International Journal of Artificial Intelligence, Data Science, and Machine Learning
Grace Horizon Publication | Volume 6, Issue 1, 254 -257, 2025
ISSN: 3050-9262 | https://doi.org/10.63282/3050-9262.IJAIDSML-V6I1P129

Original Article

A Critical Review on Ethical and Privacy Challenges of AI-Powered Heal thcare Systems

Sujit Murumkar[1], Susmit Sen[2]
[1]Director, Business Information Management, Axtria Inc.
[2]Senior Manager - Data Governance & Data Mgmt, Albertsons.

Received On: 09/01/2025 Revised On: 26/01/2025 Accepted On: 13/02/2025 Published On: 04/03/2025

Abstract *- The emergence of AI-driven healthcare systems is deeming more and more on distributed analytics to train on sensitive clinical data without centralising raw-data. Federated learning (FL) has often been framed as an answer with privacy safeguards but healthcare implementations have revealed that FL transitions instead of eliminates ethical and privacy threats. An overview of privacy enhancement techniques in FL in health care presented in a review of Scopus described 216 records retrieved and a subsequent set of approximately forty records that were included, having passed the screening procedure, which reflected not only the fast growth of the evidence base but also its dispersal. [1]. Simultaneously, empirical research in imaging, clinical prediction, and internet-scale epidemiology demonstrates that privacy mechanisms (e.g., differential privacy) may decrease performance or increase inequities when data are non-IID, institutions do not have the same resources, or the external validity is low. These tensions have transformed privacy into a socio-technical governance issue: security measures, accountability, clinical safety and fairness should be considered as a whole and not as a side-note.*

Keywords *- Ethics of AI, Medical Privacy, Federated Learning, Differential Privacy, Governance, Responsibility.*

1. Introduction

The specifics of AI in healthcare make its ethical use unique since the model outputs may influence the diagnosis, triage, and resource allocation, whereas the training data are usually highly identifying clinical features. FL has usually been marketed as a mitigation since it retains data locally, but the fundamental ethical dilemma arises of whether local training produces any significant harm reduction in a threat environment, such as inference attacks, poisoning, or institutional abuse. A broad survey on privacy preservation of FL in smart healthcare highlights the threats of confidentiality, but also adversarial model manipulation and vulnerabilities of systems at large, and considers privacy as a moving target across devices, networks, and data controllers [2]. The ensuing dilemma is not merely to make privacy a part of it, but to determine what risks will be acceptable in cases where AI becomes integrated into clinical practices.

2. Evidence Support and Review Logic.

The quality of critical reviews is determined by the selectivity of the evidence and inferences drawn out of the evidence. Deductive and inductive approaches can be combined in qualitative analysis to structure interpretation while still allowing themes to emerge from heterogeneous evidence, which is useful when synthesising diverse healthcare FL privacy studies [16]. The management-methods scholarship differentiates between inductive and deductive logic and contends that the chain of reasoning should be presented by authors in a manner that they can be evaluated by reviewers on the basis of whether the conclusions are based on the evidence or the rhetorical framing. During qualitative synthesis, deductive and inductive methods of analysing data are typically coupled to organize complex domains without preventing emergent themes to emerge due to heterogeneous studies [16]. In the case of AI-driven healthcare privacy, this is important since numerous papers report good internal performance, but fewer external validation, and ethical implications may be manipulated when privacy is synonymous with data does not share, without analysing such factors as model leakage, governance, and deployment harm.

3. Federated Learning in Healthcare: Maturity and Scope.

The literature in healthcare FL has been increasing rapidly, yet it is not evenly developed in maturity of different tasks and settings. The initial pool of 200 articles reduced to a 67-article final corpus were reported in a systematic review and architecture proposal, 47% of which were published in 2020 and 79% of which case studies were using deep learning, a surge that surpassed the normalisation of evaluation practise. The smart healthcare Survey work on FL, too, situates the research area as broad, as it deals with resource management, security and privacy, incentives, and personalised FL; it asserts a holistic taxonomy in comparison with previous more partial surveys [12]. The mix hints to a rapid space where privacy preserving is often claimed as a paradigm characteristic instead of being established within realistic limitations. A systematic review filtered an initial pool of 200 articles down to 67, reporting that 47% were published in 2020 and that deep learning appeared in about

79% of case studies, which signals rapid growth alongside methodological concentration [8]

4. Mechanisms and Trade-Offs of Privacy

Improvement of privacy in FL is often based on the toolbox that encompasses: differential privacy (DP), homomorphic encryption, secure aggregation and hybrid solutions, but is not ethical neutral due to reassigning the risk and cost. One of the reviews categorised privacy improvement strategies in the categories of DP and encryption-based interventions, and its screening pipeline reflects that the literature is largely focused on technical controls but less consistent on clinical governance and transparency to the patients [1]. A survey of FL in the area of healthcare informatics also structures issues in statistical, system, and privacy dimensions, with privacy protection pitting against performance, communication efficiency, and operational reliability [5]. As these trade-offs are ethically significant since clinicians and patients directly bear the harms of poor sensitivity, slower inference, or systematic bias than they do abstract privacy guarantees.

5. Signals of the Tension between Privacy and Utility which are Empirical in Nature

A number of studies give tangible indications that privacy protection will be capable of reducing clinical-model performance in a quantifiable manner, though the magnitude would be contingent on task, threat model, and data heterogeneity. A privacy-conserving federated Alzheimer detection system achieved 89.5% accuracy with no privacy constraints as well as 81.88% accuracy with a DP setting of ε=2 indicating a substantial utility reduction with privacy controls [3]. A medical image analysis study that trained on 2,580 lung whole-slide images (1,806 train, 774 test) and reported DP-FL to have mean test and external accuracy of 0.823±0.01 and 0.707±0.01 respectively, and found non-private FL to maintain higher external accuracy at 0.741±0.01 with the same collaboration structure [10]. These findings highlight the point that privacy concerns cannot be regarded as free ethical judgement due to the shifting of clinical risk by privacy budgets and noise calibration.

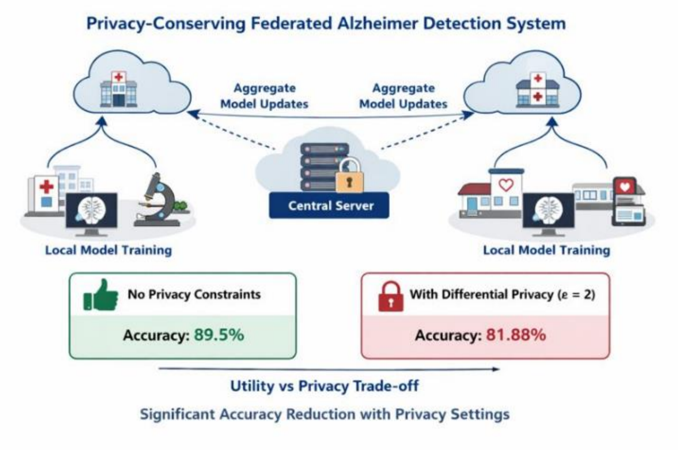
Privacy-Conserving Federated Alzheimer Detection System
Aggregate Model Updates
Aggregate Model Updates
Central Server
Local Model Training
Local Model Training
No Privacy Constraints
Accuracy: 89.5%
With Differential Privacy (ε = 2)
Accuracy: 81.88%
Utility vs Privacy Trade-off
Significant Accuracy Reduction with Privacy Settings

Fig 1: Privacy-Conserving Federated Alzheimer Detection System

[Source: 3]

6. Generalisation, Non-IID Data and External Validity.

External validity is a repeat of an ethical shortcoming given that healthcare systems differ based on population, equipment and clinical practice and that models that work well domestically may fail in different settings. Healthcare databases are stated to contain inconsistency, duplication, and logical errors, motivating a Python-based framework using profiling, rule-based validation, duplicate detection, and inconsistency checks [19]. A study using federated collaborative chest radiograph studies involved 45513 COVID-positive patients at a single site, and local models with internal AUROC of 0.94 demonstrated external AUROC of only 0.56 and 0.67 when defining the scope of 2 different sites, demonstrating a catastrophic deterioration with the switch of the distribution [6]. The same study observed that federated variants increased external AUROC (e.g., FedBN 0.78 and 0.70), but the site-to-site variability does persist which shows that the federation of variants does not necessarily resolve the issue of generalization, and may even conceal the issue when the aggregate metrics are reported [6]. This, ethically, brings about the issue of the harm accountability in the event that a model that was trained safely fails to perform well with respect to some hospitals or groups of patients.

7. Threats Models, Leakages, and the Boundaries of no Raw Data Sharing.

The main ethical fallacy is that the lack of the raw data removes the risks of privacy, whereas the gradients, updates, and model parameters may spill or be compromised. In FL, a specialised conversation about privacy sensitive medical information points to the fact that not only must FL address privacy issues across the globe but also locally but also to the fact that, in one case, trust-based mechanisms can be used to limit participation with a specific threshold, which is the threshold of 60 percent trust in one instance. Trust-based participation control is illustrated by a rule that prohibits clients with a trust value below 60%, showing that privacy governance can include exclusion thresholds rather than only encryption or DP [7]. Privacy-first privacy-scale health research studies show that meaningful modelling can be achieved with low participation, of around 2 of randomized participants reaching 99 of maximum AUC, but also report that privacy accounting can be significant with the central DP setting achieving AUC 0.83 at e=0.165 (=10 -5) and the local DP setting achieving AUC 0.83 at e=1.36 (=10 -9 per round). The ethical suggestion is that the risk of privacy must be assessed throughout the entire pipeline (client recruitment, aggregate update, and present the output) and not only on the slogan of data remains local. In a privacy-first federated health research, ~2% participation was reported as sufficient to reach 99% of maximum AUC, and a central DP configuration reported AUC 0.83 at ε=0.165 ($\delta=10^{-5}$), demonstrating that privacy accounting is measurable rather than symbolic [9].

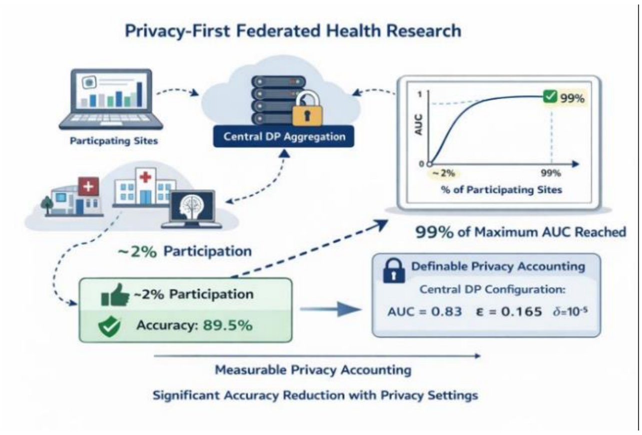
Privacy-First Federated Health Research
Particpating Sites
Central DP Aggregation
1
AUC
0
~2%
99%
% of Participating Sites
99%
99% of Maximum AUC Reached
~2% Participation
~2% Participation
Accuracy: 89.5%
Definable Privacy Accounting
Central DP Configuration:
AUC = 0.83 ε = 0.165 δ=10-5
Measurable Privacy Accounting
Significant Accuracy Reduction with Privacy Settings

Fig 2: Privacy First Federated Health Research-Measurable Accounting

[Source: 9]

8. Infrastructure Disparity and Operation Ethics

The healthcare AI based on privacy preservation carries inequality as to infrastructure: resource-abundant hospitals can engage more easily in FL, whereas the resources-poor environment can be left out or compromised, which poses an ethical threat of representational harm. An FL system preserving privacy based on a fog algorithm applied to a multi-hospital imaging dataset with a total of 21,165 images (70/30 train-test) and one image per-hospital (7,055 images/hospital) yielded the FL reports of precision/recall/F1 values of 67%/64%/93% (Hospital A), compared to local baselines of 61%/56%/87% [14]. IoMT FL system A blockchain-based enabled system defined node capabilities between 1,000 MIPS to 2,000,000 MIPS and reported energy consumption reduced by 41% and delay reduction by 28, indicating the operational incentive of optimisation of performance in the presence of heterogeneous devices [13]. Such outcomes are predictive of an ethical issue: privacy structures that presuppose highly educated compute, consistent connectivity or customised hardware can be systematic in favouring select institutions and populace.

9. Claims of Performance and Interpretability of Reported Gains.

A purported high improvement in accuracy in privacy-preserving FL should be viewed with skepticism as they vary between evaluation designs and the baselines might not reflect clinical reality. An implementation of a privacy-preservation framework within healthcare applications achieved an average accuracy of 97.69 versus 91.67 and 89.27 with a baseline method, and precision of 95.2 and a recall of 0.93, which places the framework in a more favorable perspective as being more accurate and more efficient [4]. High task performance was also reported with a privacy-preserving FL model, of 79% accuracy on IQVIA data and 98% on BC-TCGA data with privacy budgets of ε=30 and ε=10, respectively, and a statement federated training requires approximately 10 seconds [11]. Such figures must be considered as conditional on the construction of databases, label definitions, and threat models, since clinical safety is primarily reliant on failures and edge cases and not on headline accuracy.

10. Governance: Missing Accountability, Consent, and Transparency.

The governance failures even in case of privacy tools operating as intended may result in ethical harm due to lack of accountability, insufficient consent, and the lack of transparency in decision-making. Popular press Survey research underlines that healthcare FL cuts across technical levels (devices, networks, institutions) and has incentives and personalisation, making it difficult to allocate responsibility when a privacy breach or clinical error has taken place. A survey of healthcare FL also lists system and statistical issues affecting the results and comments that privacy measures can be co-optimised with compression and other operational decisions suggesting that privacy decisions are not made on ethical checklists. Various research frames healthcare data as deeply human and argues that data governance functions as a “moral architecture” so innovation does not outrun accountability [18]. The moral criterion thus becomes one of explainable governance due to which institutions are able to rationalise privacy allocations, participation principles and deployment limits in a manner that are significant to regulators, clinicians as well as patients. Traditional consent management is described as centralised, non-transparent, hard to audit, and lacking fine-grained, verifiable user control and audit trails for purpose-specific data use [17].

11. Synthesis and Critical Implications.

There are three imperative themes that come out in this evidence. To begin with, privacy mechanisms necessitate quantifiable utility costs in certain clinical situations, such as a decrease in accuracy in DP and worse performance in privacy limits indicates that privacy turns into a variable of clinical safety and not a checkbox of legal compliance [3]. Second, the methodology is an ethical risk since despite the high internal performance, there can be a complete breakage on the collapse of internal performance under hospital-to-hospital shift, and FL can only partially address it without the convincing reporting and calibration measures [6]. Third, privacy-preserving is more effectively viewed as a layered property of governance: the threat models, DP accounting, and trust thresholds as well as infrastructure capability inform who gains, who suffers harm, and who holds accountability in case of model failure.

12. Conclusion

The issue of ethical and privacy concerns in AI-powered healthcare systems remain in place as the risk is pushed not to the storage of data but rather to model updates, system design and governance decisions. There are quantitative results in various applications which indicate the existence of performance and generalisation change based on privacy budgets, aggregation, and heterogeneity, and therefore one cannot easily judge morality on the basis of architectural claims. The most justifiable way is to consider privacy as equivalent to clinical safety, equity, and accountability, and explicitly record the threat models, privacy parameters, and cross-site analysis as opposed to the use of a general-purpose statement. This area has strong ethics that is attained when both technical protective measures and institutional governance minimise harm and maintain clinically significant performance.

References

[1] X. Gu, F. Sabrina, Z. Fan, and S. Sohail, “A review of privacy enhancement methods for federated learning in healthcare systems,” *Int. J. Environ. Res. Public Health*, vol. 20, no. 15, Art. no. 6539, 2023, doi: 10.3390/ijerph20156539.

[2] M. Ali, F. Naeem, M. Tariq, and G. Kaddoum, “Federated learning for privacy preservation in smart healthcare systems: A comprehensive survey,” *IEEE J. Biomed. Health Inform.*, vol. 27, no. 2, pp. 778–789, 2023, doi: 10.1109/JBHI.2022.3181823.

[3] J. Li, Y. Meng, L. Ma, S. Du, H. Zhu, Q. Pei, and X. Shen, “A federated learning based privacy-preserving smart healthcare system,” *IEEE Trans. Ind. Informatics*, vol. 18, no. 3, pp. 2021–2031, Mar. 2022, doi: 10.1109/TII.2021.3098010.

[4] M. Abaoud, M. A. Almuqrin, and M. F. Khan, “Advancing federated learning through novel mechanism for privacy preservation in healthcare applications,” *IEEE Access*, vol. 11, pp. 83562–83579, 2023, doi: 10.1109/ACCESS.2023.3301162.

[5] 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, 2021, doi: 10.1007/s41666-020-00082-4.

[6] T. J. Loftus *et al.*, “Federated learning for preserving data privacy in collaborative healthcare research,” *Digit. Health*, vol. 8, Art. no. 20552076221134455, 2022, doi: 10.1177/20552076221134455.

[7] 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, 2023, doi: 10.1109/JBHI.2022.3185673.

[8] R. S. Antunes, C. A. da Costa, A. Küderle, I. A. Yari, and B. Eskofier, “Federated learning for healthcare: Systematic review and architecture proposal,” *ACM Trans. Intell. Syst. Technol.*, vol. 13, no. 4, Art. no. 54, pp. 54:1–54:23, 2022, doi: 10.1145/3501813.

[9] A. Sadilek *et al.*, “Privacy-first health research with federated learning,” *npj Digit. Med.*, vol. 4, no. 1, Art. no. 132, 2021, doi: 10.1038/s41746-021-00489-2.

[10] M. Adnan, S. Kalra, J. C. Cresswell, G. W. Taylor, and H. R. Tizhoosh, “Federated learning and differential privacy for medical image analysis,” *Sci. Rep.*, vol. 12, no. 1, Art. no. 1953, 2022, doi: 10.1038/s41598-022-05539-7.

[11] T. U. Islam, R. Ghasemi, and N. Mohammed, “Privacy-preserving federated learning model for healthcare data,” in *Proc. 2022 IEEE 12th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Las Vegas, NV, USA, Jan. 2022, pp. 281–287, doi: 10.1109/CCWC54503.2022.9720752.

[12] D. C. Nguyen, Q.-V. Pham, P. N. Pathirana, M. Ding, A. Seneviratne, Z. Lin, O. A. Dobre, and W.-J. Hwang, “Federated learning for smart healthcare: A survey,” *ACM Comput. Surv.*, vol. 55, no. 3, Art. no. 60, pp. 1–37, 2022, doi: 10.1145/3501296.

[13] A. Lakhan *et al.*, “Federated-learning based privacy preservation and fraud-enabled blockchain IoMT system for healthcare,” *IEEE J. Biomed. Health Inform.*, vol. 27, no. 2, pp. 664–672, 2023, doi: 10.1109/JBHI.2022.3165945.

[14] M. Butt *et al.*, “A fog-based privacy-preserving federated learning system for smart healthcare applications,” *Electronics*, vol. 12, no. 19, Art. no. 4074, 2023, doi: 10.3390/electronics12194074.

[15] J. Woiceshyn and U. Daellenbach, “Evaluating inductive vs deductive research in management studies: Implications for authors, editors, and reviewers,” *Qual. Res. Org. Manag.: Int. J.*, vol. 13, no. 2, pp. 183–195, 2018, doi: 10.1108/QROM-06-2017-1538.

[16] T. Azungah, “Qualitative research: Deductive and inductive approaches to data analysis,” *Qual. Res. J.*, vol. 18, no. 4, pp. 383–400, 2018, doi: 10.1108/QRJ-D-18-00035.

[17] S. Murumkar, “Smart contract-driven consent management for personal data sharing,” *Int. J. Artif. Intell., Data Sci., Mach. Learn.*, vol. 4, no. 2, pp. 135–141, 2023, doi: 10.63282/3050-9262.IJAIDSML-V4I2P115.

[18] C. Tayal, “Data quality assessment and cleaning framework for healthcare databases using Python,” *Int. J. Artif. Intell., Data Sci., Mach. Learn.*, vol. 3, no. 4, pp. 107–112, 2022, doi: 10.63282/3050-9262.IJAIDSML-V3I4P112