. Hofmann, and J. Schüler. Measuring the economic footprint of the biotechnology industry in europe. Technical report, WifOR Institute, 2020. URL https://www.cedefop.europa.eu/en/tools/skills-intelligence/employment-growth-high-tech-economy.

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

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

[39] M. Hao, H. Li, X. Luo, G. Xu, H. Yang, and S. Liu. Efficient and privacy-enhanced federated learning for industrial artificial intelligence. *IEEE Transactions on Industrial Informatics*, 16 (10):6532–6542, Oct. 2020. ISSN 1941-0050. doi: 10.1109/tii.2019.2945367. URL http://dx.doi.org/10.1109/TII.2019.2945367.

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

[41] 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.

[42] T. Huang, W. Lin, W. Wu, L. He, K. Li, and A. Y. Zomaya. An efficiency-boosting client selection scheme for federated learning with fairness guarantee. *IEEE Transactions on Parallel and Distributed Systems*, 32(7):1552–1564, July 2021. ISSN 2161-9883. doi: 10.1109/tpds.2020.3040887. URL http://dx.doi.org/10.1109/TPDS.2020.3040887.

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

[44] D. Huba, J. Nguyen, K. Malik, R. Zhu, M. Rabbat, A. Yousefpour, C.-J. Wu, H. Zhan, P. Ustinov, H. Srinivas, K. Wang, A. Shoumikhin, J. Min, and M. Malek. Papaya: Practical, private, and scalable federated learning. In D. Marculescu, Y. Chi, and C. Wu, editors, *Proceedings of Machine Learning and Systems*, volume 4, pages 814–832, 2022. URL https://proceedings.mlsys.org/paper_files/paper/2022/file/a8bc4cb14a20f20d1f96188bd61eec87-Paper.pdf.

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

[46] W. Jin, Y. Yao, S. Han, C. Joe-Wong, S. Ravi, S. Avestimehr, and C. He. FedML-HE: An efficient homomorphic-encryption-based privacy-preserving federated learning system. In *International Workshop on Federated Learning in the Age of Foundation Models in Conjunction with NeurIPS 2023*, 2023. URL https://openreview.net/forum?id=PuYD0fh5aq.

[47] A. Johnson, L. Bulgarelli, T. Pollard, B. Gow, B. Moody, S. Horng, L. A. Celi, and R. Mark. Mimic-iv, 2024. URL https://physionet.org/content/mimiciv/3.1/.

[48] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei. Scaling laws for neural language models, 2020. URL https://arxiv.org/abs/2001.08361.

[49] V. A. Korthikanti, J. Casper, S. Lym, L. McAfee, M. Andersch, M. Shoeybi, and B. Catanzaro. Reducing activation recomputation in large transformer models. In D. Song, M. Carbin, and T. Chen, editors, *Proceedings of Machine Learning and Systems*, volume 5, pages 341–353. Curan, 2023. URL https://proceedings.mlsys.org/paper_files/paper/2023/file/80083951326cf5b35e5100260d64ed81-Paper-mlsys2023.pdf.

[50] J. Lane and C. Schur. Balancing access to health data and privacy: A review of the issues and approaches for the future. *Health Services Research*, 45(5p2):1456–1467, Aug. 2010. ISSN 1475-6773. doi: 10.1111/j.1475-6773.2010.01141.x. URL http://dx.doi.org/10.1111/j.1475-6773.2010.01141.x.

[51] K. Lang. NewsWeeder: Learning to Filter Netnews. In A. Prieditis and S. Russell, editors, *Machine Learning Proceedings 1995*, pages 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.

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

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

[54] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith. Federated optimization in heterogeneous networks. In I. Dhillon, D. Papailiopoulos, and

V. Sze, editors, *Proceedings of Machine Learning and Systems*, volume 2, pages 429–450, 2020. URL https://proceedings.mlsys.org/paper_files/paper/2020/file/1f5fe83998a09396ebe6477d9475ba0c-Paper.pdf.

[55] O. Liu, S. Jaghouar, J. Hagemann, S. Wang, J. Wiemels, J. Kaufman, and W. Neiswanger. Metagene-1: Metagenomic foundation model for pandemic monitoring, 2025. URL https://arxiv.org/abs/2501.02045.

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

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

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

[59] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan. A survey on bias and fairness in machine learning. *ACM Comput. Surv.*, 54(6), July 2021. ISSN 0360-0300. doi: 10.1145/3457607. URL https://doi.org/10.1145/3457607.

[60] F. Mo, H. Haddadi, K. Katevas, E. Marin, D. Perino, and N. Kourtellis. Ppfl: privacy-preserving federated learning with trusted execution environments. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '21, page 94–108, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450384438. doi: 10.1145/3458864.3466628. URL https://doi.org/10.1145/3458864.3466628.

[61] F. Mo, H. Haddadi, K. Katevas, E. Marin, D. Perino, and N. Kourtellis. Ppfl: privacy-preserving federated learning with trusted execution environments. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys'21, page 94–108. Association for Computing Machinery, 2021. doi: 10.1145/3458864.3466628. URL https://doi.org/10.1145/3458864.3466628.

[62] M. Nasr, R. Shokri, and A. Houmansadr. Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In *2019 IEEE Symposium on Security and Privacy (SP)*, page 739–753. IEEE, May 2019. doi: 10.1109/sp.2019.00065. URL http://dx.doi.org/10.1109/SP.2019.00065.

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

[64] New York State Senate. New york artificial intelligence consumer protection act, 2023. URL https://www.nysenate.gov/legislation/bills/2023/S8209.

[65] J. Nguyen, K. Malik, H. Zhan, A. Yousefpour, M. Rabbat, M. Malek, and D. Huba. Federated Learning with Buffered Asynchronous Aggregation . In G. Camps-Valls, F. J. R. Ruiz, and I. Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 3581–3607. PMLR, 28–30 Mar 2022. URL https://proceedings.mlr.press/v151/nguyen22b.html.

[66] T. T. Nguyen, T. T. Huynh, Z. Ren, P. L. Nguyen, A. W.-C. Liew, H. Yin, and Q. V. H. Nguyen. A survey of machine unlearning, 2022. URL https://arxiv.org/abs/2209.02299.

[67] N. Oliver, A. Peukert, et al. First draft general-purpose ai code of practice, nov 2024. URL https://ec.europa.eu/newsroom/dae/redirection/document/109946.

[68] X. Qiu, T. Parcollet, J. Fernandez-Marques, P. P. B. Gusmao, Y. Gao, D. J. Beutel, T. Topal, A. Mathur, and N. D. Lane. 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.

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

[70] C. A. Seger. Implicit learning. *Psychological Bulletin*, 115(2):163–196, 1994. ISSN 0033-2909. doi: 10.1037/0033-2909.115.2.163. URL http://dx.doi.org/10.1037/0033-2909.115.2.163.

[71] A. Sekhari, J. Acharya, G. Kamath, and A. T. Suresh. Remember what you want to forget: algorithms for machine unlearning. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, Red Hook, NY, USA, 2024. Curran Associates Inc. ISBN 9781713845393.

[72] M. Shirts and V. S. Pande. Screen savers of the world unite! *Science*, 290(5498):1903–1904, Dec. 2000. ISSN 1095-9203. doi: 10.1126/science.290.5498.1903. URL http://dx.doi.org/10.1126/science.290.5498.1903.

[73] Texas State Senate. Texas responsible artificial intelligence governance act, 2024. URL https://capitol.texas.gov/tlodocs/89R/billtext/pdf/HB01709I.pdf.

[74] The White House. Initial rescissions of harmful executive orders and actions, 2025. URL https://www.whitehouse.gov/presidential-actions/2025/01/initial-rescissions-of-harmful-executive-orders-and-actions/.

[75] P. Villalobos, A. Ho, J. Sevilla, T. Besiroglu, L. Heim, and M. Hobbhahn. Will we run out of data? limits of llm scaling based on human-generated data, 2022. URL https://arxiv.org/abs/2211.04325.

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

[77] M. D. Ward, M. I. Zimmerman, A. Meller, M. Chung, S. J. Swamidass, and G. R. Bowman. Deep learning the structural determinants of protein biochemical properties by comparing structural ensembles with diffnets. *Nature Communications*, 12(1), May 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-23246-1. URL http://dx.doi.org/10.1038/s41467-021-23246-1.

[78] P. Wiesner, R. Khalili, D. Grinwald, P. Agrawal, L. Thamsen, and O. Kao. Fedzero: Leveraging renewable excess energy in federated learning. In *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*, e-Energy '24, page 373–385, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400704802. doi: 10.1145/3632775.3639589. URL https://doi.org/10.1145/3632775.3639589.

[79] H. Woisetschläger, A. Erben, R. Mayer, S. Wang, and H.-A. Jacobsen. Fledge: Benchmarking federated learning applications in edge computing systems. In *Proceedings of the 25th International Middleware Conference*, Middleware '24, page 88–102, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400706233. doi: 10.1145/3652892.3700751. URL https://doi.org/10.1145/3652892.3700751.

[80] H. Woisetschläger, S. Mertel, C. Krönke, R. Mayer, and H.-A. Jacobsen. Federated learning and ai regulation in the european union: Who is responsible? – an interdisciplinary analysis, 2024. URL https://blog.genlaw.org/pdfs/genlaw_icml2024/16.pdf.

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

[82] A. Yazdinejad, A. Dehghantanha, H. Karimipour, G. Srivastava, and R. M. Parizi. A robust privacy-preserving federated learning model against model poisoning attacks. *IEEE Transactions on Information Forensics and Security*, 19:6693–6708, 2024. ISSN 1556-6021. doi: 10.1109/tifs.2024.3420126. URL http://dx.doi.org/10.1109/TIFS.2024.3420126.

[83] F. Yin, Z. Lin, Q. Kong, Y. Xu, D. Li, S. Theodoridis, and S. R. Cui. Fedloc: Federated learning framework for data-driven cooperative localization and location data processing. *IEEE Open Journal of Signal Processing*, 1:187–215, 2020. ISSN 2644-1322. doi: 10.1109/ojsp.2020.3036276. URL http://dx.doi.org/10.1109/OJSP.2020.3036276.

[84] J. Yoon, W. Jeong, G. Lee, E. Yang, and S. J. Hwang. Federated continual learning with weighted inter-client transfer. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 12073–12086. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/yoon21b.html.

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

[86] H. Zhang, C. Li, W. Dai, J. Zou, and H. Xiong. FedCR: Personalized federated learning based on across-client common representation with conditional mutual information regularization. In A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 41314–41330. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/zhang23w.html.

[87] J. Zhu, B. Han, J. Yao, J. Xu, G. Niu, and M. Sugiyama. Decoupling the class label and the target concept in machine unlearning, 2024. URL https://arxiv.org/abs/2406.08288.

Appendix

A Calculatory Details on the Medical Data Learning Example in Section 5.1

We take the total of 39.9B tokens from 17.6M patient record documents [29] with an average length of 2267 tokens [47]. We use the parameter count of Llama 3.1 8B for our calculations, as it frequently forms the basis for domain-specific large language models [28]. It typically takes a total of 6 FLOP per token per model parameter to train a model [48]. Taken together, the total compute required to train a model at 100% hardware utilization is 1.9×10^{21} FLOP. Accounting for a typical hardware efficiency – measured as model-FLOP-utilization (MFU) – between 30 and 40% [49], we need a total compute of 4.8×10^{21} FLOP. A single Nvidia H100 GPU has a total compute capacity of 9.79×10^{14} FLOP/s, so it takes about 57 GPU days to train an 8B parameter model on 39.9B tokens. For a comprehensive summary, see Table 2.

	Quantity	**Unit**
Dataset size	39,899,200,000	tokens
Llama 3.1 8B	8,000,000,000	parameters
Compute requirement	6	FLOP / token / parameter
Required compute @ 100% efficiency	1.92×10^{21}	FLOP
Hardware efficiency (MFU)	40	%
Total required compute	4.79×10^{21}	FLOP
Nvidia H100 Compute Capacity	9.79×10^{14}	FLOP/s
Total Compute Time	56.6	days

Table 2: Overview of our compute effort calculation for a state-of-the-art medical text dataset.

B Experimental Details for Our Argumentation in Section 6

Table 3: 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 [19] over the 20 News Group Dataset [51] 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 using the 20 Newsgroup Dataset [51], 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 [15], 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 [3, 40, 69].

B.2 Model

We fine-tune the BERT model [19] 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 [3], 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. It 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 [56] 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 [20]. 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 [1, 57]. For the experiment with $z = [0.5; 0.6]$, we had to change the Learning Rate from 0.01 to 0.001.

Energy monitoring. In centralized DL, we often fine-tune FMs on servers with multiple GPUs and, thus, require very high bandwidth interconnects ($> 200\mathrm{GB/s}$) between the GPUs either via NVLink or Infiniband [2, 52]. FL only requires low bandwidth interconnects ($< 1\mathrm{GB/s}$) since communication happens sparingly compared to multi-GPU centralized learning [81]. This creates significant 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 [6] or deriving the energy consumption based on floating point operations and a client's system specifications [18]. 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 for a service provider to monitor. However, the bit-wide energy consumption model is available to calculate the cost of transmitting data [76]. The costs are directly tied to the number of parameters of a client update in an FL system [85]. As such, we can calculate the total energy consumption of communication as

$$P_{\mathrm{t}} = E_t \cdot \mathcal{B} = (n_{\mathrm{as}} \cdot E_{\mathrm{as}} + E_{\mathrm{bng}} + n_e \cdot E_e + n_c \cdot E_c + n_d \cdot E_d) \cdot \mathcal{B}. \tag{1}$$

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 [35]. Jalali et al. [45] 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 [9], it is currently impossible to track 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 [58, 68, 78].

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, $0.29\ \frac{€}{\mathrm{kWh}}$ [25]. The EU Commission produces quarterly reports on the electricity price trends [23]. Directly proportional to the power consumption, we emit $252\ \frac{gCO_2e}{\mathrm{kWh}}$ [24]. 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)) [45]. For the energy consumption per transmitted bit per network hop, we adopt the values from Jalali et al. [45], Vishwanath et al. [76] (Table 4).

Table 4: Energy consumption per bit network communication for our holistic energy monitoring approach. Values are adopted from Jalali et al. [45], Vishwanath et al. [76].

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 Additional Experimental Results for Section 6

We have conducted additional experiments with varying (ϵ, δ)-DP levels to outline the cost of perturbation-based privacy. Figure 4 shows that high privacy levels come at significant cost of up to $4\times$ compared to learning without DP, in our case.

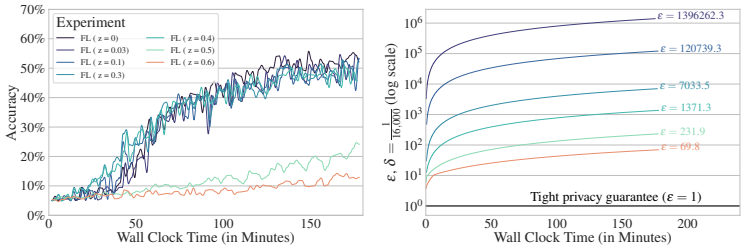

Figure 4: Privacy study. (ϵ, δ)-DP with varying noise multiplier (z) levels. ϵ is calculated based on $\delta = \frac{1}{16{,}000}$.

D Algorithmic Cost Analysis in Section 6

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. [1], SMPC as introduced by Bonawitz et al. [7], and HEC as introduced by Jin et al. [46].

D.1 (ϵ, δ)-Differential Privacy

The following algorithm (Algorithm 1) is taken verbatim from Andrew et al. [1]. 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, γ, η_c, η_s, η_C, z, σ_b, β)
  Initialize model θ^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, θ^0, η, 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 (Δ', b)
end function

D.2 Secure Multi-Party Computation

The SecAgg algorithmic costs (Table 5) are taken from Bonawitz et al. [7] Table 1. The naming convention has been adapted to our paper.

Table 5: SecAgg costs

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

D.3 Homomorphic Encryption

The following algorithm (Algorithm 2) is taken verbatim from Jin et al. [46]. 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

- $[\![\mathbf{W}]\!]$: 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**

  $\mathbf{W}_i \leftarrow Init(\mathbf{W})$;

  $\mathbf{S}_i \leftarrow Sensitivity(\mathbf{W}, \mathcal{D}_i)$;

  $[\![\mathbf{S}_i]\!] \leftarrow Enc(pk, \mathbf{S}_i)$;

  Send $[\![\mathbf{S}_i]\!]$ to server;

`// Server Encryption Mask Aggregation`

$[\![\mathbf{M}]\!] \leftarrow Select(\sum_{i=1}^{N} \alpha_i [\![\mathbf{S}_i]\!], p)$;

`// Training`

for $t = 1, 2, \ldots, T$ **do**

  for *each client* $i \in [N]$ **do in parallel**

    if $t = 1$ **then**

      Receive $[\![\mathbf{M}]\!]$ from server;

      $\mathbf{M} \leftarrow HE.Dec(sk, [\![\mathbf{M}]\!])$;

    if $t > 1$ **then**

      Receive $[\mathbf{W}_{\text{glob}}]$ from server;

      $\mathbf{W}_i \leftarrow HE.Dec(sk, \mathbf{M} \odot [\mathbf{W}_{\text{glob}}]) + (\mathbf{1} - \mathbf{M}) \odot [\mathbf{W}_{\text{glob}}]$;

    $\mathbf{W}_i \leftarrow Train(\mathbf{W}_i, \mathcal{D}_i)$;

    `// Additional Differential Privacy`

    if *Add DP* **then**

      $\mathbf{W}_i \leftarrow \mathbf{W}_i + Noise(b)$;

    $[\mathbf{W}_i] \leftarrow HE.Enc(pk, \mathbf{M} \odot \mathbf{W}_i) + (\mathbf{1} - \mathbf{M}) \odot \mathbf{W}_i$;

    Send $[\mathbf{W}_i]$ to server $\mathcal{S}$;

  `// Server Model Aggregation`

  $[\mathbf{W}_{\text{glob}}] \leftarrow \sum_{i=1}^{N} \alpha_i [\![\mathbf{M} \odot \mathbf{W}_i]\!] + \sum_{i=1}^{N} \alpha_i ((\mathbf{1} - \mathbf{M}) \odot \mathbf{W}_i)$;