Leveraging Federated Learning for Secure Transfer and Deployment of ML Models in Healthcare

Zlate Dodevski
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Skopje, Republic of N. Macedonia
zlate.dodevski@students.finki.ukim.mk

Tanja Pavleska
Laboratory for Open Systems and Networks
Jozef Stefan Institute
Ljubljana, Slovenia
atanja@e5.ijs.si

Vladimir Trajkovikj
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Skopje, Republic of N. Macedonia
vladimir.trajkovikj@finki.ukim.mk

***Abstract*—Federated learning (FL) represents a pivotal advancement in applying Machine Learning (ML) in healthcare. It addresses the challenges of data privacy and security by facilitating model transferability across institutions. This paper explores the effective employment of FL to enhance the deployment of large language models (LLMs) in healthcare settings while maintaining stringent privacy standards.**

Along a detailed examination of the challenges in applying LLMs to the healthcare domain, including privacy, security, regulatory constraints, and training data quality, we present a federated learning architecture tailored for LLMs in healthcare. This architecture outlines the roles and responsibilities of participating entities, providing a framework for secure collaboration. We further analyze privacy-preserving techniques such as differential privacy and secure aggregation in the context of federated LLMs for healthcare, offering insights into their practical implementation.

Our findings suggest that federated learning is a viable choice for enhancing he capabilities of LLMs in healthcare while preserving patient privacy. In addition, we also identify persistent challenges in areas such as computational and communicational efficiency, lack of benchmarks and tailored FL aggregation algorithms applied to LLMs, model performance, and ethical concerns in participant selection. By critically evaluating the proposed approach and highlighting its potential benefits and limitations in real-world healthcare settings, this work provides a foundation for future research in secure and privacy-preserving ML deployment in healthcare.

Keywords—Federated Learning, Large Language Models, Data Privacy, Healthcare ML, Privacy Preservation, Model Transferability

I. Introduction

The advancements in hardware and software technologies, hyper-connectivity, and the fourth industrial revolution lead to the creation of mass amounts of health-related data. Machine learning and AI, in general, are the biggest winners from this endless pool of structured and unstructured data, as these technologies thrive on large datasets to identify patterns and make predictions. [1]. The novel adaptable, predictive approach to generating insights, decision support, and assistance in tasks that have long been considered solely reserved for human expertise is based on this paradigm's capabilities to recognize patterns from the data without being explicitly programmed to do so.

Conventional machine learning implies that the data owner communicates with a specific central server with significant computational power. The central server consumes data from different sources and applies training techniques and algorithms to devise an effective model. ML requires large amounts of data to satisfy the expectations for the model's performance.

Large-language models (LLMs), as representative of ML advancements, have been a particular point of interest in recent years. They have already proven their applicability and massive potential in multiple fields [1]. LLMs are designed to understand, generate, and interact with human-like text and can understand context, making them suitable for performing a wide range of complex language-related tasks. They are trained in two main phases. First, the model learns general knowledge about language patterns in the pre-training phase. Then, it can be fine-tuned to execute downstream tasks to specialize its expertise in a specific domain [2].

However, like other machine learning models researched and implemented, this paradigm is data-hungry, meaning that it inherently requires massive training data to achieve the expected performance [3]. Thus, LLMs are designed to perform better with an increase in training data volume and computational power.

Various unfortunate scenarios related to the misuse of private and personal data cast a shadow on AI's capabilities, underscoring the growing concerns about data privacy, specifically in the phases when the models are trained [4], [5]. The year 2016 is particularly significant for two key developments aiming to overcome these challenges. The first is the attempt to regulate personal data collection, processing, and storing by introducing the General Data Protection Regulation (GDPR) in Europe [6]. The second key development was the introduction of Federated Learning (FL) by Google researchers, which provided a groundbreaking scientific approach to addressing data privacy and security concerns in ML [7]. Their approach introduced a collaborative technique for training global machine-learning models without exposing or sharing sensitive data.

This paper aims to discuss the possibility of satisfying the needs of both data owners and ML experts by leveraging the concept of federated learning. On the one hand, data owners can be supported to participate in collaborative training in a privacy-preserving manner when their data is insufficient to craft a high-performance model, such as LLM. On the other hand, ML experts can develop and advance their approaches by utilizing large volumes of real-life institutional data and access to diverse scenarios, which are essential for building a robust model.

Being aimed at investigating the FL potential for application in the health domain, the insights presented in this

work offer support in finding a more robust, secure, and effective use of AI that does not require technical proficiency of the medical experts, ultimately contributing to improved patient care and data protection.

The remainder of this paper is organized as follows: Sections II and III introduce FL and LLMs, and their relevance to healthcare; section IV presents our proposed FL architecture for LLMs in healthcare, detailing the system components and their roles; section V discusses challenges in implementing LLMs in healthcare using FL; and section VI explores privacy-preserving techniques for integration with our FL architecture.

II. LLM-BASED HEALTHCARE APPLICATIONS

LLMs are usually trained on high-quality public data, but their performance is often limited when tasked with specialized or narrower-spectrum tasks. With specific expertise in mind, such as healthcare, different approaches should be considered to build that corpus. Healthcare institutions can use their local datasets, leading to less efficient LLM, or they can join other collaborative efforts to make high-quality training data facing the inevitable challenges of privacy and regulations.

LLMs are attractive in the healthcare area because of their capability to simplify the interaction with an intelligent system without needing technical expertise. Lack of technological proficiency of medical experts can decrease the adoption of a specific software solution and can be marked as overly complex. The core principles of the LLMs allow users to interact with their domain rules, persistent knowledge, and past experiences without the need to rely on their computer literacy. One of the enormous benefits of utilizing LLMs in medical workflows is due to the nature of the output, which is in an understandable form of natural language. The ease of use of natural language to provide instructions and ask for decision support bridges the gap between the domain experts and the utilization of an intelligent computer system. In the past several years, we have already witnessed the potential of LLMs in healthcare in many areas, such as interpreting images from a specific medical domain, summarizing reports and medical history, identifying patterns in electronic health records (EHR), and offering support for decision-making processes. The use of natural language can also influence patient engagement processes [16].

Furthermore, much of the data that persists within healthcare institutions is in unstructured formats, such as clinical notes, conversations, diagnoses, prescriptions, and research articles. LLMs are particularly effective at processing and using these natural language texts. In that way, the transparency is increased, and the expert can examine the reasoning behind the provided answers in a straightforward manner. Even though efforts to provide suitable LLMs for healthcare are already in place and the community is verifying the theoretical and conceptual findings, the decision of one healthcare institution to utilize such a system can face a lot of hesitation. Most of the training data for the LLMs comes from publicly available sources, lacking the nuances that bring the real-life data isolated in the infrastructure of a single healthcare institution. Additionally, each institution has many characteristics that make it unique in how it works. Workflows, dictionaries, specific characteristics of the population it serves, or particular domain attributes can result in difficulties for off-the-shelf LLMs in providing the correct output to the instruction given. This calls for the institutions to further tailor and tune the capabilities of the LLM. Although state-of-the-art LLMs allow for such modifications and fine-tunings and making this process feasible, this comes with a heavy involvement and effort by the institution representatives and with extensive computational resources. Finally, even if one institution is capable of making efforts to adapt a generalized LLM for its use, it faces the inevitable obstacle of data insufficiency. In general, a single institution either cannot provide enough data to receive proper, correct output for the downstream task or is incapable of solving instruction of so-called new events or conditions.

Clearly, healthcare institutions need support in multiple areas to make the process easier to follow and adopt. As a result, processes related to finding a suitable LLM model, maintaining it, and keeping it up to date should be outsourced to a separate body owning the expertise. To effectively adapt LLMs in the healthcare domain, every proposed solution must address and guarantee the resolution of the challenges discussed in section II. Therefore, collaboration among institutions in compliance with the industry regulations should be established to build a rich training corpus.

III. OVERVIEW OF FEDERATED LEARNING PRINCIPLES AND THEIR RELEVANCE TO HEALTHCARE DATA PRIVACY AND SECURITY

A. Overview of Federated Learning Principles

In healthcare, data is often distributed across multiple institutions, each possessing unique and valuable patient information. Traditional approaches to AI model training require centralizing this data, which poses significant privacy and security risks. Federated learning provides a solution by enabling collaborative model training without exchanging raw data. Instead, each institution trains the model locally and shares only aggregated updates with a central server. This method ensures that sensitive patient data remains within the institution, facilitating the transfer and deployment of AI models across different settings without compromising data security. FL is an iterative process, and each communication round aims to improve the model's performance. A typical FL scenario consists of two main phases in each round: local training of the model done on the participant side and aggregation of updates, which aims to create the most accurate consensus model.

There are three main types of FL based on how the data is distributed across participants. In horizontal federated learning, the datasets share the same feature space but differ in the samples they contain. Vertical federated learning, on the other hand, involves datasets with the same samples but different features. Lastly, federated transfer learning encompasses datasets that vary in both their feature and sample spaces [8].

B. Relevance to healthcare data privacy and security

In the context of machine learning (ML) applications involving healthcare data, there are three critical vulnerability points that require attention: the data itself, the training of ML models, and the communication and transfer of data. Each area carries specific challenges and risks that must be mitigated to ensure the privacy, security, and efficacy of ML systems in healthcare.

Health-related data is inherently complex, with characteristics such as high dimensionality, variance over time, heterogeneity, difficult interoperability, sparsity, and isolation [9]. Protecting the privacy of patients' personal and sensitive health information is crucial. Due to the sensitive nature of healthcare data, security breaches can lead to severe consequences, including identity theft, fraud, and violation of patient confidentiality.

Healthcare data often comes from various sources, such as hospitals, clinics, wearable devices, and electronic health records (EHRs). This data is typically non-independent, identically distributed (non-iid), unbalanced, and fragmented across different systems. Additionally, data may be sparse or isolated, making it challenging to build comprehensive patient profiles or conduct large-scale analyses.

Federated learning offers a promising approach to overcoming these challenges by allowing ML models to be trained across multiple decentralised data sources while keeping data local. This technique improves data privacy and security by not requiring raw data to be transferred to a central location. In an FL environment, each data controller defines its governance processes and privacy policies. This includes setting conditions for data access, training, and validation phases [10], [11], [12], [13] .

Communication between institutions, especially when dealing with healthcare data, must adhere to strict regulatory requirements, such as the Health Insurance Portability and Accountability Act (HIPAA) in the U.S. and GDPR in Europe. Thus, privacy-preserving mechanisms should be implemented "by design" and "by default" to ensure that sensitive patient information is processed securely. A compliant ML system requires secure data transfer mechanisms, consent management, and audit trails. The FL setting can offer advancement in this area by letting institutions keep sensitive information, prevent unnecessary data transfers and processing that could violate regulatory requirements, and minimize the risk of data breaches [14].

Training ML models with healthcare data presents unique challenges, including addressing data bias, limited sample sizes, and ensuring model performance. Healthcare data may be biased due to demographic imbalances, socio-economic factors, or varying levels of care access across populations.

Training ML models with diverse datasets enhances their generalizability and robustness. By incorporating data from various sources and populations, models can better adapt to new and unforeseen health events, improving their predictive power and reliability. Federated learning, in particular, enables the use of diverse datasets while maintaining privacy, thus improving overall model performance [15] [16].

The following section introduces a FL-based architecture devised to address the challenges outlined above, detailing the role of each of its components in that process.

IV. Presentation of a Typical FL Architecture that Can be Deployed for LLM-Based Healthcare Applications

Figure 1 depicts the three major components of a typical FL architecture. The participants involved in our cross-silo FL setting are the healthcare institutions, the manager (e.g., aggregation server or global server), and the communication-computation layer, which aggregates local updates and orchestrates communication phases in the ecosystem. Each component has its own responsibilities, which are essential for the model to satisfy the preset expectations.

Leveraging FL in utilizing LLMs adds a layer of complexity and implies different approaches based on the level of decentralization that needs to be achieved [17]. FL can help in both the pre-training and fine-tuning phases of LLM, and it is up to the requirements' specific characteristics and the parties involved computational power to choose the right strategy [18], [19]. We will cover the different approaches while examining the three major architecture components.

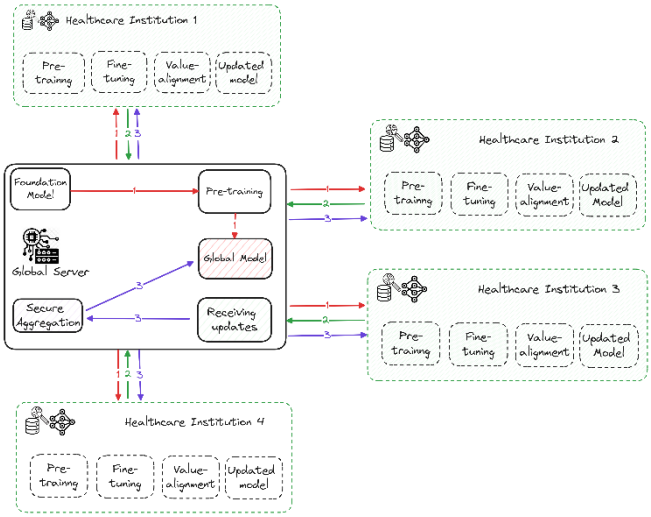

Figure 1 Typical FL Architecture in LLM. The global server is responsible for the aggregation of local updates and orchestration of communication phases, and the participants in the form of healthcare institutions are responsible for training the LLM

A. Global server

The global server plays a central role, and due to the lack of properly established taxonomy and standards, this component in the literature is also considered as the manager or creator of the whole ecosystem. It is often referred to as the "manager" or "aggregation server" tasked with overseeing the entire collaboration and ensuring its smooth functioning. In healthcare, the manager can be a single healthcare institution that holds a lot of data and wants to leverage the FL setting to collaborate with other institutions, either to complete the missing domains and dimensions (by utilizing Vertical FL) or to enrich and expand the feature set in the same dimension (Horizontal FL). The global server's responsibilities can be broken down into several distinct areas: strategy for choosing a foundation model, strategy for exploiting data distribution and the client selection.

The foundation model, also called the base model, forms the initial point of the LLM training in the FL setting. It represents a starting point for institutions to leverage the pre-training process with their own data or fine-tune it to perform specific downstream tasks since foundation models are usually trained on publicly available datasets. The pre-training process is computationally and time-consuming, meaning that the global server must find the most suitable scenario for satisfying requirements.

There are two main kinds of pre-training models: one based on the BERT model and the other on the GPT model. Both perform differently for different tasks and scenarios

[20], [21]. Many attempts are made to use publicly available literature specific to the medical domain and create ready LLM models for usage, such as BioBert, PubMedBert, and ClinicalBert, which show superior performance than general pre-trained models [22], [23].

One approach is to select a suitable foundation model based on the options examined before or to initialize FL pre-training, where each party will contribute to the pre-training of the foundation LLM model. The latter approach requires the institutions to have expertise and sufficient computational power in their infrastructure to complete the assignment. Another consideration is the model's size and complexity, which will influence the following steps if not chosen according to the participant's IT infrastructure.

Federated learning can work differently depending on how data is spread and distributed across entities. In horizontal FL, each institution has data with similar features (for example, multiple hospitals with similar patient data). In vertical FL, institutions have different features for the same set of patients (for example, one entity has clinical data, and another has genetic data). Transfer learning can also be used when the model needs to generalize across different datasets [24]. The global server is responsible for choosing the appropriate strategy based on the data distribution and the desired outcomes.

Client selection in FL refers to choosing which institutions to participate in each training round. The global server must ensure that diverse institutions contribute to model updates without overloading the communication system. Institutions with more data or better computational resources might participate more frequently. Still, the system should be flexible enough to rotate clients or dynamically adjust client participation based on resource availability.

B. Participants

The entities participating in this collaboration technique are also tasked with significant responsibilities. In the medical domain, these institutions own huge structured or unstructured datasets and are willing to participate in a distributed training process. Their responsibilities can be broken down into the pre-training process, fine-tuning, value alignment, and strategy for local updates.

If the FL-specific training approach is adopted, as discussed previously in the strategy for choosing the foundation model, then each institution may pre-train the selected model on its data and ensure that the initial model updates sent to the global server are more relevant and valuable. This capability in an FL setting can depend significantly on the institution's computational power, and even though possible and theoretically feasible, it can require a lot of additional expertise for the healthcare institution to engage in this kind of activity.

Fine-tuning is crucial in adopting an LLM in the healthcare domain. The approach of LLM fine-tuning is to make downstream tasks required by the domain be instructed with human feedback [25]. Each institution should provide input-output pairs where instructions are explicitly offered to solve some already defined downstream tasks. These datasets are designed to give the model an idea of what kind of output is expected. The expectation is that the LLM will learn to generalize and can handle novel instructions even though they were not a part of the fine-tuning instruction dataset.

The variety of downstream tasks that LLMs can perform in the healthcare domain is often the critical reason institutions engage with this concept. Based on a benchmark for generalist biomedical AI, some of the most frequently performed medical-relevant tasks suitable for the LLM domain are question answering, visual question answering (for example, based on radiology or pathology images), report summarization and generation, and medical image and medical documents classification [26]. Additionally, relation extraction in combination with named entity recognition can be added to the list of medical-relevant tasks. This is helpful in the medical domain to extract medical terms such as diseases, conditions, procedures, and symptoms from unstructured data and find suitable interpretations and connections in the unstructured data pool.

Thanks to the LLaMA, each institution can make significant attempts to build its domain-specific instruction set and contribute to global instruction tuning in the FL setting. With the FL paradigm, each downstream task can be trained on multiple datasets instead of a single dataset, giving more suitable responses and outputs [27].

In the FL setting, the value-alignment step occurs on the participant's side during local training. Its purpose is to ensure that the model's objectives are aligned with each institution's values and goals. This step is particularly crucial in the medical field, where ethical guidelines and patient care standards are of the utmost importance.

Technically, value-alignment is solved similarly to instruction tuning, with each participant's preference dataset containing combinations of instruction, preferred, and misreferred responses.

In FL, participants typically have far fewer computational resources than centralized cloud servers and fine-tuning all parameters of LLMs can be an obstacle. Parameter-efficient tuning techniques, such as Lora, are used to address this limitation [28]. Instead of updating the entire pre-trained model weights to obtain local updates, participants modify only specific parameters and send them back to the global server for aggregation.

C. Communication-computation layer

As presented above, the global server is responsible for managing the whole ecosystem, and one of the most complex tasks is related to the communication-computation layer. The global server should manage the aggregation process of local model updates and ensure that the global LLM and updates are securely transmitted across the system.

Choosing the suitable FL algorithm for combining all findings and improvements made by each participant in the form of parameter weights is a step that has attracted many researchers and experts. One of the first and most used algorithms is Federated Averaging (FedAvg), but more sophisticated approaches may be necessary in different scenarios [7]. The model's performance relies significantly on how model updates are aggregated.

Even though the only data transmitted through the network in an FL setting are the model and its updates, the communication layer is responsible for ensuring that the transfer is done securely and continuously. The communication layer component must develop a strategy for creating a pipeline from a live data connection to the model and inference to transmitting new model parameters via

secure channels to the aggregating server. Size and complexity of the model must be considered as well, since they can cause a bottleneck.

In addition, the communication layer also ensures that the data transfer is seamless and uninterrupted. This component is tasked with developing a robust strategy to create an efficient pipeline, from managing real-time data connections to facilitating model utilization and transmitting updated model parameters securely to the central aggregation server. A key consideration for the communication layer is the size and complexity of the used model. Large models with huge parameter lists can introduce significant bottlenecks during transmission, especially when dealing with limited bandwidth or less powerful devices. As such, the communication layer must be adept at handling these challenges, ensuring that updates are transferred efficiently without compromising the speed or security of the system.

V. Challenges for LLM application in healthcare from a federated learning perspective

Implementing LLMs in healthcare using FL presents a set of intertwined challenges when viewed through the lenses of privacy and security. There is a foundational challenge between the need for diverse and high-quality data generated by institutions in the specific domain and the importance of protecting sensitive information. FL enables availability and access to a broader spectrum of data sources while maintaining privacy. Still, the inability to directly act upon raw data can impact the convergence of the model and model performance. Data transfer needs in FL, even though minimized to just model updates, still introduce a risk for security attacks. This risk increases with the communication overhead caused by distributing complex and large LLMs. By introducing a central figure in the architecture in the name of the global aggregation server, the FL setting in LLM opens up a single point of failure in the ecosystem. Adversarial attacks can be performed, compromising model integrity, which could lead to data breaches and incorrect outputs.

FL is still a young and immature topic in the context of LLM. One of the biggest challenges is the lack of benchmarks and comprehensive reviews that can examine the solution's success based on different tasks, architectures, the number of clients, network bandwidth, computational resources, etc. These reviews and benchmarks can further expose security and privacy-preserving issues and initiate proper risk mitigation strategies. Multiple algorithms exist in the literature for aggregating local updates, but no specific algorithm is proposed or adapted for LLMs.

The analysis of the three major components in the previous section pointed out the responsibilities, approaches, and strategies that need to be considered in order to collaboratively design and implement training, and utilize LLM properly. The analysis emphasized that training LLMs in a federated learning setting requires a thoughtful, tailored approach to address the unique challenges. Additionally, there are various approaches to take, depending on factors such as participant resources, data distribution, model size and complexity, data transfer, etc.

The client selection process, in which the ecosystem manager decides which participants should be involved, can raise many ethical concerns, such as fairness. The purpose of the collaboration is to make the LLM more robust. Still, some participants' data volume and computational power can squeeze out institutions that are not on that level but still can add to the diversity and offer unique cultural, ethical, and contextual values. While FL addresses many privacy concerns by design, it also introduces new security considerations that must be carefully managed. Successfully navigating these challenges requires a detailed approach that balances privacy protection, security enhancement, and the pursuit of practical and robust LLM in healthcare.

VI. Privacy-preserving techniques

The deployment of LLMs in the healthcare field through FL promises advancements in preparing models to react to given domain-specific downstream tasks. The FL can enhance LLMs' effectiveness and proper application while safeguarding patient confidentiality and ensuring regulatory compliance, providing medical professionals greater confidence in adopting these tools.

However, while FL enables collaborative learning without direct data sharing, it's not immune to privacy threats. With this approach, raw data remains local, but the model updates shared during training can still leak information. In addition, LLMs trained with healthcare data could memorize and potentially regenerate sensitive patient information. A privacy breach in this context can cause severe consequences, including exposure to medical history, compromising patient confidentiality, and misuse of sensitive health information [29].

During this collaborative process, the model or its updates could become targets for various attacks. For instance, model inversion attacks performed on the global model might allow the reconstruction of individual patient records. Similarly, membership inference attacks could reveal the presence of specific institutions or patient data in the training, potentially exposing the entire medical history. Malicious participants in the process could poison the model by introducing biases or backdoors, potentially leading to improper results generated by the LLMs [30], [31].

To counter these risks and threats, researchers and practitioners evaluate the effects of several privacy-preserving techniques, such as secure aggregation and differential privacy. Secure aggregation, a cryptographic protocol, allows the central server to observe aggregated results without accessing individual model updates. This approach maintains accuracy but adds significant communication costs. Differential privacy, on the other hand, adds calibrated noise to data or model parameters, offering statistical privacy guarantees. While effective against inference attacks, it may reduce model accuracy and require additional workload in the parameter-tuning process [32].

The choice of privacy-preserving techniques must be made with a thorough understanding of the specific use case, the sensitivity of the data involved, and the potential impacts of privacy breaches. The tailored approach should calibrate the trade-off between model performance and data protection. More robust privacy protection might require limiting the model’s access to much-needed data for LLMs to offer a proper answer to a specific task, degrading the model performance and increasing the computational and communicational overhead. Finding the right balance between privacy, system performance, and efficiency will be crucial for deploying LLMs in healthcare using FL.

VII. Conclusion

This paper explored the potential of federated learning (FL) in enhancing the deployment of large language models (LLMs) in healthcare settings. By enabling privacy-preserving collaboration, FL allows healthcare institutions to collectively train and improve LLMs without compromising sensitive patient data. This approach not only addresses fundamental privacy concerns but also enhances model performance by leveraging diverse datasets across institutions, potentially improving the generalizability and robustness of LLMs in healthcare applications. To facilitate the implementation of healthcare LLM with FL, we examined a tailored architectural framework that outlines the roles and responsibilities of participating entities. In addition, challenges and consideration of risks and threats were reviewed, especially in combination with privacy-preserving techniques.

Looking ahead, several areas require further research and development. Optimization of computational and communication efficiency for LLMs, development of standardized benchmarks, establishment of ethical frameworks for participant selection, and exploration of advanced privacy-preserving techniques are crucial for future work.

VIII. Acknowledgments

The research within this paper has been funded by the project "Bridging Research Institutions to Catalyze Generative AI Adoption by the Health Sector in the Widening Countries (ChatMED)" under the Horizon Europe - Widening Participation and Spreading Excellence program (Grant Agreement ID: 101159214), supported by the European Union.

References

[1] C. Zhou *et al.*, "A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT," May 01, 2023, *arXiv*: arXiv:2302.09419. doi: 10.48550/arXiv.2302.09419.

[2] J. Yang *et al.*, "Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond," *ACM Trans Knowl Discov Data*, vol. 18, no. 6, p. 160:1-160:32, Apr. 2024, doi: 10.1145/3649506.

[3] J. Kaplan *et al.*, "Scaling Laws for Neural Language Models," Jan. 22, 2020, *arXiv*: arXiv:2001.08361. doi: 10.48550/arXiv.2001.08361.

[4] G. Laurie, K. Jones, L. Stevens, and C. Dobbs, *A Review of Evidence Relating to Harm Resulting from Uses of Health and Biomedical Data*. 2015.

[5] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané, "Concrete Problems in AI Safety," arXiv.org. Accessed: Sep. 05, 2024. [Online]. Available: https://arxiv.org/abs/1606.06565v2

[6] C. Kuner, L. A. Bygrave, C. Docksey, and L. Drechsler, Eds., "The EU General Data Protection Regulation (GDPR): A Commentary," Oxford University PressNew York, Feb. 2020. doi: 10.1093/oso/9780198826491.001.0001.

[7] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, PMLR, Apr. 2017, pp. 1273–1282. Accessed: Sep. 01, 2024. [Online]. Available: https://proceedings.mlr.press/v54/mcmahan17a.html

[8] P. Kairouz *et al.*, "Advances and Open Problems in Federated Learning," *Found. Trends® Mach. Learn.*, vol. 14, no. 1–2, pp. 1–210, Jun. 2021, doi: 10.1561/2200000083.

[9] "Deep learning for healthcare: review, opportunities and challenges - PMC." Accessed: Sep. 12, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6455466/

[10] N. Rieke *et al.*, "The future of digital health with federated learning," *Npj Digit. Med.*, vol. 3, no. 1, pp. 1–7, Sep. 2020, doi: 10.1038/s41746-020-00323-1.

[11] B. Pfitzner, N. Steckhan, and B. Arnrich, "Federated Learning in a Medical Context: A Systematic Literature Review," *ACM Trans. Internet Technol.*, vol. 21, no. 2, pp. 1–31, Jun. 2021, doi: 10.1145/3412357.

[12] K. Dasaradharami Reddy and T. R. Gadekallu, "A Comprehensive Survey on Federated Learning Techniques for Healthcare Informatics," *Comput. Intell. Neurosci.*, vol. 2023, no. 1, p. 8393990, Jan. 2023, doi: 10.1155/2023/8393990.

[13] M. Joshi, A. Pal, and M. Sankarasubbu, "Federated Learning for Healthcare Domain - Pipeline, Applications and Challenges," *ACM Trans. Comput. Healthc.*, vol. 3, no. 4, pp. 1–36, Oct. 2022, doi: 10.1145/3533708.

[14] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, "Federated Learning for Healthcare Informatics," *J. Healthc. Inform. Res.*, vol. 5, no. 1, pp. 1–19, Mar. 2021, doi: 10.1007/s41666-020-00082-4.

[15] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, "Secure and Robust Machine Learning for Healthcare: A Survey," Jan. 21, 2020, *arXiv*: arXiv:2001.08103. doi: 10.48550/arXiv.2001.08103.

[16] D. C. Nguyen *et al.*, "Federated Learning for Smart Healthcare: A Survey," *ACM Comput Surv*, vol. 55, no. 3, p. 60:1-60:37, Feb. 2022, doi: 10.1145/3501296.

[17] R. Ye *et al.*, "OpenFedLLM: Training Large Language Models on Decentralized Private Data via Federated Learning," 2024, doi: 10.48550/ARXIV.2402.06954.

[18] W. Kuang *et al.*, "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*, in KDD '24. New York, NY, USA: Association for Computing Machinery, Aug. 2024, pp. 5260–5271. doi: 10.1145/3637528.3671573.

[19] C. Chen, X. Feng, J. Zhou, J. Yin, and X. Zheng, "Federated Large Language Model: A Position Paper," Jul. 17, 2023, *arXiv*: arXiv:2307.08925. Accessed: Aug. 30, 2024. [Online]. Available: http://arxiv.org/abs/2307.08925

[20] M. Sallam, "ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns," *Healthcare*, vol. 11, no. 6, Art. no. 6, Jan. 2023, doi: 10.3390/healthcare11060887.

[21] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," May 24, 2019, *arXiv*: arXiv:1810.04805. doi: 10.48550/arXiv.1810.04805.

[22] R. Luo *et al.*, "BioGPT: generative pre-trained transformer for biomedical text generation and mining," *Brief. Bioinform.*, vol. 23, no. 6, p. bbac409, Nov. 2022, doi: 10.1093/bib/bbac409.

[23] J. Lee *et al.*, "BioBERT: a pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, Feb. 2020, doi: 10.1093/bioinformatics/btz682.

[24] Prayitno *et al.*, "A Systematic Review of Federated Learning in the Healthcare Area: From the Perspective of Data Properties and Applications," *Appl. Sci.*, vol. 11, no. 23, Dec. 2021, doi: 10.3390/app112311191.

[25] L. Ouyang *et al.*, "Training language models to follow instructions with human feedback," *Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 27730–27744, Dec. 2022.

[26] T. Tu *et al.*, "Towards Generalist Biomedical AI," *NEJM AI*, vol. 1, no. 3, p. AIoa2300138, Feb. 2024, doi: 10.1056/AIoa2300138.

[27] H. Touvron *et al.*, "LLaMA: Open and Efficient Foundation Language Models," Feb. 27, 2023, *arXiv*: arXiv:2302.13971. doi: 10.48550/arXiv.2302.13971.

[28] E. J. Hu *et al.*, "LoRA: Low-Rank Adaptation of Large Language Models," Oct. 16, 2021, *arXiv*: arXiv:2106.09685. doi: 10.48550/arXiv.2106.09685.

[29] O. Aouedi, A. Sacco, K. Piamrat, and G. Marchetto, "Handling Privacy-Sensitive Medical Data With Federated Learning: Challenges and Future Directions," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 2, pp. 790–803, Feb. 2023, doi: 10.1109/JBHI.2022.3185673.

[30] C. Thapa and S. Camtepe, "Precision health data: Requirements, challenges and existing techniques for data security and privacy," *Comput. Biol. Med.*, vol. 129, p. 104130, Feb. 2021, doi: 10.1016/j.compbiomed.2020.104130.

[31] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nat. Mach. Intell.*, vol. 2, no. 6, pp. 305–311, Jun. 2020, doi: 10.1038/s42256-020-0186-1.

[32] Q. Li *et al.*, "A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 3347–3366, Apr. 2023, doi: 10.1109/TKDE.2021.3124599.