2025 International Wireless Communications and Mobile Computing (IWCMC) | 979-8-3315-0887-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/IWCMC65282.2025.11059634

Selective Privacy-Preserving Federated Learning for Large Language Model Fine-Tuning

Qianqian Pan and Jun Wu*
Graduate School of Information, Production and Systems,
Waseda University, Fukuoka, Japan
* jun.wu@ieee.org

***Abstract*—The emergence of the large language model (LLM) accelerates network intelligence and supports numerous applications across various areas. Pre-trained LLM is trained based on massive amounts of public data, and domain-specific data is required to fine-tune LLM when it is adopted to specific fields. However, fine-tuning LLM with domain-specific data faces isolated data silos and data security issues. Although the federated learning framework mitigates the data silos issue and avoids direct exposure of local data, there are still the following challenges: i) The contradiction between the high computing/storage resources requirements of LLM fine-tuning and resource-constrained local devices, ii) privacy leakage through fine-tuning parameters. Efficient privacy-preserving LLM fine-tuning is becoming increasingly important and is still an open issue. To solve the above problems and challenges, we propose a selective privacy-preserving federated LLM fine-tuning mechanism. First, a selective privacy-preserving federated learning framework is designed to fine-tune LLM to specific domains efficiently while protecting sensitive information. Second, we propose a selective privacy-preserving exponential mechanism, which adds customized noise to private tokens of local data to protect sensitive information. Third, an adapter-enabled privacy-preserving LLM federated fine-tuning mechanism is proposed for high efficiency and security. Finally, experimental evaluations verify the effectiveness of our proposed mechanism.**

***Index Terms*—Large language model, fine-tuning, selective privacy preservation, federated learning**

I. Introduction

In the upcoming 6G era, artificial intelligence (AI) is expected to be deeply integrated with wireless networks. Large language model (LLM) has astonishing achievements in a broad spectrum of real-world applications across wide fields [1]–[3]. The rise of LLM accelerates network intelligence, providing numerous intelligent services in 6G, e.g. chatbots, semantic analysis, and writing assistance. LLM is trained based on vast amounts of public data from large and diverse sources, having up to trillions of parameters. When LLM is applied to specific areas, e.g. health care and education, it is required to fine-tune LLM with domain-specific data [4]–[6]. Compared with LLM pre-training using public data, LLM fine-tuning with domain-specific data confronts the following two issues: 1) Domain-specific data are often stored in isolated data silos. 2) Security and privacy concerns make data owners unwilling to share their data with the LLM training server. Therefore, privacy-preserving LLM fine-tuning remains a challenge.

Federated learning (FL) is a distributed learning framework to solve the data silos issue, where dataset owners train intelligent models locally and only transmit parameters to the central server [7], [8]. However, the traditional FL framework is not suitable for LLM fine-tuning for the following reasons: First, LLM fine-tuning requires a lot of computing resources and domain-specific data, which is inconsistent with resource-constrained local devices. Second, although the FL framework avoids direct data exposure, privacy information can be extracted by hacking the shared LLM fine-tuning parameters [9], [10]. Recently, the first challenge can be mitigated by fine-tuning LLM with a limited number of parameters. For example, the adapter tuning adopted in [11] inserts a small adapter module into the pre-trained LLM model and updates the adapter parameters merely. The authors of [12] utilize the Low-Rank Adaptation (LoRA) technique to decompose the LLM weight matrix into two low-domain matrices and train the low-domain matrix only. Protecting privacy during LLM fine-tuning is becoming increasingly important and is still an open issue.

Differential privacy (DP) technology is a feasible solution to protect privacy in the FL fine-tuning process, which introduces noise to local data. With characteristics of a solid mathematical foundation and high efficiency, DP has been widely applied in FL training. In [13], the authors add artificial noise to the model parameters by local clients before transmitting them to the central server for aggregation. The authors of [14] design a DP-based FL framework to protect sensitive information by adding customized demand-based noise. However, one challenge of traditional DP is that it reduces the availability and utility of local data due to the introduction of extra noise. Multiple variations have been proposed to mitigate this problem. The authors of [15] design the partial DP mechanism, which protects only sensitive attributes of datasets. In [16], selective DP is designed to protect sensitive parts of attributes in a dataset, where private and non-private attributes are distinguished in a data sample, and only the sensitive parts of the data sample are protected. To tackle the above issues and challenges, we propose a selective privacy-preserving federated LLM fine-tuning mechanism, taking both security and utility into account. The contribution of this work is summarized as follows:

- A selective privacy-preserving federated learning frame-

work for LLM fine-tuning is proposed, which improves LLM's ability in specific domains while protecting sensitive information.
- We design a selective privacy-preserving exponential mechanism, where private tokens in local data are perturbed to noisy tokens based on customized privacy budgets.
- An adapter-enabled privacy-preserving LLM federated fine-tuning mechanism is devised, where only inserted adapter parameters are updated based on privacy-preserving local datasets for high efficiency and security.

The remainder of this paper is organized as follows: In section II, related works are discussed. Section III proposes a privacy-preserving FL framework for LLM fine-tuning. The selective privacy-preserving exponential mechanism is designed in section IV. In section V, the adapter-enabked privacy-preserving LLM federated fine-tuning mechanism is designed. Experimental results and analysis are shown in section VI. This work is concluded in section VII.

II. Related Work

A. Federated Learning-Enabled LLM Fine-Tuning

The fine-tuning process of LLM requires massive amounts of data, which is unfeasible to share due to security issues and transmission limitations. Multiple studies have integrated the federated learning framework with LLM fine-tuning to mitigate these issues. In [17], the authors investigate the wireless federated learning mechanism for LLM fine-tuning, where personalized federated instruction tuning and task tuning are studied. The authors of [18] propose a collaborative LLM training framework based on FL, where distributed private data are used for training without transmission. In [19], the authors develop an FL-empowered open-source package for LLMs fine-tuning, called FederatedScope-LLM (FS-LLM), where extensive experiments are conducted to demonstrate the effectiveness of their work. Although these works prevent local data from being exposed to attackers directly, there are still privacy leakage threats through analyzing parameters sent by local devices.

The authors of [20] try to address the privacy issue, which is achieved mainly based on key encryption during the client and server communication. However, due to the complexity of encryption and frequent client-server interaction, the efficiency of this work is reduced.

B. DP for Intelligent Model Training

DP is a privacy-preserving technique with low computing complexity and a solid mathematical foundation, which is widely applied in federated model training. The authors of [21] investigate DP applications in AI models, which is used for privacy preservation, security improvement, stabilizing learning, and fair models. In [22], the authors develop an open differentially private deep learning framework. This open framework is realized based on the private stochastic gradient descent algorithm and designed for medical image classification and segmentation. Stacey *et al.* construct a novel federated learning system with a privacy guarantee, which is developed based on local differential privacy [23]. Although DP plays a great role in privacy protection during model training, it degrades the accuracy of AI models as noise is added to training data or model parameters.

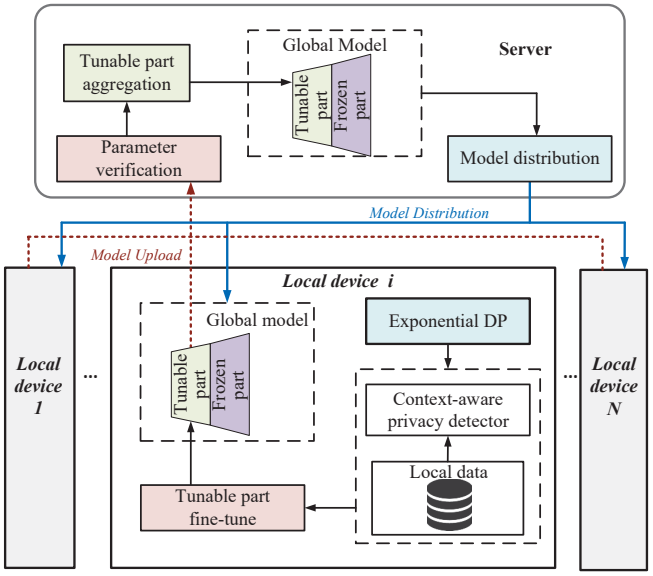

Fig. 1. Framework of selective privacy-preserving federated LLM fine-tuning.

Some variations of traditional DP are studied, among which partial DP and selective DP are famous DP variations. These DP variations have the ability to add artificial noise only to the sensitive parts of data samples. Hence, selective DP is able to protect private information while guaranteeing data availability. Shi *et al.* design the selective DP scheme and apply it to the language models and large language models [16], [24]. These existing SDP works only focus on the centralized learning system without considering the distributed learning system, which is more vulnerable to privacy attacks. Besides, unlike redacting private information with preset tokens, our proposed selective privacy-preserving mechanism perturbs sensitive information on demand based on the exponential DP scheme.

III. Selective Privacy-Preserving Federated LLM Fine-Tuning Framework

We propose a selective privacy-preserving federated LLM fine-tuning framework and describe its main components.

A. Proposed Framework

The proposed framework of selective privacy-preserving federated LLM fine-tuning is shown in Fig. 1, which consists of a central server and multiple local devices. Details are described as follows:

- Central server: The central server has high computing power and sufficient storage resources. Thus, the central server has ability to pre-train an LLM model, which is usually achieved via self-supervised learning on unlabeled public data. In the designed federated fine-tuning

framework, the central server is also responsible for identity verification, model distribution, and aggregation to fine tune LLM models for specific domains by collaborating with multiple local devices.

- Local devices: Local devices have local datasets. Compared with the public data used for LLM pre-training, local data on devices is of higher quality and is usually labeled. However, local data inevitably contains private information, which makes it unshareable due to privacy concerns. In our framework, multiple local devices assist LLM fine-tuning based on their local data collaboratively. As computing and memory resources at local devices are limited, only a part of LLM parameters are updated. Besides, local devices also adopt the selective privacy-preserving scheme to protect their sensitive information from attackers during LLM fine-tuning.

After LLM pre-training, the central server adjusts LLM with the assistance of multiple local devices. Our proposed framework aims to utilize local data on devices to fine-tune LLM for specific-domain applications while protecting sensitive information of the data.

B. Main Components

To avoid data silo and privacy leakage issues, the central server and multiple devices collaborate to fine-tune LLM. Our proposed selective privacy-preserving federated LLM fine-tuning framework mainly includes the following two components:

- Federated LLM fine-tuning: After the pre-training stage of LLM, the central server distributes the pre-trained LLM to verified local devices to realize fine-tuning of the pre-trained LLM. Local devices receive the global LLM and store it in their memory. Then, the global LLM parameters are divided into two parts: One is the frozen part, whose parameters are not updated by local devices. Another is the tunable part, and local devices only fine-tune the parameters of this part. Then, local devices fine-tune the LLM parameters of the tunable part based on their local dataset. Meanwhile, the privacy-preserving technique is adopted to avoid sensitive information leakage during LLM fine-tuning. After fine-tuning, local devices transmit the parameters of the tunable part to the central server. Next, the central server verifies the received parameters and aggregates them to update the global LLM. These steps are operated iteratively until the convergence of LLM in specific domains.
- Customized selective privacy preservation: Based on privacy budgets, local devices add artificial noise to the sensitive parts of their own training data. In this component, a context-aware privacy detector is utilized to recognize the sensitive parts of a data sample in the local dataset. Next, exponential DP noise is only added to these sensitive parts of the data sample according to the privacy demands of local devices and the sensitivity level of private information. The purpose of this component is to preserve the privacy of local devices while guaranteeing data utility.

Algorithm 1 selective DP-based exponential mechanism

Input: $\mathcal{D}_i, \epsilon, d$
Output: $\mathcal{D}_i^{\mathrm{s}}$
Initialization: $\mathcal{D}_i^{\mathrm{s}} = \emptyset$
for $\boldsymbol{s}_{i,j} \in \mathcal{D}_i$ **do**
  $\{\boldsymbol{s}_{i,j}^{\mathrm{p}}, \boldsymbol{s}_{i,j}^{\mathrm{n}}\} \leftarrow PrivacyDector(\boldsymbol{s}_{i,j})$
  for $\boldsymbol{s}_{i,j,k} \in \boldsymbol{s}_{i,j}^{\mathrm{p}}$ **do**
    Token dataset: $\mathcal{D}_{i,j,k}^{\mathrm{p}} \leftarrow \{\boldsymbol{a} | \parallel \boldsymbol{a} - \boldsymbol{s}_{i,j,k}^{\mathrm{p}} \parallel \leq d\}$
    Score function $g(\mathcal{D}_{i,j,k}^{\mathrm{p}}, \boldsymbol{a})$ as (3)
    Calculate sensitivity $\Delta g(\mathcal{D}_{i,j,k}^{\mathrm{p}}, \boldsymbol{a})$ based on (4)
    Calculate probability $Pr(\mathcal{D}_{i,j,k}^{\mathrm{p}}, \boldsymbol{a})$ based on (5)
    Select noisy token $\boldsymbol{a}_{i,j,k}$ for $\boldsymbol{s}_{i,j,k}$ according to the probability distribution in (5)
    Replace $\boldsymbol{s}_{i,j,k}$ with $\boldsymbol{a}_{i,j,k}$: $\boldsymbol{s}_{i,j,k} \leftarrow \boldsymbol{a}_{i,j,k}$
  end for
  Add $\boldsymbol{s}_{i,j}$ to $\mathcal{D}_i^{\mathrm{s}}$: $\mathcal{D}_i^{\mathrm{s}} \leftarrow \mathcal{D}_i^{\mathrm{s}} + \boldsymbol{s}_{i,j}$
end for

IV. Selective DP-based Exponential Mechanism for LLM Fine-tuning Privacy Preservation

LLM is fine-tuned based on datasets of local devices. To protect the privacy of devices, a selective DP-based exponential mechanism is proposed. Inspired by the selective DP in [25], we first detect the sensitive parts of each sample in the local dataset. Then, based on our designed exponential DP mechanism, we mask the sensitive parts of all samples belonging to the local dataset.

In the proposed system, we denote the central server as S and consider that there are N local devices participating in the privacy-preserving LLM federated fine-tuning expressed as the set $\mathcal{N} = \{1, 2, \cdots, N\}$. We denote the local dataset of the device $i \in \mathcal{N}$ as $\mathcal{D}_i$. First, a context-aware privacy detector is utilized for each sample $\boldsymbol{s}_{i,j}$ of the local dataset $\mathcal{D}_\rangle$ to distinguish its sensitive part $\boldsymbol{s}_{i,j}^{\mathrm{p}}$ and the normal part $\boldsymbol{s}_{i,j}^{\mathrm{n}}$, that is,

$$\boldsymbol{s}_{i,j} = \{\boldsymbol{s}_{i,j}^{\mathrm{p}}, \boldsymbol{s}_{i,j}^{\mathrm{n}}, \quad \forall \boldsymbol{s}_{i,j} \in \mathcal{D}_i\}. \tag{1}$$

For the k-th private token of the sensitive part $\boldsymbol{s}_{i,j}^{\mathrm{p}}$ in the sample $\boldsymbol{s}_{i,j} \in \mathcal{D}_i$, we denote its original token as $\boldsymbol{s}_{i,j,k}^{\mathrm{p}}$. We define the token dataset $\mathcal{D}_{i,j,k}^{\mathrm{p}}$ for the original token $\boldsymbol{s}_{i,j,k}^{\mathrm{p}}$ and it satisfies the following condition:

$$\mathcal{D}_{i,j,k}^{\mathrm{p}} = \{\boldsymbol{a} | \parallel \boldsymbol{a} - \boldsymbol{s}_{i,j,k}^{\mathrm{p}} \parallel \leq d\}, \tag{2}$$

where $\parallel \boldsymbol{a} - bms_{i,j,k}^{\mathrm{p}} \parallel$ is the distance between the token $\boldsymbol{a}$ and the original token $\boldsymbol{s}_{i,j,k}^{\mathrm{p}}$. The score function for the original token $\boldsymbol{s}_{i,j,k}^{\mathrm{p}}$ is formulated as

$$g(\mathcal{D}_{i,j,k}^{\mathrm{p}}, \boldsymbol{a}) = -\alpha \parallel \boldsymbol{a} - \boldsymbol{s}_{i,j,k}^{\mathrm{p}} \parallel, \quad \boldsymbol{a} \in \mathcal{D}_{i,j,k}^{\mathrm{p}} \tag{3}$$

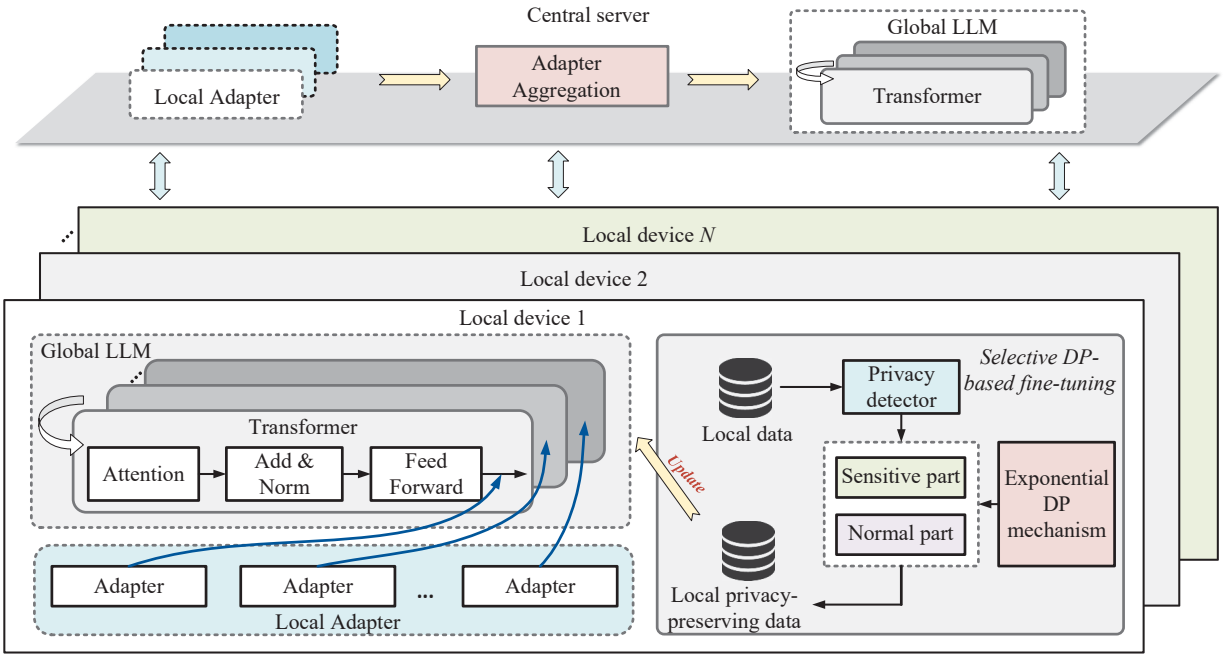

Fig. 2. Adapter-based privacy-preserving LLM federated fine-tuning mechanism.

where α is the positive factor. According to the definition of sensitivity [25], we obtain the sensitivity as

$$\begin{aligned}\Delta g(\mathcal{D}^{\mathrm{p}}_{i,j,k}, \boldsymbol{a}) &= \max_{\mathcal{D}^{\mathrm{p}}_{i,j,k}, \mathcal{D}^{'\mathrm{p}}_{i,j,k}} \parallel g(\mathcal{D}^{\mathrm{p}}_{i,j,k}, \boldsymbol{a}) - g(\mathcal{D}^{'\mathrm{p}}_{i,j,k}, \boldsymbol{a}) \parallel, \\ &= 2d\end{aligned} \tag{4}$$

where $\mathcal{D}^{'\mathrm{p}}_{i,j,k}$ is the adjacent set of $\mathcal{D}^{\mathrm{p}}_{i,j,k}$, which has only one different sample of $\mathcal{D}^{\mathrm{p}}_{i,j,k}$. Therefore, given the privacy budget ϵ, the original token $\boldsymbol{s}^{\mathrm{p}}_{i,j,k}$ selects the token $\boldsymbol{a}$ as the perturbed token by the following probability:

$$Pr(\mathcal{D}^{\mathrm{p}}_{i,j,k}, \boldsymbol{a}) = \frac{e^{-\epsilon\alpha \| \boldsymbol{a} - \boldsymbol{s}^{\mathrm{p}}_{i,j,k} \| / (4d)}}{\sum_{\boldsymbol{a}' \in \mathcal{D}^{\mathrm{p}}_{i,j,k}} e^{-\epsilon\alpha \| \boldsymbol{a}' - \boldsymbol{s}^{\mathrm{p}}_{i,j,k} \| / (4d)}} \tag{5}$$

In our designed selective DP-based exponential mechanism, every private token $\boldsymbol{s}^{\mathrm{p}}_{i,j,k}$ selects its noisy token $\boldsymbol{a}_{i,j,k}$ as the alternative in LLM fine-tuning based on (5). Applying the designed selective DP-based exponential mechanism to every sample $\boldsymbol{s}_{i,j} \in \mathcal{D}_i$ in the local data for the local device $i \in \mathcal{N}$, we obtain the selective privacy-preserving data set denoted as $\mathcal{D}^{\mathrm{s}}_i$. The details of the proposed selective DP-based exponential mechanism are presented in Algorithm 1.

V. Adapter-based Privacy-Preserving LLM Federated Fine-Tuning

The proposed adapter-based privacy-preserving LLM federated fine-tuning mechanism is shown in Fig. 2.

A. Adapter-based Federated LLM Fine-Tuning

The pre-trained LLM is denoted as $\boldsymbol{w}_{\mathrm{G}}$, which is distributed to local devices before the fine-tuning phase. To realize effective LLM fin-tuning, we only update a small part of the parameters rather than the whole LLM based on the adapter technique [26]. The adapter is a promising method to achieve parameter-efficient fine-tuning of modern language models and integrated with federated learning systems [27].

The adapter aims to freeze the entire pre-trained LLM and insert some small modules at different locations of the LLM. As shown in Fig. 2, adapter modules are inserted into the position after the *Fed & Forward* module of each transformer. For the input $\boldsymbol{x}_i \in \mathbb{R}^{1\times r}$, the inserted adapter layer can be expressed with a pair of weight matrix $\boldsymbol{w}_{\mathrm{d}} \in \mathcal{R}^{r\times h}$ and $\boldsymbol{w}_{\mathrm{d}} \in \mathcal{R}^{h\times r}$. The weight matrix $\boldsymbol{w}_{\mathrm{d}} \in \mathcal{R}^{r\times h}$ projects the input $\boldsymbol{x}_i \in \mathbb{R}^{1\times r}$ down to the h dimension, where $h < r$. Then, an activate function is adopted to the $\boldsymbol{x}_i\boldsymbol{w}_{\mathrm{d}}$ and we have $f(\boldsymbol{x}_i\boldsymbol{w}_{\mathrm{d}})$. After that, the weight matrix $\boldsymbol{w}_{\mathrm{d}} \in \mathcal{R}^{h\times r}$ is used to project $f(\boldsymbol{x}_i\boldsymbol{w}_{\mathrm{d}})$ up to the r dimension. Adapters are connected to the original LLM via residual links. Thus, the output of the adapter can be formulated as:

$$\boldsymbol{x}'_i = \boldsymbol{x}_i + f(\boldsymbol{x}_i\boldsymbol{w}_{\mathrm{d}})\boldsymbol{w}_{\mathrm{u}} \tag{6}$$

During the process of LLM fine-tuning, only the parameters of the adapters are adjusted. After the privacy-preserving training at local devices based on their dataset, fine-tuned adapter parameters are uploaded to the central server. The central server aggregates uploaded local adapters and distributes the aggregated global adapter to the participating local devices. Then, local devices update their local adapter parameters with the distributed aggregated global adapter.

B. Overall Workflow

The proposed adapter-based privacy-preserving LLM federated fine-tuning mechanism consists of the following four stages:

1) Initialization: The central server S publishes the information about LLM to the system. Local devices willing to participate in the selective privacy-preserving federated LLM fine-tuning transmit their identities and requests to the central server.
2) LLM distribution: The central server first verifies the identities of local devices willing to participate in the federated fine-tuning mechanism. Then, the central server distributes its pre-trained LLM $\boldsymbol{w}_{\mathrm{G}}$ to those verified local devices.
3) Local fine-tuning: Local devices insert adapters to the pre-trained LLM according to section V-A. Then, local devices perturb the sensitive parts of each sample in their local data based on the designed selective DP-based exponential mechanism in section IV. Next, the parameters of adapters are adjusted based on the selective privacy-preserving dataset $\mathcal{D}_i^{\mathrm{s}}, i \in \mathcal{N}$ at local devices. After that, updated local adapter parameters are uploaded to the central server.
4) Global aggregation: After receiving local parameters of adapters, the central server aggregates them as follows:

$$\begin{aligned} \boldsymbol{w}_{\mathrm{d}}^{(m)} &= \frac{1}{N} \sum_{i \in \mathcal{N}} \boldsymbol{w}_{i,\mathrm{d}}^{(m)}, \\ \boldsymbol{w}_{\mathrm{u}}^{(m)} &= \frac{1}{N} \sum_{i \in \mathcal{N}} \boldsymbol{w}_{i,\mathrm{u}}^{(m)}, \end{aligned} \tag{7}$$

where $m \in \mathcal{M}$ and $\mathcal{M}$ is the set of adapter.

The above LLM distribution, local fine-tuning, and global aggregation stages are operated iteratively until the model converges in the specific application domain or satisfies the preset end conditions.

VI. Experimental Evaluation

A. Experimental Setup

Experimental evaluations are conducted based on the GPT-2 LLM and the wikitext-2 dataset. We investigate the accuracy of the fine-tuned LLM, which is indicated by the metric perplexity (PPL). PPL is an important indicator to evaluate the accuracy of language models and lower PPL means higher accuracy performance. We compare our proposed selective privacy-preserving federated LLM fine-tuning scheme with the following two baselines:

- Traditional FL: This scheme fine-tunes LLM based on the traditional FL framework without any privacy-preserving measurements.
- FL with DP: This scheme is also based on the federated fine-tuning framework. Moreover, DP-based artificial noise is added to both sensitive and normal parts.

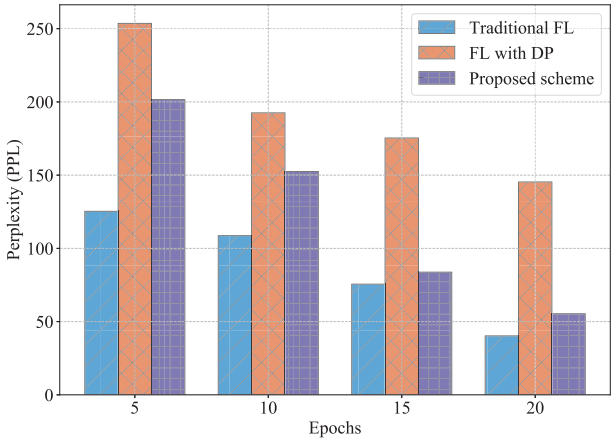

Fig. 3. PPL performance of the proposed method and baselines.

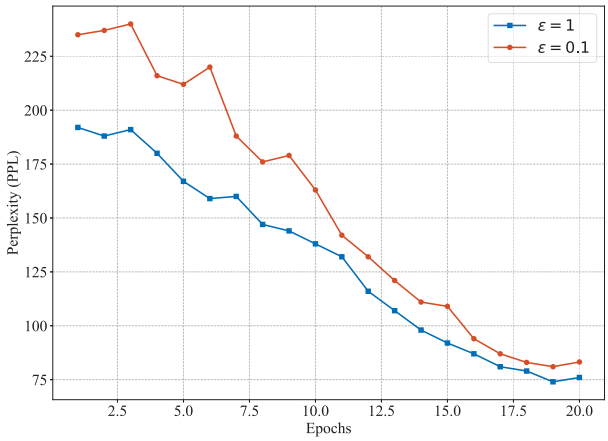

Fig. 4. PPL performance between various privacy budgets.

B. Experimental Results

The PPL performance of our proposed method and baselines is shown in Fig. 3. Along with the epoch, the metric PPL of all the methods decreases gradually. This phenomenon shows that the accuracy of LLM gradually improves and converges with the increase of fine-tuning epochs. The traditional FL scheme without privacy-preserving measurements had the lowest PPL, which means the highest accuracy performance among these three schemes. In contrast, the FL with DP scheme behaves the worst accuracy performance with the highest PPL. Our proposed scheme achieves the medium accuracy performance. These experimental results are reasonable. Since the traditional FL scheme does not add any privacy-preserving noise, it achieves the best accuracy performance. However, privacy leakage threats exist in the traditional FL scheme. As the FL with DP scheme adds noise to both sensitive and normal parts, its accuracy performance reduces. Our proposed scheme takes both utility and privacy into consideration, aiming to find a trade-off between accuracy performance and privacy

preservation.

The PPL performance of our proposed method under various privacy budgets is presented in Fig. 4. We set the privacy budget as $\epsilon = \{0.1, 1\}$. From the results in Fig. 4, we can obtain that the proposed scheme with the lower privacy budget exhibits higher PPL, indicating lower accuracy performance. This is because the lower privacy budget means higher privacy demands and better preservation. Thus, the accuracy performance of LLM decreases.

VII. Conclusion

This paper proposes a selective privacy-preserving federated LLM fine-tuning mechanism. First, we propose a selective privacy-preserving federated learning framework for LLM fine-tuning, which includes the federated LLM fine-tuning component and the customized selective privacy preservation component. Then, a selective privacy-preserving exponential mechanism is designed, where private information of local data is perturbed with noise on demand. After that, we design an adapter-based privacy-preserving LLM federated fine-tuning mechanism to guarantee high efficiency and security. Finally, experimental evaluation demonstrates the effectiveness of our proposed selective privacy-preserving federated LLM fine-tuning mechanism. Future works will focus on the privacy preservation of LLM in the collaboration among decentralized local devices.

Acknowledgment

This work was supported in part by the JSPS KAKENHI under Grants 24KF0259 and 23K11072, in part by the Telecommunications Advancement Foundation. Qianqian Pan is currently the JSPS International Research Fellow.

References

[1] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier *et al.*, "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education," *Learning and individual differences*, vol. 103, p. 102274, 2023.

[2] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong *et al.*, "A Survey of Large Language Models," *arXiv preprint arXiv:2303.18223*, 2023.

[3] Y. Chang, X. Wang, J. Wang, Y. Wu, L. Yang, K. Zhu, H. Chen, X. Yi, C. Wang, Y. Wang *et al.*, "A Survey on Evaluation of Large Language Models," *ACM Transactions on Intelligent Systems and Technology*, vol. 15, no. 3, pp. 1–45, 2024.

[4] N. Ding, Y. Qin, G. Yang, F. Wei, Z. Yang, Y. Su, S. Hu, Y. Chen, C.-M. Chan, W. Chen *et al.*, "Parameter-Efficient Fine-Tuning of Large-Scale Pre-Trained Language Models," *Nature Machine Intelligence*, vol. 5, no. 3, pp. 220–235, 2023.

[5] T. Susnjak, P. Hwang, N. H. Reyes, A. L. Barczak, T. R. McIntosh, and S. Ranathunga, "Automating Research Synthesis with Domain-Specific Large Language Model Fine-Tuning," *ACM Transactions on Knowledge Discovery from Data*, 2024.

[6] X. Yang, A. Chen, N. PourNejatian, H. C. Shin, K. E. Smith, C. Parisien, C. Compas, C. Martin, A. B. Costa, M. G. Flores *et al.*, "A Large Language Model for Electronic Health Records," *NPJ digital medicine*, vol. 5, no. 1, p. 194, 2022.

[7] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated Learning: Challenges, Methods, and Future Directions," *IEEE signal processing magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[8] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings *et al.*, "Advances and Open Problems in Federated Learning," *Foundations and trends® in machine learning*, vol. 14, no. 1–2, pp. 1–210, 2021.

[9] L. Lyu, H. Yu, X. Ma, C. Chen, L. Sun, J. Zhao, Q. Yang, and P. S. Yu, "Privacy and Robustness in Federated Learning: Attacks and Defenses," *IEEE transactions on neural networks and learning systems*, vol. 35, no. 7, pp. 8726–8746, 2022.

[10] Z. Wang, M. Song, Z. Zhang, Y. Song, Q. Wang, and H. Qi, "Beyond Inferring Class Representatives: User-Level Privacy Leakage from Federated Learning," in *IEEE INFOCOM 2019-IEEE conference on computer communications*. IEEE, 2019, pp. 2512–2520.

[11] R. He, L. Liu, H. Ye, Q. Tan, B. Ding, L. Cheng, J.-W. Low, L. Bing, and L. Si, "On the Effectiveness of Adapter-based Tuning for Pretrained Language Model Adaptation," *arXiv preprint arXiv:2106.03164*, 2021.

[12] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-Rank Adaptation of Large Language Models," *arXiv preprint arXiv:2106.09685*, 2021.

[13] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. Quek, and H. V. Poor, "Federated Learning with Differential Privacy: Algorithms and Performance Analysis," *IEEE transactions on information forensics and security*, vol. 15, pp. 3454–3469, 2020.

[14] Q. Pan, J. Wu, A. K. Bashir, J. Li, W. Yang, and Y. D. Al-Otaibi, "Joint Protection of Energy Security and Information Privacy for Energy Harvesting: An Incentive Federated Learning Approach," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 5, pp. 3473–3483, 2021.

[15] Y. Zhu and Y.-X. Wang, "Poission Subsampled Rényi Differential Privacy," in *International Conference on Machine Learning*. PMLR, 2019, pp. 7634–7642.

[16] W. Shi, A. Cui, E. Li, R. Jia, and Z. Yu, "Selective Differential Privacy for Language Modeling," *arXiv preprint arXiv:2108.12944*, 2021.

[17] F. Jiang, L. Dong, S. Tu, Y. Peng, K. Wang, K. Yang, C. Pan, and D. Niyato, "Personalized Wireless Federated Learning for Large Language Models," *arXiv preprint arXiv:2404.13238*, 2024.

[18] R. Ye, W. Wang, J. Chai, D. Li, Z. Li, Y. Xu, Y. Du, Y. Wang, and S. Chen, "Openfedllm: Training Large Language Models on Decentralized Private Data via Federated Learning," in *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*, 2024, pp. 6137–6147.

[19] W. Kuang, B. Qian, Z. Li, D. Chen, D. Gao, X. Pan, Y. Xie, Y. Li, B. Ding, and J. Zhou, "FederatedScope-LLM: A Comprehensive Package for Fine-Tuning Large Language Models in Federated Learning," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 5260–5271.

[20] J. Zheng, H. Zhang, L. Wang, W. Qiu, H. Zheng, and Z. Zheng, "Safely Learning with Private Data: A Federated Learning Framework for Large Language Model," *arXiv preprint arXiv:2406.14898*, 2024.

[21] T. Zhu, D. Ye, W. Wang, W. Zhou, and S. Y. Philip, "More Than Privacy: Applying Differential Privacy in Key Areas of Artificial Intelligence," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 6, pp. 2824–2843, 2020.

[22] A. Ziller, D. Usynin, R. Braren, M. Makowski, D. Rueckert, and G. Kaissis, "Medical Imaging Deep Learning with Differential Privacy," *Scientific Reports*, vol. 11, no. 1, p. 13524, 2021.

[23] S. Truex, L. Liu, K.-H. Chow, M. E. Gursoy, and W. Wei, "LDP-Fed: Federated Learning with Local Differential Privacy," in *Proceedings of the third ACM international workshop on edge systems, analytics and networking*, 2020, pp. 61–66.

[24] W. Shi, R. Shea, S. Chen, C. Zhang, R. Jia, and Z. Yu, "Just Fine-Tune Twice: Selective Differential Privacy for Large Language Models," *arXiv preprint arXiv:2204.07667*, 2022.

[25] Q. Pan, J. Wu, X. Zheng, W. Yang, and J. Li, "Differential Privacy and IRS Empowered Intelligent Energy Harvesting for 6G Internet of Things," *IEEE Internet of Things Journal*, vol. 9, no. 22, pp. 22 109–22 122, 2021.

[26] Y.-L. Sung, J. Cho, and M. Bansal, "VL-Adapter: Parameter-Efficient Transfer Learning for Vision-and-Language Tasks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5227–5237.

[27] D. Cai, Y. Wu, S. Wang, F. X. Lin, and M. Xu, "FedAdapter: Efficient Federated Learning for Modern NLP," *arXiv preprint arXiv:2205.10162*, 2022.