

REVIEW

Current Research and Open Gaps in Federated Large Language Models: A Systematic Review

Roberto Gomes Ribeiro   [Federal University of Ouro Preto | roberto.gomes@ufop.edu.br]

Frederico Gadelha Guimarães  [Federal University of Minas Gerais | fredericoguimaraes@dcc.ufmg.br]

Hélder Seixas Lima  [Federal Institute of Education, Science and Technology of Northern Minas Gerais | helder.seixas@ifnmg.edu.br]

Vinicius Antonio de Oliveira Martins  [Federal University of Ouro Preto | vinicius.martins@ufop.edu.br]

Allana Tavares Bastos  [Federal University of Minas Gerais | allanatavares@dcc.ufmg.br]

Larissa Duarte Santana  [Federal University of Minas Gerais | larissa.santana@dcc.ufmg.br]

Beatriz Camilly Gulart Pereira  [Federal University of Minas Gerais | beatrizcamilly@dcc.ufmg.br]

Carla Beatriz Ferreira  [Federal University of Minas Gerais | carlaferreira@dcc.ufmg.br]

Júlia Borba Fonseca de Souza  [Federal University of Minas Gerais | juliaborba@dcc.ufmg.br]

Melissa Costa Corrêa de Lima  [Federal University of Minas Gerais | melissaccl@ufmg.br]

 Departamento de Computação e Sistemas (DECSI), Instituto de Ciências Exatas e Aplicadas (ICEA), Universidade Federal de Ouro Preto (UFOP), João Monlevade-MG, MG, CEP: 35931-088, Brazil.

Received: 30 September 2025 • Accepted: 14 August 2026 • Published: 30 September 2026

Abstract The integration of Large Language Models (LLMs) with Federated Learning (FL), known as FedLLMs, enables decentralized model training while preserving data privacy. However, this integration introduces novel challenges while also inheriting existing issues from FL and LLMs. This systematic review addresses the research question: *How are the key challenges in Federated Large Language Models being addressed in current research, and what gaps remain to be explored?* We analyzed 168 studies across four major databases, applying eligibility criteria that resulted in 24 studies for detailed synthesis. Findings were organized across twelve key challenges: multimodal foundation model integration; data erasure, legal, ethical, and property rights; environmental sustainability and energy efficiency; federated learning over black-box LLMs; personalization and domain adaptation; continual learning and continuous data streams; communication overhead and efficiency constraints; security and robustness; heterogeneous data, devices, and resources; computational efficiency and resource constraints; and data privacy preservation. While some challenges have received attention, solutions are often limited to narrow domains or experimental datasets, and several remain unaddressed. These insights highlight critical gaps and provide a foundation for future research toward generalizable, responsible, and sustainable LLM technologies.

Keywords: FedLLMs, LLM technologies, Systematic Review, Federated Learning.

1 Introduction

Recent advances in Large Language Models (LLMs) have driven significant progress in Natural Language Processing (NLP), powering applications ranging from intelligent assistants to machine translation. As researchers push the boundaries of language understanding, the future of LLMs lies in making them more grounded, efficient, and better aligned with human intent—enabling responsible deployment across diverse real-world applications [Bommasani *et al.*, 2021]. However, training and personalizing these models require vast amounts of data, raising critical concerns regarding privacy, security, and ethical use.

In this scenario, Federated Learning has emerged as a decentralized machine learning paradigm that allows multiple clients to jointly train a shared model while keeping their data localized. Despite its promise, Federated Learning faces significant challenges, including non-IID data distributions, limited local datasets, high communication costs, privacy vulnerabilities, and the complexity of delivering personalized and scalable models.

Considering these factors, the integration of Federated

Learning and LLMs presents novel opportunities alongside distinct and interconnected challenges. Training in a federated environment enables the personalization of language models in a distributed manner while preserving user data confidentiality. Moreover, pretrained LLMs can help overcome FL limitations by providing robust initialization, facilitating efficient transfer learning, and supporting effective personalization across heterogeneous clients.

The integration of Federated Learning and Large Language Models (LLMs), commonly referred to as FedLLM, has been receiving increasing attention from the research community due to its potential to enable decentralized training while preserving data privacy and complying with regulatory requirements.

Reflecting this growing interest, recent review studies have surveyed the emerging field of FedLLMs, highlighting both its potential and the challenges it entails. Chen *et al.* [2024] offer an overview of the current state of research on FedLLMs, discussing typical applications and the benefits of such integration. Similarly, Cheng *et al.* [2024] present a comprehensive review by systematically examining existing research across the entire lifecycle of federated

LLMs, including pre-training, fine-tuning, and practical applications. Hu *et al.* [2024] provide a systematic introduction to FedLLMs as a distributed learning approach under the FL paradigm, introducing the concept which enables collaborative fine-tuning by updating only a small subset of parameters in local models. Complementing these perspectives, Yao *et al.* [2024] survey recent advances with a particular focus on fine-tuning and prompt learning in federated settings.

Despite their potential, FedLLMs encounter a variety of novel challenges that have been highlighted in the literature, with the four aforementioned studies addressing them from distinct viewpoints (see a summary description in subsection 1.1). While these reviews provide important overviews of the FedLLM landscape, most focus primarily on architectural aspects, fine-tuning strategies, system perspectives, or lifecycle stages of federated LLMs. In contrast, this work adopts a challenge-oriented perspective, synthesizing how the current literature addresses key FedLLM challenges and identifying underexplored and unaddressed research gaps across the selected studies.

Motivated by these open challenges, this paper examines how they have been addressed in the existing literature. Our goal is to synthesize the approaches, identify existing strategies, and highlight unresolved gaps that warrant further investigation. Accordingly, we present a systematic review guided by the following central research question: How are the key challenges in Federated Large Language Models being addressed in current research, and what gaps remain to be explored?

Following the taxonomy proposed by Braun *et al.* [2021], this research question goes beyond purely descriptive findings by incorporating explanatory aspects, as it seeks to understand how existing studies have addressed FedLLM challenges through recurring strategies, methodological trends, and associated trade-offs. Moreover, it integrates a gap-identification component, aiming to highlight unresolved issues that require further investigation.

The remainder of this paper is organized as follows. Subsection 1.1 further discusses the key challenges identified in the state of the art. Section 2 provides background on Federated Learning, Large Language Models, and FedLLMs. Section 3 describes the systematic search methodology, based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) protocol. Section 4 presents the main findings and discusses the insights derived from the reviewed literature. Finally, Section 5 concludes the paper by summarizing key contributions and outlining future research directions.

1.1 Identified challenges in the literature

[Yao *et al.*, 2024] emphasize that deploying LLMs in FL environments entails high communication costs. Similarly, [Chen *et al.*, 2024] note that integrating FedLLMs introduces substantial communication overhead due to the exchange of large model parameters, which may cause communication time to exceed training time. They also point out that this condition can negatively impact the environment by increasing carbon emissions, highlighting environmental sustainability and energy efficiency as emerging challenges.

Another challenge involves model property rights, as

training LLMs relies on vast, domain-specific datasets that make the resulting models commercially valuable. Recent FedLLM review studies [Chen *et al.*, 2024; Hu *et al.*, 2024] highlight that this increases the risk of model leaks, intellectual property infringement, and privacy violations in federated settings. Cheng *et al.* [2024] further emphasize that addressing these challenges requires a multidisciplinary approach that integrates advances in machine learning and statistics with insights from law, ethics, and the social sciences. Moreover, Yao *et al.* [2024] note that integrating LLMs into Federated Learning raises legal and ethical concerns over privacy, intellectual property, and compliance, thereby demanding responsible deployment mechanisms.

Closely related to these issues, data erasure in FedLLMs remains an open and underexplored challenge. As noted by Cheng *et al.* [2024], the reliance on large-scale datasets may introduce significant risks, including privacy breaches, accuracy limitations, copyright infringement, and the reinforcement of societal biases. Moreover, recent legal disputes have intensified the debate regarding the applicability of fair use principles to LLM training, further emphasizing the urgent need for research into effective data erasure mechanisms within federated learning settings.

Heterogeneity in data, devices, and resources emerges as a key challenge in the FedLLM context. As noted by Cheng *et al.* [2024], differences in client data and system resources can cause model divergence, uneven training progress, and suboptimal performance. Yao *et al.* [2024] underscore that such heterogeneity constitutes a central research barrier by undermining global model convergence. Similarly, Hu *et al.* [2024] emphasize that variations in local dataset distributions and client hardware/software capabilities can compromise model performance.

Despite the use of federated learning, preserving data privacy remains a significant challenge for LLMs. According to Cheng *et al.* [2024], existing privacy methods face limitations in handling complex tasks, heterogeneous client resources, and high computation and communication costs, highlighting the need for robust, lifecycle-spanning privacy solutions. Chen *et al.* [2024] further note that LLMs gather vast amounts of internet data during pretraining, including private and copyrighted information, which is difficult to trace. Their strong memory capabilities can cause generated outputs to closely resemble sensitive training data, posing legal and privacy risks for federated models.

Another challenge arises when combining FL with black-box LLMs. As noted by Chen *et al.* [2024], high-performance models operate as APIs, providing only input-output access, which limits traditional federated learning methods and complicates their integration. Similarly, Yao *et al.* [2024] emphasize that many state-of-the-art LLMs function as black-box APIs, restricting access to model parameters and posing challenges for conventional federated learning techniques.

Federated LLMs face ongoing challenges in security threats and robustness issues. Hu *et al.* [2024] highlight vulnerabilities to backdoor and inference attacks, while heterogeneity in data, devices, and resources further complicates model reliability, emphasizing the need for resilient and privacy-preserving solutions. Chen *et al.* [2024] note that,

despite the privacy benefits of federated learning, FedLLMs remain exposed to additional threats, including poisoning and inference attacks. Moreover, Cheng *et al.* [2024] emphasize that the large-scale and complex nature of federated LLM systems increases susceptibility to malicious behaviors, underscoring the necessity for comprehensive assessment and robust defense mechanisms.

Personalization and domain adaptation remain open challenges for Federated LLMs. Chen *et al.* [2024] note that, although federated learning enables models to leverage local data, the heterogeneity of client datasets and system resources can hinder effective personalization. Furthermore, the application of LLMs in domain-specific contexts highlights the need for models that can adapt to varying data distributions and domain requirements. Cheng *et al.* [2024] further emphasize that domain-specific LLM agents require integration with highly sensitive and distributed data. While federated learning allows models to be trained and fine-tuned without centralizing data, significant challenges remain in ensuring that these models are effectively personalized and adapted to diverse domain contexts.

Computational efficiency and resource constraints emerge as additional challenges in FedLLMs. Chen *et al.* [2024] note that the domain mismatch between LLM pre-training data and client data in federated learning poses a critical challenge, potentially reducing model transfer effectiveness and introducing bias if LLM-generated synthetic data misaligns with local distributions. These issues are further compounded by the computational demands of large LLMs on local devices, highlighting the need for efficient training and optimization strategies. Moreover, Hu *et al.* [2024] highlights that, despite the privacy and decentralization advantages of federated learning, implementing LLMs faces significant challenges due to computational overhead and local device resource constraints.

Continual learning and continuous data streams are critical open challenges for Federated LLMs. Cheng *et al.* [2024] state that these models must constantly update their domain knowledge in dynamic real-world scenarios, where changing data distributions can lead to domain and concept drift. Similarly, Chen *et al.* [2024] note that continuous data streams pose difficulties, as clients may join or leave with differing data distributions, and integrating new data efficiently is challenging given the large size of LLMs and limited local resources.

Foundation models, like LLMs, require large-scale data, parameters, and computational resources, making them suitable for Federated Learning but also subject to similar challenges in heterogeneity, privacy, and robustness [Cheng *et al.*, 2024]. Therefore, extending solutions developed for LLMs to multimodal foundation models remains a significant barrier.

2 Background

2.1 Federated Learning (FL)

Federated Learning (FL) is a distributed machine learning framework introduced by McMahan *et al.* [2017] at Google AI with the objective of collaborative training models without the need to share data between the participants. The ini-

tial implementation for language prediction models on smartphones, using the Gboard as test, was proven promising when allowing clients to do the training locally, sending only the model updates to a central server which gathers all of them and redistributes the global model for the clients, instead of transferring huge volumes of data to a single processing place, such as the centralized methods do [Kim *et al.*, 2023]. Therefore, FL aims to address key challenges related to privacy, data ownership, and data locality in scenarios where heterogeneous data from various distributed devices contribute to the learning process [Mothukuri *et al.*, 2021].

There are two main architectural settings of FL: cross-device and cross-silo [Bao and Guo, 2022; Pais *et al.*, 2025]. In the cross-device setting, a large number of edge devices, such as smartphones, wearables, or IoT sensors, participate in the training process. These devices are often unreliable, intermittently connected, and have limited computational resources, which makes this setting highly scalable but also more challenging in terms of communication efficiency and client availability [Bao and Guo, 2022; Pais *et al.*, 2025]. In contrast, the cross-silo setting involves a smaller number of more reliable participants, such as hospitals, banks, or enterprise data centers. These clients typically have stable network connections and greater computational capacity, allowing for more consistent training rounds and easier coordination [Bao and Guo, 2022; Pais *et al.*, 2025]. While both settings aim to preserve data privacy by keeping data localized, they differ significantly in their system constraints, trust assumptions, and optimization strategies.

The applications of FL usually implement the Federated Averaging (FedAvg) algorithm [McMahan *et al.*, 2017], which performs local training on clients and then averages the resulting model parameters on a central server. While FedAvg is simple and communication-efficient, it is known to suffer from performance degradation in non-IID settings due to model divergence during aggregation [Ahmad *et al.*, 2023; Kim *et al.*, 2023]. To address this issue, various techniques have been proposed, such as FedProx [Li *et al.*, 2020], which prevents training instability of local model updates by altering the loss function to penalize updates that deviate too far from the global model. Other approaches include SCAFFOLD [Karimireddy *et al.*, 2020], which uses control variates to correct client drift caused by data heterogeneity, and FedDyn [Acar *et al.*, 2021], which incorporates dynamic regularization to align local and global objectives. Strategies, as pFedMe [T. Dinh *et al.*, 2020], and clustering-based methods, such as IFCA [Ghosh *et al.*, 2020], have also shown promise by adapting models to client-specific data distributions or grouping clients with similar characteristics. These techniques aim to mitigate the negative effects of non-IID data and improve the convergence and generalization of federated models.

Privacy preservation, which is the core motivation in FL, is another challenge, and the researchers proposed several techniques to enhance the confidentiality of client data throughout the training process [Saha *et al.*, 2024]. One widely adopted method is Differential Privacy (DP), which introduces carefully calibrated noise into model updates or gradients before sharing them with the central server, ensuring that individual data points cannot be reverse-engineered

[Saha *et al.*, 2024; Wei *et al.*, 2020]. Another important strategy is Secure Aggregation, a cryptographic protocol that allows the server to compute aggregated model updates without accessing individual client updates in plain form [Saha *et al.*, 2024; Bonawitz *et al.*, 2017; So *et al.*, 2022]. Homomorphic Encryption (HE) also plays a role, enabling computations to be performed on encrypted data, although it is often computationally expensive [Saha *et al.*, 2024; Truex *et al.*, 2020; Han *et al.*, 2022].

2.2 Large Language Models (LLMs)

Large Language Models (LLMs) have revolutionized Natural Language Processing (NLP) by presenting significant performance in various tasks, such as language translation, summarization, text generation, and question answering [Sadot *et al.*, 2023; Wu *et al.*, 2024a]. These models, which have an extensive number of parameters, are trained on massive volumes of textual data, covering multiple domains, styles, and linguistic patterns [Wu *et al.*, 2024b; Iftikhar and Davy, 2024; Jiang *et al.*, 2024]. As a result, they are capable of capturing semantic nuances, learning complex language structures, and generating highly accurate and context-aware responses, often with quality comparable to that of humans [Iftikhar and Davy, 2024; Liu *et al.*, 2023].

These capabilities are related to the introduction of the Transformer architecture, proposed by Vaswani *et al.* [2017], which became the foundational building block for all LLMs. Based on self-attention mechanisms, Transformers associate different positions within the same sequence to generate contextualized representations of the data [Vaswani *et al.*, 2017]. This mechanism allows the model to identify how each element relates to others, regardless of the distance between input or output positions, efficiently capturing long-range dependencies [Vaswani *et al.*, 2017]. Furthermore, these models benefit from parallel computing capabilities, contributing to their high performance in NLP tasks.

In this context, models such as BERT, T5, GPT-2, GPT-3, and LLaMA illustrate the rapid evolution of LLMs in terms of capacity and performance. Launched in 2018, Bidirectional Encoder Representations from Transformers (BERT) [Tenney *et al.*, 2019] stood out for introducing bidirectional training in the Transformer architecture, with around 340 million parameters [Kim *et al.*, 2023]. The following year, the Text-to-Text Transfer Transformer (T5) [Raffel *et al.*, 2020] expanded on this by reformulating all NLP tasks into a text-to-text format, where both input and output are text sequences, reaching up to 11 billion parameters [Kim *et al.*, 2023]. In 2022, OpenAI launched the Chat Generative Pre-trained Transformer (ChatGPT), a large-scale application [Wu *et al.*, 2024b] based on GPT-3.5. The release of ChatGPT popularized the use of LLMs by making them accessible to the general public through conversational interfaces. The development of these generative models continued to grow rapidly, and shortly after, OpenAI integrated GPT-4 into ChatGPT, a version with expanded capabilities, including multimodal functionalities and significantly larger context windows [Wu *et al.*, 2024b]. In parallel, the emergence of open models like Large Language Model Meta AI (LLaMA 2) [Touvron *et al.*, 2023] and DeepSeek [DeepSeek-AI *et al.*, 2025] contributed to the democratization of this

technology by making robust, high-performance models accessible for research, development, and adaptation in diverse contexts.

Despite considerable advancements, LLMs still face significant limitations. Training these models demands high computational power [Yun *et al.*, 2023; Iftikhar and Davy, 2024; Sun *et al.*, 2025], mainly due to the massive data volumes involved. This characteristic also raises privacy and security concerns, as sensitive information may be inadvertently memorized and reproduced by the models [Yun *et al.*, 2023; Su *et al.*, 2024]. Additionally, the large number of parameters makes the fine-tuning process harder [Wu *et al.*, 2024b; Kumar and Chimalakonda, 2024], which increases computational demands [Jiang *et al.*, 2024], a critical issue for devices with hardware constraints, such as embedded systems or edge devices [Sun *et al.*, 2025]. Moreover, this process often requires domain-specific datasets that are not publicly available due to legal, ethical, or confidentiality concerns [Wu *et al.*, 2024b; Kuang *et al.*, 2024; Wu *et al.*, 2024a]. Under such conditions, data scarcity may prevent LLMs from delivering satisfactory performance [Kumar and Chimalakonda, 2024; Wu *et al.*, 2024a].

2.3 Federated Large Language Models (FedLLMs)

Federated Learning of Large Language Models (FedLLM) is a paradigm in which distributed clients collaboratively train or adapt language models without sharing raw data [Ye *et al.*, 2024]. As illustrated in Figure 1, each client is characterized by its local dataset, input text, and training procedure, while the federated server serves as an abstract global model that aggregates updates and distributes collective knowledge. This enables clients to benefit from global improvements while preserving data privacy.

Depending on the training objectives and resource constraints, FedLLM can be realized through pre-training, fine-tuning, or prompt learning. It has been applied in both pre-training and fine-tuning of models in distributed environments, which are fundamental stages in their architecture. This paradigm generates a range of opportunities and challenges because the significant size of LLMs, the complexity of their training, and the need for privacy protection have driven growing interest in FedLLM solutions. Furthermore, FedLLM can support the adjustment and parameterization of training models in federated settings, leveraging LLM outputs to improve performance. Depending on the desired outcome and methodology, these approaches can be categorized into distinct strategies, which are discussed in the following subsections.

2.3.1 FedLLM Pre-Training

The pre-training refers to the initial phase of training an LLM. In a federated learning context, clients participate in pre-training by processing their data locally, ensuring data privacy, while the model learns general language representations. In addition, pre-training techniques in LLMs are used to address various challenges in FL, such as reducing convergence time, dealing with non-independent and identically distributed (non-iid) data issues, and enabling FL models to perform multiple tasks.

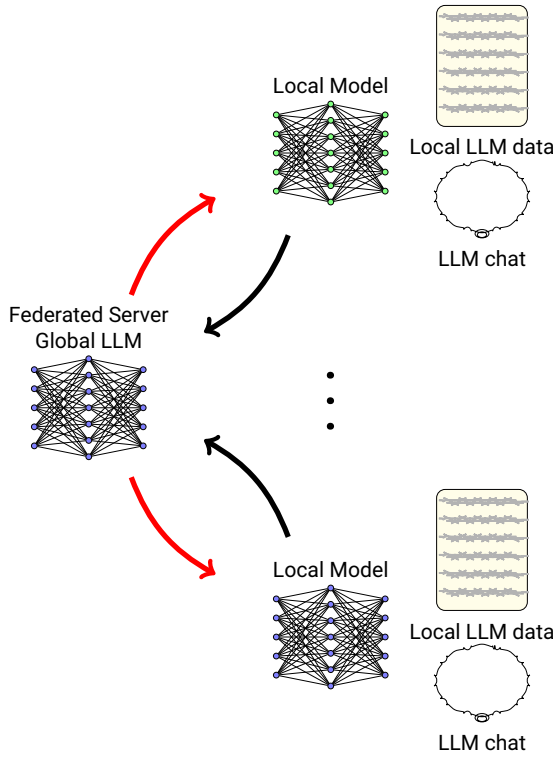


Figure 1. Schematic representation of FedLLM illustrating client-specific data, local training, and the abstract global model.

2.3.2 Federated Fine-Tuning and Prompt Learning

Fine-tuning aims to adapt a pre-trained model to specific tasks and is commonly used in LLMs to achieve better performance. In the federated context, several studies present new perspectives on the subject, such as the risks of data leakage and privacy breaches arising from fine-tuning, approaches to overcome limited computing capacity, and tools such as FwdLLM (*Efficient FedLLM using Forward Gradient*) [Xu et al., 2023], FedIT (*Federated Instruction Tuning*), among others presented in [Yao et al., 2024].

Among these resources, FwdLLM [Xu et al., 2023] is a framework that facilitates fine-tuning using the gradient descent method in FedLLM without backpropagation, motivated by issues such as high memory usage. The tool presents good results, such as scalability in convergence speed according to the increase in the number of training devices, and better energy efficiency compared to other methods. FedIT [Zhao et al., 2024], in turn, incorporates a fine-tuning technique, Low-Rank Adaptation (LoRA), with the goal of facilitating local training and reducing communication and computation costs, being extremely useful for federated environments with billions of clients.

2.3.3 Federated Knowledge Distillation and Transfer Learning

The integration of FedLLM can be classified into two distinct approaches. The first involves FKD (*Federated Knowledge Distillation*), while the second refers to TL (*Transfer Learning*). The first approach enables collaboration among multiple devices or servers by sharing model outputs instead of their weights. In the context of FedLLM, FKD can be used to create smaller and more efficient versions of an LLM, which is considered a larger and well-trained model, through knowl-

edge distillation within an FL system [Cheng et al., 2024]. The second approach aims to improve model performance in the target domain by transferring knowledge contained in the source domain. In the case of FedLLM, this process typically occurs by transferring weights or representations from a pre-trained model to a new target domain, where the model can be fine-tuned locally without the need to centralize the data.

2.3.4 Federated Hyperparameter Optimization and Reinforcement Learning

FHO (*Federated Hyperparameter Optimization*) and RFL (*Reinforcement Federated Learning*) emerge as complementary approaches applicable in the context of FedLLM. To reduce communication overhead in unstable networks, [Wang, 2024] presents an FHO strategy that combines GPT-4 with PSO (*Particle Swarm Optimization*), improving training performance and efficiency in FL environments.

3 Methodology

This section examines the use of FedLLMs from multiple perspectives, with particular emphasis on the challenges that underpin the research question. The systematic search methodology employed in this review follows the PRISMA guidelines [Moher et al., 2010], ensuring a rigorous, transparent, and reproducible approach to study selection and analysis, consistent with the methodological standards recommended by Higgins et al. [2019].

3.1 Data Collection Protocol and Search Strategy

The choice of databases and eligibility requirements was a crucial step before the review began. For this, notable platforms such as ACM Digital Library¹, Scopus², IEEE Xplore³ and Web of Science⁴ were used. The search string “(“federated learning” OR “distributed learning”) AND (“large language models” OR “transformers”)” was selected for the search. We considered peer-reviewed studies published between 2021 and 2024. Because some journals make accepted articles available online before assigning them to a specific journal issue, peer-reviewed studies that were electronically available during the search period (late 2024) but were subsequently assigned to 2025 issues, were also considered eligible.

Furthermore, we used the Parsifal platform with the aim of helping to organize the studies obtained.

All articles retrieved from the databases were initially considered. Eligibility criteria were then applied, and only studies meeting all inclusion criteria and none of the exclusion criteria were retained. These criteria are detailed in the following subsections.

3.1.1 Exclusion Criteria

In this review, studies were excluded if they met at least one of the predefined criteria: (1) articles discussing LLMs in non-federated contexts, including training on centralized servers,

¹<https://dl.acm.org/search/advanced>

²<https://www.scopus.com/>

³<https://ieeexplore.ieee.org/>

⁴<https://mjl.clarivate.com/>

blockchain-based frameworks, or any other distributed training approaches that differ from Federated Learning; (2) duplicate documents published in multiple versions, without substantial methodological changes regarding FedLLMs; (3) studies addressing Federated Learning in contexts other than LLMs (e.g., computer vision, IoT, healthcare), since the research question specifically investigates the integration of FL with LLMs; (4) works that do not clearly address the integration of federated models with LLMs, where such integration is not part of the proposed methodology; (5) studies not written in English; and (6) grey literature.

Given the rapidly evolving nature of FedLLM research, recently released preprints were exceptionally considered when no peer-reviewed version was yet available, as they were likely still undergoing the peer-review process. Conversely, preprints that had remained publicly available for more than two years without a subsequent peer-reviewed publication were considered grey literature and excluded from the final analysis.

Additionally, the last exclusion criterion (7) was secondary and tertiary studies, which synthesize reviews rather than providing original contributions. Secondary and tertiary studies not excluded by the other criteria were recorded separately for reference, but were not included in the systematic review process.

3.1.2 Inclusion Criteria

The inclusion criteria for this review were defined following established methodological standards for systematic literature reviews. Specifically, the inclusion criteria were: (1) studies had to be complete academic publications, ensuring that methods and results could be critically assessed and replicated, as recommended by Petticrew and Roberts [2006]; and (2) primary studies reporting methods in which the integration of FedLLM is explicitly part of the proposed methodology. This ensures that the selected studies directly address the research question and provide sufficient methodological detail for comparison and synthesis, in line with best practices for systematic reviews [Higgins *et al.*, 2019].

3.2 Data Collection Process

After retrieving studies using the search string described in subsection 3.1, the articles underwent four evaluation stages following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines Page *et al.* [2021] to ensure transparency and reproducibility. These stages included duplicate removal, title and abstract screening, full-text screening, and final eligibility assessment based on the predefined inclusion and exclusion criteria. All evaluation stages were conducted through peer cross-review.

In addition to the authors, other FutureLab researchers interested in the subject also participated in the exclusion stages of the review process, with credits provided in the acknowledgements. In total, each phase involved a subset of 9 reviewers.

3.3 Data Extraction and Synthesis

Following the careful selection of studies, the next phase focused on extracting relevant information from the included articles to support a systematic synthesis of findings. Since this review follows a challenge-oriented structure, the extrac-

tion process was designed to directly address the central research question.

Accordingly, the extraction process focused primarily on identifying and categorizing the FedLLM challenges addressed by each study. The extracted information was then synthesized through a cross-study challenge mapping approach to support the identification of research trends, underexplored challenges, and open gaps in the current FedLLM literature.

3.4 Quality Assessment

Following the study selection phase, the next step focused on assessing the methodological robustness of the included works. This assessment did not serve as an exclusion criterion, but rather supported the interpretation of the findings.

In this phase, we evaluated each study based on five criteria adapted to the context of Federated Learning and Large Language Models: (1) clarity of the study objective, (2) adequacy of methodological description, (3) presence of experimental validation, (4) discussion of limitations, and (5) explicit connection to FedLLM challenges. Each study was scored in a binary manner (0 or 1), resulting in a total score ranging from 0 to 5.

4 Results and Discussion

Figure 2 provides an overview of the systematic review process. The search across the four selected databases retrieved 234 studies. After removing duplicates with the Parsifal⁵ review tool, we reduced this number to 168 studies. We then applied the exclusion criteria in two stages: peer cross-review of titles and abstracts reduced the set to 63 studies, and peer cross-review of full-text articles reduced it further to 34 studies. Finally, applying the inclusion criteria through full-text eligibility assessment in a peer cross-review, we identified 24 studies for the final selection, which are presented in Table 1.

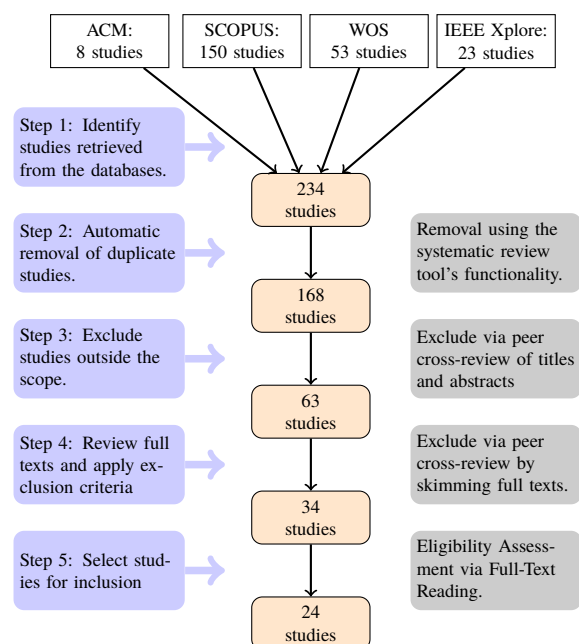


Figure 2. Schematic overview of the conducted systematic review.

⁵Parsifal: a software tool designed to support and manage systematic literature reviews.

4.1 Key challenges in FedLLMs identified in our systematic review

This subsection shows how the challenges related to FedLLMs are addressed across the studies included in our systematic review. Table 2 correlates the challenges (rows) with the individual articles (columns). A checkmark (✓) indicates that a particular challenge has been explicitly addressed in the corresponding article, allowing readers to quickly identify which studies contribute to the mitigation of each challenge. Each challenge is subsequently discussed, providing a detailed account of the strategies or solutions proposed in the literature.

4.1.1 Legal, Ethical, and Property Rights Aspects

The integration of LLMs into Federated Learning entails legal, ethical, and intellectual property challenges, requiring multidisciplinary approaches to ensure privacy, compliance, and responsible deployment. To address these issues, Ye *et al.* [2024] introduce OpenFedLLM, a research-oriented framework that establishes a practical setting termed Federated Value Alignment (FedVA). FedVA seeks to align LLMs with clients' ethical guidelines and societal values. Nevertheless, the heterogeneity of value preferences across clients remains a critical challenge for effective alignment.

4.1.2 Impact of Heterogeneous Data, Devices, and Resources

Heterogeneity in data, devices, and resources emerges as one of the most pervasive challenges in the FedLLM context. As noted by Cheng *et al.* [2024], differences in client data and system resources can cause model divergence, uneven training progress, and suboptimal performance. Yao *et al.* [2024] underscore that such heterogeneity constitutes a central research barrier by undermining global model convergence. Similarly, Hu *et al.* [2024] emphasize that variations in local dataset distributions and client hardware/software capabilities can compromise model performance. Building on these general observations, the literature has addressed heterogeneity in more specific ways, spanning statistical non-IID distributions, system and resource constraints, and advanced mitigation mechanisms.

Several works explore the impact of statistical heterogeneity on federated learning performance. Kumar and Chimalakonda [2024] simulate heterogeneity by splitting a dataset among clients with different repository ratios, reflecting real-world data imbalances. Along the same lines, Thapa *et al.* [2023] examine phishing-to-legitimate sample ratios and varying dataset sizes, showing that FL remains resilient through weighted averaging despite asymmetric distributions. Vasconcelos *et al.* [2023] further study heterogeneity by partitioning the eRisk2021 dataset under homogeneous, heterogeneous, and intermediate conditions to analyze the effects of class imbalance. Expanding on this, Gholamiangonabadi and Grolinger [2024] explicitly highlight how non-IID data can degrade FL performance, particularly in edge scenarios where privacy and security must also be preserved. Other studies emphasize the sensitivity of FL to the number of clients: Hilmkil *et al.* [2021] demonstrate that smaller partitions and contradictory models can destabilize performance, while Sujaykumar Reddy *et al.* [2024]

illustrate how inherent class imbalance in Amazon Product datasets interacts with communication overhead to exacerbate heterogeneity effects.

Another line of research focuses on disparities in resource availability and device capabilities. Kim *et al.* [2023] identify that while pre-trained models can mitigate data heterogeneity, they also introduce large communication costs. Their adapter-based mechanism reduces transmission requirements and training time while preserving efficiency in diverse datasets. Complementarily, Gholamiangonabadi and Grolinger [2024] also acknowledge computational limitations at the edge, linking them with privacy and robustness challenges. Iftikhar and Davy [2024] stress that FL with LLMs magnifies these problems due to the considerable demands of memory, communication, and computation, further complicating deployment in real-world heterogeneous environments.

Beyond characterizing heterogeneity, recent work proposes strategies to address it directly. Seo *et al.* [2023] explore the novel idea of leveraging LLMs themselves to adapt submitted code to heterogeneous client environments while simultaneously detecting potential threats, thus tackling both statistical and system heterogeneity. Meanwhile, Yun *et al.* [2023] introduce FCLM, a framework based on clustered sampling that groups clients with similar distributions. This approach reduces bias and accelerates convergence, demonstrating effectiveness in mitigating privacy-preserving federated learning challenges under non-IID settings.

4.1.3 Communication Overhead and Efficiency Constraints

Lowering communication costs is one of the main draws of Federated Learning, as this methodology negates the need for data transmission. This is specially impactful for LLM tasks, as textual data tends to be resource intensive. For many of these tasks, Transformers are the state-of-the-art architecture, which is famously data intensive. Integrating Federated Learning with Transformer models offers significant advantages in terms of privacy preservation and personalized recommendation capabilities. Experimental results Kim *et al.* [2023]; Wu *et al.* [2024a]; Sujaykumar Reddy *et al.* [2024] show a remarkable improvement in transmission speed. Such improvements enable the use of LLM in real-world Federated Learning environments. One example of this is presented in Vasconcelos *et al.* [2023], where the authors use Federated Learning techniques to train a classifier for depression - a needed application - while still preserving the data privacy of individuals. Despite these advances, deploying FL for LLMs on edge devices still faces key challenges, such as limited access to model parameters, high computational demands on local clients, and excessive communication overhead, hindering its practical use Li *et al.* [2024].

Another research focus consists of optimization in the updating of weights of the neural network parameters to further improve communication costs Wang [2024]; Zhao *et al.* [2024]. This optimization can also be achieved via Low-Rank Adaptation (LoRA) in the fine-tuning step Sun *et al.* [2025]. A significant research gap remains in fine-tuning LLMs in FL environments, which persists to be computationally expensive on the client side, even with the parameter-

Table 1. Summary of studies retained for final analysis and quality assessment.

Nº	Article	Reference	QA1	QA2	QA3	QA4	QA5	Total
1	A Novel Approach to Efficient Multilabel Text Classification: BERT-Federated Learning Fusion	Sadot <i>et al.</i> [2023]	1	1	1	1	1	5
2	An API Sensitive Data Identification Method Based on the Federated Large Language Model	Wu <i>et al.</i> [2024b]	1	1	1	0	1	4
3	Code Summarization without Direct Access to Code - Towards Exploring Federated LLMs for Software Engineering	Kumar and Chimalakonda [2024]	1	1	1	1	1	5
4	Efficient Federated Intrusion Detection in 5G Ecosystem Using Optimized BERT-Based Model	Adjewa <i>et al.</i> [2024]	1	1	1	0	1	4
5	Efficient Federated Learning with Pre-Trained Large Language Model Using Several Adapter Mechanisms	Kim <i>et al.</i> [2023]	1	1	1	1	1	5
6	Evaluation of Federated Learning in Phishing Email Detection	Thapa <i>et al.</i> [2023]	1	1	1	1	1	5
7	Exploring FL to Trace Depression in Social Media with Language Models	Vasconcelos <i>et al.</i> [2023]	1	1	1	1	1	5
8	FedBiOT: LLM Local Fine-tuning in Federated Learning without Full Model	Wu <i>et al.</i> [2024a]	1	1	1	0	1	4
9	FL for Sentiment Analysis with Non-IID Data: DL Model Sensitivity	Gholamiangonabadi and Grolinger [2024]	1	1	1	1	1	5
10	FL Using GPT-4 Boosted PSO for Compact NAS	Wang [2024]	1	1	1	0	1	4
11	Federated Low-Rank Adaptation for Large Models Fine-Tuning Over Wireless Networks	Sun <i>et al.</i> [2025]	1	1	1	0	1	4
12	Federated Prompting and Chain-of-Thought Reasoning for Improving LLMs Answering	Liu <i>et al.</i> [2023]	1	1	1	1	1	5
13	FederatedScope-LLM: A Comprehensive Package for Fine-tuning Large Language Models in Federated Learning	Kuang <i>et al.</i> [2024]	1	1	1	1	1	5
14	FedNPC: A FL Framework for LLMs in Game NPCs	Hong <i>et al.</i> [2023]	1	1	0	1	1	4
15	FedQAS: Privacy-Aware MRC with FL	Ait-Mlouk <i>et al.</i> [2022]	1	1	1	0	1	4
16	FedsLLM: Federated Split Learning for LLMs Over Communication Networks	Zhao <i>et al.</i> [2024]	1	1	1	0	1	4
17	Flexible and Secure Code Deployment in Federated Learning using LLMs: Prompt Engineering to Enhance Malicious Code Detection	Seo <i>et al.</i> [2023]	1	1	1	1	1	5
18	Low-Parameter Federated Learning with Large Language Models	Jiang <i>et al.</i> [2024]	1	1	1	0	1	4
19	OpenFedLLM: Training LLMs on Decentralized Private Data via FL	Ye <i>et al.</i> [2024]	1	1	1	1	1	5
20	Privacy-Preserving Federated Learning through Clustered Sampling on Fine-Tuning Distributed Non-IID Large Language Models	Yun <i>et al.</i> [2023]	1	1	1	0	1	4
21	Reducing Carbon Footprint in AI: A Framework for Sustainable Training of Large Language Models	Iftikhar and Davy [2024]	1	1	0	0	1	3
22	Scaling FL for Fine-Tuning of LLMs	Hilmkil <i>et al.</i> [2021]	1	1	1	1	1	5
23	Towards Confidential Chatbot Conversations: A Decentralised Federated Learning Framework	Su <i>et al.</i> [2024]	1	1	1	0	1	4
24	Transformer-Based FL Models for Recommender Systems	Sujaykumar Reddy <i>et al.</i> [2024]	1	1	1	1	1	5

efficient, energy-preserving fine-tuning (PEFT) algorithms Iftikhar and Davy [2024]. Experimental results in Kuang *et al.* [2024] show that algorithms with federated fine-tuning can significantly outperform their local counterparts, with very competitive results against those under the global scenario.

4.1.4 Preserving Data Privacy

Data privacy preservation is a central challenge in the integration of LLMs with FL, since the sharing of information during training can expose sensitive data. To mitigate this risk, the analyzed works explore different strategies. The most fundamental consists of using the decentralized structure of federated learning itself, which ensures that training data remains stored locally, while only model updates are shared [Kumar and Chimalakonda, 2024; Adjewa *et al.*,

2024; Gholamiangonabadi and Grolinger, 2024; Sun *et al.*, 2025; Liu *et al.*, 2023; Ait-Mlouk *et al.*, 2022]. Beyond this standard approach, other research proposes complementary techniques to reinforce data security. The study in Wu *et al.* [2024b], for example, employs an LLM to identify sensitive data in APIs in real-time, allowing for the preemptive blocking of risky APIs. The work in Vasconcelos *et al.* [2023], to ensure privacy and efficiency in the exchange of information, uses the FedOpt algorithm so that clients can share compressed gradients protected with additive homomorphic encryption. Another strategy is presented in Yun *et al.* [2023], which proposes the Privacy-Preserving Federated Learning through Clustered Sampling (FCLM) framework, a clustered sampling technique based on gradient similarities to reduce the possibility of information leakage. Finally, Su *et al.* [2024] introduces an architectural change with Decentralized

Table 2. Challenges in Integrating Federated Learning with Large Language Models Across 24 Articles.

Challenge Description	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Multimodal Foundation Models Integration																								
Data Erasure for Federated LLMs																								
Legal, Ethical and Property Rights Aspects																			✓					
Environmental sustainability and energy efficiency																					✓			
Federated Learning over Black-Box LLMs										✓							✓							
Personalization and Domain Adaptation								✓			✓													
Continual Learning and Continuous Data Streams		✓																					✓	
Communication Overhead and Efficiency Constraints					✓		✓	✓						✓										✓
Security Threats and Robustness Issues	✓	✓		✓			✓		✓			✓					✓							
Heterogeneous Data, Devices, and Resources			✓		✓	✓	✓		✓				✓				✓			✓	✓	✓		✓
Computational Efficiency and Resource Constraints	✓	✓	✓	✓				✓			✓	✓	✓			✓		✓			✓	✓		
Preserving Data Privacy	✓	✓	✓	✓			✓		✓		✓	✓			✓				✓			✓	✓	✓

Federated Learning (DFL), a blockchain-based model where participants collaborate directly in a peer-to-peer manner, sending encrypted model updates. This approach is strengthened by the use of differential privacy, utilized during the evaluation phase to protect against inversion attacks and prevent the leakage of private data, and secure multi-party computation (SMPC), a protocol that allows for the collaborative aggregation of models without participants revealing the raw parameters of their individual models.

4.1.5 Federated Learning over Black-Box LLMs

The proprietary nature of black-box LLMs constrains their direct integration with federated learning, since model weights and internal architectures are inaccessible. Nevertheless, Seo *et al.* [2023] demonstrated that prompt-based techniques can partially bridge this gap. They used OpenAI's GPT-3.5 and GPT-4 to automatically tailor a single codebase to heterogeneous federated nodes and to detect potentially malicious code.

Similarly, Wang [2024] incorporated a black-box LLM into a neural-architecture-search framework for federated learning, where GPT-4 generated the initial population of architectures in a Particle Swarm Optimization (PSO) strategy.

These studies suggest that, although black-box LLMs cannot be directly trained in federated settings due to restricted access to internal parameters, they can still support federated learning workflows through auxiliary roles such as prompt engineering, automated code adaptation, and architecture generation. However, the lack of direct access to model weights remains a major limitation for conventional federated optimization strategies.

4.1.6 Security Threats and Robustness Issues

While LLM training benefits from the decentralized architecture of FL, security threats and robustness issues become

key factors to successful training. Data poisoning and non-IID data, for example, can compromise model convergence and reliability.

To address the robustness issues, different strategies have been proposed. Sadot *et al.* [2023] suggest a combination of Horizontal Federated Learning (HFL) with a pre-trained BERT model to perform preprocessing to check and correct imbalances in the training data. Complementarily, Gholamiangonabadi and Grolinger [2024] investigate the behavior of architectures such as FFNN, LSTMs, and Transformers when faced with different types and degrees of non-IID data, using the FedAvg algorithm to reduce information exchange. Vasconcelos *et al.* [2023] use BERT as a baseline model for detecting symptoms of depression, a context in which the impact of data heterogeneity, computational efficiency, and privacy were evaluated while still preserving the data privacy of individuals. Furthermore, considering that some models are too large to be trained locally, Liu *et al.* [2023] investigate how to improve the accuracy of LLMs in the cloud to answer frequently asked questions without the need for complex fine-tuning.

In the context of security threats, other approaches are explored. The work Adjewa *et al.* [2024] also uses BERT, but focuses on anomaly-based intrusion detection. Along the same lines of security, Seo *et al.* [2023] validate code adaptability in heterogeneous FL environments that employ LLMs, detecting and correcting critical instances of malicious code with high probability. Finally, addressing privacy and efficiency, Wu *et al.* [2024b] identify sensitive data during processing and then apply fine-tuning with LoRA to adapt the model.

4.1.7 Personalization and Domain Adaptation

While LLMs have demonstrated general-purpose capabilities, personalization and domain adaptation remain a signif-

ificant challenge, particularly in a federated learning framework, since it introduces complexities to an effective personalization, such as data heterogeneity and varying resource constraints. To address this, Wu *et al.* [2024a] have proposed Federated Bi-level Offsite Tuning (FedBiOT), a method that separates the model into distinct components for domain-specific adaptation and general knowledge. FedBiOT is centered on the division of a compressed LLM into an emulator and an adapter. This involves the server “aligning its performance with the full model”, as described by the authors, and is achieved through knowledge distillation to transfer the original LLM’s general linguistic patterns to the emulator. Simultaneously, clients “fine-tune a lightweight yet important part of the compressed model, referred to as an adapter” using techniques like LoRA (Low-Rank Adaptation), allowing them to assimilate domain-specific patterns without updating the entire model. This effectively minimizes the negative impact of data discrepancies, ensuring that the model can be personalized and adapted to diverse contexts while maintaining the capabilities of the full model.

Furthermore, Sun *et al.* [2025] note that “it is imperative to fine-tune LLMs for delivering personalized and accurate services” as well as for making them “more precise and tailored to the demands of different downstream tasks.” To address this, a novel approach integrates LoRA (Low-Rank Adaptation) and AirComp (Over-the-Air Computation). LoRA reduces the number of parameters needing to be trained and is combined with AirComp, which leverages wireless communication to more efficiently aggregate updates from multiple devices simultaneously, reducing communication costs. This framework, called AirFL-LoRA, handles the challenge of optimizing wireless resources through techniques like MIMO (Multiple-Input Multiple-Output) and OFDM (Orthogonal Frequency Division Multiplexing), as well as a new rank selection scheme. The entire system is optimized to minimize the optimality gap by using a complex alternating algorithm, demonstrating an approach to making personalized and accurate LLM services viable at scale.

4.1.8 Computational Efficiency and Resource Constraints

Computational overhead and local device resource constraints are common concerns when dealing with FL. Furthermore, computational efficiency in LLM tasks is a critical challenge intrinsic to the problem [Hu *et al.*, 2024]. Therefore, in FedLLM, researchers address these problems with different strategies. Wu *et al.* [2024b] present a solution with parameter-efficient LoRA fine-tuning method and continual learning to address the computational resource exhaustion issue in their context. Other authors such as Kumar and Chimalakonda [2024], Sun *et al.* [2025], Hilmkil *et al.* [2021] and Jiang *et al.* [2024] also present solutions using fine-tuning techniques, Iftikhar and Davy [2024] and Kuang *et al.* [2024] incorporated a parameter efficient fine-tuning technique (PEFT) and Adjewa *et al.* [2024] compressed the BERT-based model in order to reduce the required storage size.

Zhao *et al.* [2024] proposed FedsLLM, an approach that combines Split Learning (SL), Federated Learning and

LoRA to deal with computational efficiency constraints. This integration enables parallel processing on distributed clients, while the model is divided into client-server subnets. Another strategy presented by Wu *et al.* [2024a] formulates a bi-level optimization, using an emulator and an adapter, avoiding full model fine-tuning. Even with different approaches for each specific context, the challenge is still an open one and difficult to deal with, as mentioned by Liu *et al.* [2023] that tries to improve the responses of LLMs and also by Sadot *et al.* [2023] with multi-label classification problems.

4.1.9 Continual Learning and Continuous Data Streams

The challenge of continual learning in Federated Large Language Models (FedLLM) lies in enabling models to assimilate knowledge from continuous, non-stationary, and heterogeneous data streams without succumbing to catastrophic forgetting. To address this challenge, Wu *et al.* [2024b] propose an API-sensitive data identification framework based on FedLLM, which combines Low-Rank Adaptation (LoRA) with federated training. Their method explicitly introduces a continual learning mechanism “to prevent crucial knowledge in the global model from being forgotten during local model training” [Wu *et al.*, 2024b], thereby preserving knowledge stability while accommodating new data. Complementarily, Su *et al.* [2024] present a decentralized federated learning framework for confidential chatbot conversations, where blockchain and Secure Multi-Party Computation (SMPC) enable iterative training. According to the authors, “this iterative process allows the model to learn from various data distributions and adapt to diverse user preferences, which promote continuous model refinement” [Su *et al.*, 2024]. Together, these works illustrate that addressing continual learning and continuous data streams in FedLLM demands both knowledge-preservation mechanisms and systemic safeguards. Wu *et al.* [2024b] address the risk of catastrophic forgetting to ensure stable adaptation, while Su *et al.* [2024] reinforce the continuity of learning through decentralized, privacy-preserving refinement.

4.1.10 Environmental sustainability and energy efficiency

Iftikhar and Davy [2024] tackle energy efficiency challenges in FedLLM by proposing a framework that combines federated learning with renewable energy sources. By prioritizing clients running on clean energy during training, it reduces carbon emissions while preserving collaborative model development, promoting more sustainable and responsible AI practices.

Despite the relevance of sustainability concerns in large-scale AI systems, this remains a largely underexplored topic in the current FedLLM literature.

4.2 Strategies and Trade-offs in FedLLM Research

Across the selected studies, some recurring strategies emerge as dominant approaches for mitigating FedLLM challenges. A significant portion of the literature focuses on improving computational and communication efficiency through parameter-efficient fine-tuning techniques such as LoRA,

PEFT, adapters, and split learning. Studies such as Jiang *et al.* [2024], Ye *et al.* [2024], Sun *et al.* [2025], Adjewa *et al.* [2024], and Wu *et al.* [2024a] adopt these strategies to reduce the number of trainable parameters and communication overhead while preserving acceptable performance levels.

Another recurring trend involves addressing heterogeneous and resource-constrained federated environments. Approaches based on clustered sampling, lightweight adapters, emulator-based optimization, and compressed BERT models attempt to mitigate limitations associated with non-IID data distributions, unstable communication conditions, and limited client-side computational resources [Yun *et al.*, 2023; Wu *et al.*, 2024a; Adjewa *et al.*, 2024; Gholamiangonabadi and Grolinger, 2024]. Similarly, Wang [2024] explores GPT-4-assisted neural architecture search to improve compactness and robustness under unstable network conditions.

Privacy preservation and secure collaboration also appear as central concerns in FedLLM research. Different studies investigate federated prompting, secure multi-party computation, decentralized training, and privacy-aware learning mechanisms to avoid direct data sharing while maintaining collaborative model optimization [Liu *et al.*, 2023; Su *et al.*, 2024; Ait-Mlouk *et al.*, 2022]. However, these approaches often introduce additional computational complexity, communication overhead, or system design constraints.

Overall, although the proposed methodologies demonstrate promising results in their respective application domains, most solutions remain highly context-dependent and are commonly validated in restricted experimental settings, specialized datasets, or narrow operational scenarios. This indicates that, despite recent advances, several FedLLM strategies still face important trade-offs involving efficiency, scalability, robustness, privacy preservation, and generalizability.

4.3 Recent Advances in FedLLMs

While conducting this systematic review covering the 2021–2024 period, research on FedLLMs continued to evolve rapidly. To contextualize our findings with the most recent developments, we performed an updated search across the same databases using identical search strings. In this search, we observed that FedLLM research has become a significantly more prominent topic in recent years. This trend is highlighted by the temporal analysis of retrieved publications from 2021 to 2026 (Figure 3), which demonstrates remarkable growth in scientific output across all four analyzed databases (ACM, Scopus, Web of Science, and IEEE).

During the initial period (2021–2022), research activity remained modest and was primarily concentrated in Scopus, with limited representation in Web of Science and no records identified in ACM or IEEE. A clear turning point occurred in 2023, marked by active contributions across all four databases and a substantial increase in total publication volume. This upward trajectory escalated through 2024 and peaked in 2025, where Scopus led significantly with 376 articles, complemented by strong expansions in Web of Science (104) and IEEE (97). Although the figures for 2026 reflect a slight decline in raw numbers compared to 2025, it is important to note that data collection was conducted at the beginning of the second semester of 2026. Given this partial

indexing window, the annual total for 2026 is highly likely to surpass the record set in 2025 as remaining publications are indexed throughout the rest of the year, confirming the sustained momentum and rapid maturation of this research domain.

The objective of this complementary analysis is not to extend the original systematic review, but rather to provide readers with an overview of the major developments that have emerged since the review period. To provide this complementary perspective, we analyzed a representative sample of recent peer-reviewed studies published in high-impact journals. These studies were selected to illustrate recent advances in FedLLMs and to examine how the latest research relates to the challenges identified in our systematic review.

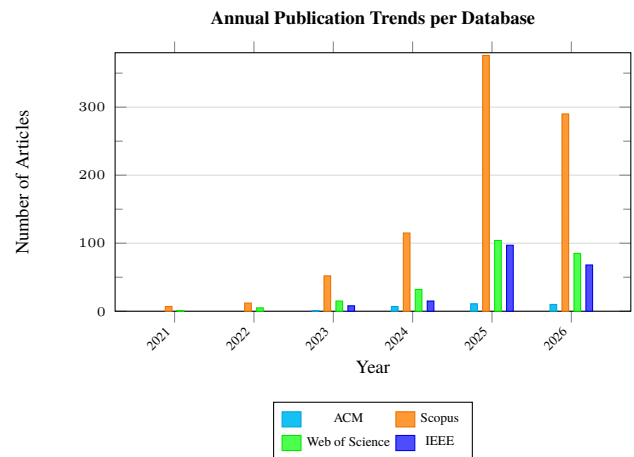


Figure 3. Annual publication trends across databases (2021–2026).

The recent literature indicates that the challenges identified in our systematic review remain central to current FedLLM research. Most recent studies focus on improving communication efficiency, computational efficiency, and resource management through parameter-efficient fine-tuning and adaptive model aggregation strategies. Representative studies include FedFA [Zhang *et al.*, 2025a], FedLoDrop [Xie *et al.*, 2026], FedALoRA [Yi *et al.*, 2025], Federated LoRA Fine-Tuning of LLMs with Only Transmitting Matrix A or B [Chen *et al.*, 2026], LaVFL [Zhang *et al.*, 2025c], FLGuard Zhang *et al.* [2025b], Tool-Aided Evolutionary LLM for Generative Policy Toward Efficient Resource Management in Wireless Federated Learning [Tan *et al.*, 2025], and A Web-Based Solution for Federated Learning With LLM-Based Automation [Mawela *et al.*, 2025].

Beyond communication and computational efficiency, recent studies also address heterogeneous environments, privacy preservation, security, personalization, fairness, and trustworthy federated intelligence. These aspects are addressed in A Privacy-Preserving and Trustworthy Inference Framework for LLM-IoT Integration [Han *et al.*, 2025], FedShieldLLM [Sun *et al.*, 2026a], Efficient and Privacy-Preserving Federated Knowledge Learning for Distributed LLM [Sun *et al.*, 2026b], R-SFLLM Djuhera *et al.* [2025], Energy- and AoI-Aware Hierarchical Personalized Federated Learning [Solat and Lee, 2026], Fine-Tuning LLMs in Federated Learning with Fairness-Aware Prompt Selection Jiang *et al.* [2025], When Fine-Tuning LLMs Meets Data Privacy Luo *et al.* [2026], Federated TrustChain Zuo *et al.*

[2026], and AES Cryptography Enabled Responsible Federated Foundation Model [Hasan *et al.*, 2025]. Overall, these studies corroborate the challenges previously identified in this review while demonstrating their applicability across emerging scenarios.

In addition to reinforcing the existing taxonomy, the recent literature also points to new research directions. In particular, LaVFL Zhang *et al.* [2025c] and FLAGuard Zhang *et al.* [2025b] introduce mechanisms for efficiently verifying the integrity and correctness of aggregated model updates, suggesting that trustworthy verification may emerge as a new research challenge for future FedLLM systems.

Recent surveys further corroborate the relevance of the challenges identified in our systematic review. Fan *et al.* [2025] discuss ten open research problems for federated foundation models, several of which directly align with our taxonomy, including communication efficiency, heterogeneous resources, privacy preservation, personalization, and trustworthy federated intelligence. Likewise, Yan *et al.* [2025] provide a comprehensive overview of federated LLM fine-tuning frameworks and identify communication efficiency, computational constraints, and scalable adaptation strategies as key research directions. The consistency between these independent surveys and our findings provides additional evidence that the taxonomy proposed in this review captures the fundamental research challenges that continue to drive the evolution of FedLLMs.

As a complementary perspective, we performed a bibliometric analysis based on the keyword co-occurrence network generated using VOSviewer [Arruda *et al.*, 2022]. As illustrated in Figure 4, the network is centered on the term federated learning, confirming that the recent literature remains predominantly grounded in this distributed learning paradigm. Strong connections between this central node and terms such as language model, fine tuning, and natural language processing further indicate that a substantial portion of recent research is devoted to adapting and optimizing large language models in federated environments.

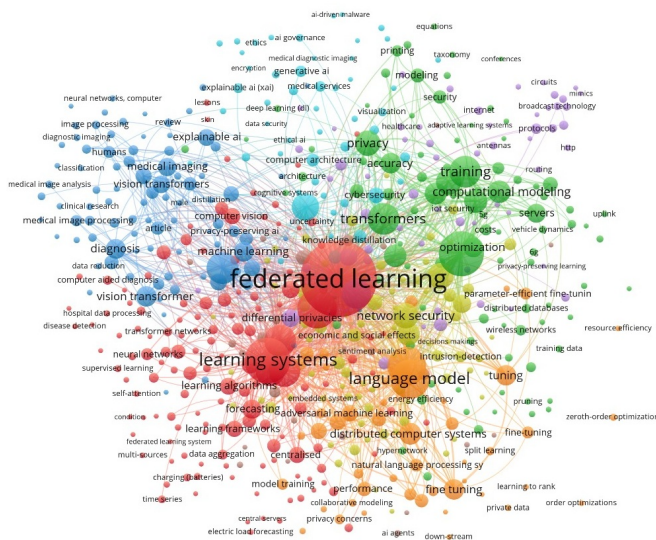


Figure 4. Bibliometric analysis based on the keyword co-occurrence network generated using VOSviewer.

The co-occurrence network also provides complemen-

tary evidence supporting the challenges identified in this review. Clusters containing terms such as optimization, parameter-efficient fine-tuning, pruning, and energy efficiency reflect the continued emphasis on communication efficiency, computational efficiency, and resource constraints. Likewise, keywords including privacy, differential privacy, and cybersecurity highlight the central role of privacy preservation and security in recent FedLLM research. Overall, the bibliometric analysis reinforces that the primary research directions observed in the recent literature remain strongly aligned with the challenges identified throughout this systematic review.

4.4 Unaddressed challenges, gaps and open opportunities

From these results, we observe that the integration of multimodal foundation models in a federated environment constitutes a significant challenge. Consequently, extending solutions developed for FedLLMs to multimodal foundation models remains a major barrier. Furthermore, none of the works included in our systematic review addressed data erasure, which is closely related to the *right to be forgotten*, as emphasized by Cheng *et al.* [2024]. Together, these gaps highlight clear open challenges and opportunities for future research in the FedLLM context.

Furthermore, although mentioned, some challenges were relatively neglected. Several FedLLM-related issues appeared in only a few studies and therefore remain underexplored. Specifically, challenges related to legal, ethical, and property rights have been explored by Ye *et al.* [2024], while environmental sustainability and energy efficiency have been examined by Iftikhar and Davy [2024]. Federated learning over black-box LLMs has been investigated by [Wang, 2024] and [Seo *et al.*, 2023], personalization and domain adaptation by [Wu *et al.*, 2024a] and [Sun *et al.*, 2025], and continual learning with continuous data streams by [Wu *et al.*, 2024b] and [Su *et al.*, 2024]. Overall, these topics have received limited attention in the current literature. As such, they represent opportunities for future research aimed at developing more comprehensive solutions in the FedLLM context.

The remaining challenges were more frequently addressed and thus emerge as particularly relevant topics in the current literature. Most proposed methodologies to mitigate these challenges were developed for narrow application contexts, specialized domains, limited operational settings, or experimental evaluations on particular datasets, which limits their broader applicability. For example, Vasconcelos *et al.* [2023] addressed communication overhead and efficiency constraints by training a classifier for trace depression detection in social media, while Kuang *et al.* [2024] tackled security threats and robustness issues in federated prompting and chain-of-thought experiments. Similar limitations can be observed in studies addressing computational efficiency and resource constraints, data privacy preservation, and heterogeneous data, devices, and resources. Nevertheless, these studies provide valuable insights that can guide future research, inspiring new directions and offering possibilities for mitigating the impact of these remaining challenges.

5 Conclusion

Guided by the central research question *How are the key challenges in Federated Large Language Models being addressed in current research, and what gaps remain to be explored?* this systematic review examined twelve key challenges in the FedLLM context. The analysis shows that some remain entirely unaddressed, namely the integration of multimodal foundation models and the implementation of data erasure mechanisms aligned with the right to be forgotten. Others have been only marginally explored, including legal, ethical, and property rights, environmental sustainability and energy efficiency, black-box learning, personalization and domain adaptation, and continual learning with continuous data streams. In contrast, challenges such as communication overhead and efficiency constraints, security threats and robustness issues, heterogeneous data, devices and resources, computational efficiency and resource constraints, and data privacy preservation have received more frequent attention, although proposed solutions are often limited to narrow contexts or experimental evaluations.

By consolidating these insights, this review contributes to a clearer understanding of the state of the art in FedLLMs, highlighting both areas of progress and substantial gaps that persist. Addressing these challenges requires advancing beyond domain-specific approaches toward more generalizable, responsible, and sustainable solutions, which reinforces the critical relevance of FedLLMs in the evolving landscape of LLM technologies.

It is important to note that, due to publication and indexing delays, very recent studies may not be fully captured. To partially mitigate this limitation, we included a complementary analysis based on a sample of recent peer-reviewed studies published in 2025–2026, aiming to highlight key recent advances in FedLLMs. However, we emphasize that this subset does not constitute a new systematic review and does not cover all recent publications on the topic, which continues to grow rapidly with the emergence of new approaches. Therefore, the findings of this review primarily reflect the state of the literature at the time the systematic review was conducted. Nevertheless, this limitation does not affect the identification of the main open challenges, which are expected to persist as the field continues to evolve. Another limitation of this review is the exclusion of non-English studies, which may introduce language bias.

Declarations

Authors' Contributions

Roberto G. Ribeiro: Conceptualization, Methodology, Data Curation, Formal Analysis, Writing – Original Draft, Visualization. Frederico G. Guimarães: Conceptualization, Supervision. Hélder S. Lima, Vinicius A. O. Martins, Allana T. Bastos, Larissa D. Santana, Beatriz C. G. Pereira, Carla B. Ferreira, Júlia B. F. Souza and Melissa C. C. Lima: Methodology, Data Curation, Formal Analysis, Writing – Original Draft. All authors read and approved this version of the manuscript.

Competing interests

The authors declare that they have no conflicts of interest.

Acknowledgements

We would like to thank the other FutureLab researchers that support the idea: Julia Zibetti Castilhos de Mello, Ester de Lima Pontes Andrade, Isabela Fernandes Guerra de Moraes and Marcus Vinicius Moraes Oliveira.

Funding

This work has been supported by the Brazilian agencies (i) National Council for Scientific and Technological Development (CNPq), Grant no. 304856/2025-8, “*Aprendizado de Máquina Colaborativo e Proteção à Privacidade*”; (ii) Coordination for the Improvement of Higher Education Personnel (CAPES) through the Academic Excellence Program (PROEX) and (iii) Foundation for Research of the State of Minas Gerais (FAPEMIG, in Portuguese).

Partially funded by the R&D&I agreement between UFMG and Fundep – Research Development Foundation: “Federated Machine Learning Program for Connected Vehicles”.

Availability of research artifacts

The spreadsheets related to the systematic review process, including title and abstract screening, full-text screening, eligibility assessment, and quality assessment, are provided as supplementary material.

References

- Acar, D. A. E., Zhao, Y., Navarro, R. M., Mattina, M., Whatmough, P. N., and Saligrama, V. (2021). Federated learning based on dynamic regularization. *arXiv preprint arXiv:2111.04263*. DOI: 10.48550/arxiv.2111.04263.
- Adjewa, F., Esseghir, M., and Merghem-Boulahia, L. (2024). Efficient federated intrusion detection in 5g ecosystem using optimized bert-based model. In *2024 20th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 62–67. DOI: 10.1109/WiMob61911.2024.10770340.
- Ahmad, A., Luo, W., and Robles-Kelly, A. (2023). Robust federated learning under statistical heterogeneity via hessian spectral decomposition. *Pattern Recognition*, 141:109635. DOI: 10.1016/j.patcog.2023.109635.
- Ait-Mlouk, A., Alawadi, S. A., Toor, S., and Hellander, A. (2022). Fedqas: Privacy-aware machine reading comprehension with federated learning. *Applied Sciences*, 12(6). DOI: 10.3390/app12063130.
- Arruda, H., Silva, E. R., Lessa, M., Proença Jr, D., and Bartholo, R. (2022). Vosviewer and bibliometrix. *Journal of the Medical Library Association: JMLA*, 110(3):392. DOI: 10.5195/jmla.2022.1434.
- Bao, G. and Guo, P. (2022). Federated learning in cloud-edge collaborative architecture: key technologies, applications and challenges. *Journal of Cloud Computing*, 11(1):94. DOI: 10.1186/s13677-022-00377-4.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., et al. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*. DOI: 10.48550/arxiv.2108.07258.
- Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., and Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017*

- ACM SIGSAC Conference on Computer and Communications Security, CCS '17, page 1175–1191, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3133956.3133982.
- Braun, V., Clarke, V., Boulton, E., Davey, L., and McEvoy, C. (2021). The online survey as a qualitative research tool. *International journal of social research methodology*, 24(6):641–654. DOI: 10.1080/13645579.2020.1805550.
- Chen, C., Feng, X., Li, Y., Lyu, L., Zhou, J., Zheng, X., and Yin, J. (2024). Integration of large language models and federated learning. *Patterns*, 5(12). DOI: 10.1016/j.pat-ter.2024.101098.
- Chen, Y., Li, R., Shao, J., Sun, L., Zhang, C., Zhang, J., and Li, X. (2026). Federated lora fine-tuning of llms with only transmitting matrix a or b. *IEEE Transactions on Mobile Computing*. DOI: 10.1109/tmc.2026.3696859.
- Cheng, Y., Zhang, W., Zhang, Z., Zhang, C., Wang, S., and Mao, S. (2024). Towards federated large language models: Motivations, methods, and future directions. *IEEE Communications Surveys & Tutorials*. DOI: 10.1109/COMST.2024.3503680.
- DeepSeek-AI, Liu, A., Feng, B., Xue, B., Wang, B., Wu, B., Lu, C., Zhao, C., Deng, C., Zhang, C., Ruan, C., Dai, D., Guo, D., Yang, D., Chen, D., Ji, D., Li, E., Lin, F., Dai, F., Luo, F., Hao, G., Chen, G., Li, G., Zhang, H., Bao, H., Xu, H., Wang, H., Zhang, H., Ding, H., Xin, H., Gao, H., Li, H., Qu, H., Cai, J. L., Liang, J., Guo, J., Ni, J., Li, J., Wang, J., Chen, J., Chen, J., Yuan, J., Qiu, J., Li, J., Song, J., Dong, K., Hu, K., Gao, K., Guan, K., Huang, K., Yu, K., Wang, L., Zhang, L., Xu, L., Xia, L., Zhao, L., Wang, L., Zhang, L., Li, M., Wang, M., Zhang, M., Zhang, M., Tang, M., Li, M., Tian, N., Huang, P., Wang, P., Zhang, P., Wang, Q., Zhu, Q., Chen, Q., Du, Q., Chen, R. J., Jin, R. L., Ge, R., Zhang, R., Pan, R., Wang, R., Xu, R., Zhang, R., Chen, R., Li, S. S., Lu, S., Zhou, S., Chen, S., Wu, S., Ye, S., Ye, S., Ma, S., Wang, S., Zhou, S., Yu, S., Zhou, S., Pan, S., Wang, T., Yun, T., Pei, T., Sun, T., Xiao, W. L., Zeng, W., Zhao, W., An, W., Liu, W., Liang, W., Gao, W., Yu, W., Zhang, W., Li, X. Q., Jin, X., Wang, X., Bi, X., Liu, X., Wang, X., Shen, X., Chen, X., Zhang, X., Chen, X., Nie, X., Sun, X., Wang, X., Cheng, X., Liu, X., Xie, X., Liu, X., Yu, X., Song, X., Shan, X., Zhou, X., Yang, X., Li, X., Su, X., Lin, X., Li, Y. K., Wang, Y. Q., Wei, Y. X., Zhu, Y. X., Zhang, Y., Xu, Y., Xu, Y., Huang, Y., Li, Y., Zhao, Y., Sun, Y., Li, Y., Wang, Y., Yu, Y., Zheng, Y., Zhang, Y., Shi, Y., Xiong, Y., He, Y., Tang, Y., Piao, Y., Wang, Y., Tan, Y., Ma, Y., Liu, Y., Guo, Y., Wu, Y., Ou, Y., Zhu, Y., Wang, Y., Gong, Y., Zou, Y., He, Y., Zha, Y., Xiong, Y., Ma, Y., Yan, Y., Luo, Y., You, Y., Liu, Y., Zhou, Y., Wu, Z. F., Ren, Z. Z., Ren, Z., Sha, Z., Fu, Z., Xu, Z., Huang, Z., Zhang, Z., Xie, Z., Zhang, Z., Hao, Z., Gou, Z., Ma, Z., Yan, Z., Shao, Z., Xu, Z., Wu, Z., Zhang, Z., Li, Z., Gu, Z., Zhu, Z., Liu, Z., Li, Z., Xie, Z., Song, Z., Gao, Z., and Pan, Z. (2025). Deepseek-v3 technical report. DOI: 10.48550/arxiv.2412.19437.
- Djuhera, A., Andrei, V. C., Li, X., Mönich, U. J., Boche, H., and Saad, W. (2025). R-sflm: Jamming resilient framework for split federated learning with large language models. *IEEE Transactions on Information Forensics and Security*. DOI: 10.1109/tifs.2025.3594107.
- Fan, T., Gu, H., Cao, X., Chan, C. S., Chen, Q., Chen, Y., Feng, Y., Gu, Y., Geng, J., Luo, B., et al. (2025). Ten challenging problems in federated foundation models. *IEEE Transactions on Knowledge and Data Engineering*. DOI: 10.1109/tkde.2025.3555328.
- Gholamiangonabadi, D. and Grolinger, K. (2024). Federated learning for sentiment analysis in presence of non-iid data: Sensitivity of deep learning models. *IEEE Access*, 12:128049–128060. DOI: 10.1109/ACCESS.2024.3453068.
- Ghosh, A., Chung, J., Yin, D., and Ramchandran, K. (2020). An efficient framework for clustered federated learning. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 19586–19597. Curran Associates, Inc.. DOI: 10.1109/tit.2022.3192506.
- Han, C., Yang, T., Cui, Z., and Sun, X. (2025). A privacy-preserving and trustworthy inference framework for llm-iiot integration via hierarchical federated collaborative computing. *IEEE Internet of Things Journal*. DOI: 10.1109/jiot.2025.3583764.
- Han, G., Zhang, T., Zhang, Y., Xu, G., Sun, J., and Cao, J. (2022). Verifiable and privacy preserving federated learning without fully trusted centers. *Journal of Ambient Intelligence and Humanized Computing*, 13(3):1431–1441. DOI: 10.1007/s12652-020-02664-x.
- Hasan, M. K., Kabir, S. R., Islam, S., Abdullah, S., Abbas, H. S., Pandey, B., and Gadekallu, T. R. (2025). Aes cryptography enabled responsible federated foundation model using transformer llm and lstm for smart grid iiot networks. *IEEE Internet of Things Journal*. DOI: 10.1109/jiot.2025.3608807.
- Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., and Welch, V. A. (2019). *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley-Blackwell, 2nd edition. Book.
- Hilmkil, A., Callh, S., Barbieri, M., Sütfeld, L. R., Zec, E. L., and Mogren, O. (2021). Scaling federated learning for fine-tuning of large language models. In *International Conference on Applications of Natural Language to Information Systems*, pages 15–23. Springer. DOI: 10.1007/978-3-030-80599-9_2.
- Hong, M., Zhang, K., Zhang, S., and He, Z. (2023). Fednpc: A federated learning framework for large language models in game npcs. In *2023 IEEE 21st Student Conference on Research and Development (SCORED)*, pages 363–368. DOI: 10.1109/SCORED60679.2023.10563705.
- Hu, J., Wang, D., Wang, Z., Pang, X., Xu, H., Ren, J., and Ren, K. (2024). Federated large language model: Solutions, challenges and future directions. *IEEE Wireless Communications*. DOI: 10.1109/mwc.009.2400244.
- Iftikhar, S. and Davy, S. (2024). Reducing carbon footprint in ai: A framework for sustainable training of large language models. In Arai, K., editor, *Proceedings of the Future Technologies Conference (FTC) 2024, Volume 1*, pages 325–336, Cham. Springer Nature Switzerland. DOI: 10.1007/978-3-031-73110-5_2.
- Jiang, J., Jiang, H., Ma, Y., Liu, X., and Fan, C. (2024). Low-

- parameter federated learning with large language models. In Jin, C., Yang, S., Shang, X., Wang, H., and Zhang, Y., editors, *Web Information Systems and Applications*, pages 319–330, Singapore. Springer Nature Singapore. DOI: 10.1007/978-981-97-7707-5_28.
- Jiang, Y., Li, Z., and Song, B. (2025). Fine-tuning large language models in federated learning with fairness-aware prompt selection. *Neural Networks*, page 108160. DOI: 10.1016/j.neunet.2025.108160.
- Karimireddy, S. P., Kale, S., Mohri, M., Reddi, S., Stich, S., and Suresh, A. T. (2020). SCAFFOLD: Stochastic controlled averaging for federated learning. In III, H. D. and Singh, A., editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 5132–5143. PMLR. DOI: 10.48550/arxiv.1910.06378.
- Kim, G., Yoo, J., and Kang, S. (2023). Efficient federated learning with pre-trained large language model using several adapter mechanisms. *Mathematics*, 11(21). DOI: 10.3390/math11214479.
- Kuang, W., Qian, B., Li, Z., Chen, D., Gao, D., Pan, X., Xie, Y., Li, Y., Ding, B., and Zhou, J. (2024). 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, KDD '24*, page 5260–5271, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3637528.3671573.
- Kumar, J. and Chimalakonda, S. (2024). Code summarization without direct access to code - towards exploring federated llms for software engineering. In *Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering, EASE '24*, page 100–109, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3661167.3661210.
- Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., and Smith, V. (2020). Federated optimization in heterogeneous networks. In Dhillon, I., Papailiopoulos, D., and Sze, V., editors, *Proceedings of Machine Learning and Systems*, volume 2, pages 429–450. DOI: 10.48550/arxiv.1812.06127.
- Li, Y., Sun, J., Liu, Y., Zhang, Y., Li, A., Chen, B., Roth, H. R., Xu, D., Chen, T., and Chen, Y. (2024). Federated black-box prompt tuning system for large language models on the edge. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking, ACM MobiCom '24*, page 1775–1777, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3636534.3698856.
- Liu, X., Pang, T., and Fan, C. (2023). Federated prompting and chain-of-thought reasoning for improving llms answering. In Jin, Z., Jiang, Y., Buchmann, R. A., Bi, Y., Ghiran, A.-M., and Ma, W., editors, *Knowledge Science, Engineering and Management*, pages 3–11, Cham. Springer Nature Switzerland. DOI: 10.1007/978-3-031-40292-0_1.
- Luo, W., Keung, J. W., Yang, B., Ye, H., Goues, C. L., Bissyande, T. F., Tian, H., and Le, B. (2026). When fine-tuning llms meets data privacy: An empirical study of federated learning in llm-based program repair. *ACM Transactions on Software Engineering and Methodology*, 35(3):1–46. DOI: 10.1145/3733599.
- Mawela, C., Issaid, C. B., and Bennis, M. (2025). A web-based solution for federated learning with llm-based automation. *IEEE Internet of Things Journal*. DOI: 10.1109/jiot.2025.3542897.
- McMahan, B., Moore, E., Ramage, D., Hampson, S., and Arcas, B. A. y. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 1273–1282. PMLR. DOI: 10.48550/arxiv.1602.05629.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Group, P., et al. (2010). Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *International journal of surgery*, 8(5):336–341. DOI: 10.7326/0003-4819-151-4-200908180-00135.
- Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., and Srivastava, G. (2021). A survey on security and privacy of federated learning. *Future Generation Computer Systems*, 115:619–640. DOI: 10.1016/j.future.2020.10.007.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Glasziou, P. P., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., McGowan, J., Miller, C. M., Mockford, K., and Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372:n71. DOI: 10.1136/bmj.n71.
- Pais, V., Rao, S., and Muniyal, B. (2025). Strategies for reducing the communication and computation costs in cross-silo federated learning: A comprehensive review. *IEEE Access*, 13:93385–93416. DOI: 10.1109/ACCESS.2025.3573933.
- Petticrew, M. and Roberts, H. (2006). *Systematic Reviews in the Social Sciences: A Practical Guide*. Blackwell Publishing.
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67. DOI: 10.48550/arxiv.1910.10683.
- Sadot, A. A. I. M., Maliha Mehjabin, M., and Mahafuz, A. (2023). A novel approach to efficient multilabel text classification: Bert-federated learning fusion. In *2023 26th International Conference on Computer and Information Technology (ICCIT)*, pages 1–6. DOI: 10.1109/ICCIT60459.2023.10441264.
- Saha, S., Hota, A., Chattopadhyay, A. K., Nag, A., and Nandi, S. (2024). A multifaceted survey on privacy preservation of federated learning: progress, challenges, and opportunities. *Artificial Intelligence Review*, 57(7):184. DOI: 10.1007/s10462-024-10766-7.
- Seo, J., Zhang, N., and Rong, C. (2023). Flexible and secure code deployment in federated learning using large language models: Prompt engineering to enhance ma-

- icious code detection. In *2023 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, pages 341–349. DOI: 10.1109/CloudCom59040.2023.00062.
- So, J., He, C., Yang, C.-S., Li, S., Yu, Q., E Ali, R., Guler, B., and Avestimehr, S. (2022). Lightsecagg: a lightweight and versatile design for secure aggregation in federated learning. *Proceedings of Machine Learning and Systems*, 4:694–720. DOI: 10.48550/arxiv.2109.14236.
- Solat, F. and Lee, J. (2026). Energy-and aoi-aware hierarchical personalized federated learning for distributed edge systems with selective llm module exchange. *IEEE Internet of Things Journal*. DOI: 10.1109/jiot.2026.3695103.
- Su, H., Xiang, C., and Ramesh, B. (2024). Towards confidential chatbot conversations: A decentralised federated learning framework. *The Journal of The British Blockchain Association*. DOI: 10.31585/jbba-7-1-(3)2024.
- Sujaykumar Reddy, M., Karnati, H., and Mohana Sundari, L. (2024). Transformer-based federated learning models for recommendation systems. *IEEE Access*, 12:109596–109607. DOI: 10.1109/ACCESS.2024.3439668.
- Sun, H., Tian, H., Ni, W., Zheng, J., Niyato, D., and Zhang, P. (2025). Federated low-rank adaptation for large models fine-tuning over wireless networks. *IEEE Transactions on Wireless Communications*, 24(1):659–675. DOI: 10.1109/TWC.2024.3497998.
- Sun, W., Wang, X., Liang, Z., Chen, J., Lan, W., Chen, Y., and Wang, F. (2026a). Fedshieldllm: Measurement, detection and protection of privacy leakage in federated llms. *IEEE Transactions on Mobile Computing*. DOI: 10.1109/tmc.2026.3657229.
- Sun, W., Wang, X., Liang, Z., Gong, T., Lan, W., Chen, Y., Li, H., and Wang, F. (2026b). Efficient and privacy-preserving federated knowledge learning for distributed llm. *IEEE Transactions on Mobile Computing*. DOI: 10.1109/tmc.2026.3674545.
- T. Dinh, C., Tran, N., and Nguyen, J. (2020). Personalized federated learning with moreau envelopes. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 21394–21405. Curran Associates, Inc.. DOI: 10.48550/arxiv.2006.08848.
- Tan, C., Wen, R., Li, R., Zhao, Z., Hossain, E., and Zhang, H. (2025). Tool-aided evolutionary llm for generative policy toward efficient resource management in wireless federated learning. *IEEE Journal on Selected Areas in Communications*. DOI: 10.1109/jsac.2025.3645754.
- Tenney, I., Das, D., and Pavlick, E. (2019). BERT rediscovers the classical NLP pipeline. In Korhonen, A., Traum, D., and Màrquez, L., editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4593–4601, Florence, Italy. Association for Computational Linguistics. DOI: 10.18653/v1/P19-1452.
- Thapa, C., Tang, J. W., Abuadbbba, A., Gao, Y., Camtepe, S., Nepal, S., Almashor, M., and Zheng, Y. (2023). Evaluation of federated learning in phishing email detection. *Sensors*, 23(9):4346. DOI: 10.3390/s23094346.
- Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C. C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., Fuller, B., Gao, C., Goswami, V., Goyal, N., Hartshorn, A., Hosseini, S., Hou, R., Inan, H., Kardas, M., Kerkez, V., Khabsa, M., Kloumann, I., Korenev, A., Koura, P. S., Lachaux, M.-A., Lavril, T., Lee, J., Liskovich, D., Lu, Y., Mao, Y., Martinet, X., Mihaylov, T., Mishra, P., Molybog, I., Nie, Y., Poulton, A., Reizenstein, J., Rungta, R., Saladi, K., Schelten, A., Silva, R., Smith, E. M., Subramanian, R., Tan, X. E., Tang, B., Taylor, R., Williams, A., Kuan, J. X., Xu, P., Yan, Z., Zarov, I., Zhang, Y., Fan, A., Kambadur, M., Narang, S., Rodriguez, A., Stojnic, R., Edunov, S., and Scialom, T. (2023). Llama 2: Open foundation and fine-tuned chat models. Available at: <https://arxiv.org/abs/2307.09288>.
- Truex, S., Liu, L., Chow, K.-H., Gursoy, M. E., and Wei, W. (2020). Ldp-fed: Federated learning with local differential privacy. In *Proceedings of the third ACM international workshop on edge systems, analytics and networking*, pages 61–66. DOI: 10.1145/3378679.3394533.
- Vasconcelos, A. B., Drummond, L. M. d. A., Brum, R. C., and Paes, A. (2023). Exploring federated learning to trace depression in social media with language models. In *2023 International Symposium on Computer Architecture and High Performance Computing Workshops (SBAC-PADW)*, pages 24–30. DOI: 10.1109/SBAC-PADW60351.2023.00014.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc. Available at: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
- Wang, D. (2024). Federated learning using gpt-4 boosted particle swarm optimization for compact neural architecture search. *Journal of Advances in Information Technology*, 15(9). DOI: 10.12720/jait.15.9.1011-1018.
- Wei, K., Li, J., Ding, M., Ma, C., Yang, H. H., Farokhi, F., Jin, S., Quek, T. Q. S., and Vincent Poor, H. (2020). Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Transactions on Information Forensics and Security*, 15:3454–3469. DOI: 10.1109/TIFS.2020.2988575.
- Wu, F., Li, Z., Li, Y., Ding, B., and Gao, J. (2024a). Fedbiot: Llm local fine-tuning in federated learning without full model. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '24*, page 3345–3355, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3637528.3671897.
- Wu, J., Chen, L., Fang, S., and Wu, C. (2024b). An application programming interface (api) sensitive data identification method based on the federated large language model. *Applied Sciences*, 14(22). DOI: 10.3390/app142210162.

- Xie, S., Wen, D., You, C., Chen, Q., Bennis, M., and Huang, K. (2026). Fedldrop: Federated lora with dropout for generalized llm fine-tuning. *IEEE Journal on Selected Areas in Communications*. DOI: 10.1109/jsac.2026.3660935.
- Xu, M., Cai, D., Wu, Y., Li, X., and Wang, S. (2023). Fwdllm: Efficient fedllm using forward gradient. *arXiv preprint arXiv:2308.13894*. DOI: 10.48550/arxiv.2308.13894.
- Yan, N., Su, Y., Deng, Y., and Schober, R. (2025). Federated fine-tuning of llms: Framework comparison and research directions. *IEEE Communications Magazine*, 63(10):52–58. DOI: 10.1109/mcom.001.2400770.
- Yao, Y., Zhang, J., Wu, J., Huang, C., Xia, Y., Yu, T., Zhang, R., Kim, S., Rossi, R., Li, A., et al. (2024). Federated large language models: Current progress and future directions. *arXiv preprint arXiv:2409.15723*. DOI: 10.1007/978-981-92-1468-6_12.
- Ye, R., Wang, W., Chai, J., Li, D., Li, Z., Xu, Y., Du, Y., Wang, Y., and Chen, S. (2024). Open-fedllm: 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*, pages 6137–6147. DOI: 10.1145/3637528.3671582.
- Yi, X., Hu, C., Cai, B., Huang, H., Chen, Y., and Wang, K. (2025). Fedalora: Adaptive local lora aggregation for personalized federated learning in llm. *IEEE Internet of Things Journal*. DOI: 10.1109/jiot.2025.3582427.
- Yun, S., Bhuiyan, Z. A., Sadi, M. T. A. H., and Su, S. (2023). Privacy-preserving federated learning through clustered sampling on fine-tuning distributed non-iid large language models. In *2023 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCLOUD/SocialCom/SustainCom)*, pages 531–538. DOI: 10.1109/ISPA-BDCLOUD-SocialCom-SustainCom59178.2023.00101.
- Zhang, L., Ao, X., Guo, S., and Qiao, D. (2025a). Fedfa: Efficient federated large language models with feature adapters. *Expert Systems with Applications*, page 130894. DOI: 10.1016/j.eswa.2025.130894.
- Zhang, T., Yu, H., Chen, Y., Wang, S., and Yang, Z. (2025b). Flaguard: Efficient verifiable federated lora of large language models. *IEEE Transactions on Mobile Computing*. DOI: 10.1109/tmc.2025.3641570.
- Zhang, T., Yu, H., Yang, Z., Chen, Y., and Yu, S. (2025c). Lavfi: Efficient verifiable federated learning for large language models. *IEEE Transactions on Dependable and Secure Computing*. DOI: 10.1109/tdsc.2025.3581728.
- Zhao, K., Yang, Z., Huang, C., Chen, X., and Zhang, Z. (2024). Fedslm: federated split learning for large language models over communication networks. In *2024 International Conference on Ubiquitous Communication (Ucom)*, pages 438–443. IEEE. DOI: 10.1109/ucom62433.2024.10695888.
- Zuo, X., Wang, M., Zhu, T., Zhang, L., Ye, D., Yu, S., and Zhou, W. (2026). Federated trustchain: Blockchain-enhanced llm training and unlearning. *IEEE Trans-*
- actions on Dependable and Secure Computing*. DOI: 10.1109/tdsc.2026.3665277.