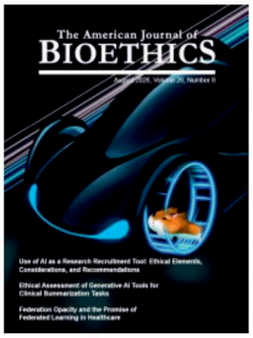

The American Journal of Bioethics

ISSN: 1526-5161 (Print) 1536-0075 (Online) Journal homepage: www.tandfonline.com/journals/uajb20

Does Swarm Learning Solve Federation Opacity? The Need for Governance Beyond Decentralization

Marieke Bak, Stefanie Warnat-Herresthal, Theresa Willem, Bettina Zimmermann, Jonas Schulte-Schrepping, Lucas Secchim Ribeiro, Joachim Schultze & Stuart McLennan

To cite this article: Marieke Bak, Stefanie Warnat-Herresthal, Theresa Willem, Bettina Zimmermann, Jonas Schulte-Schrepping, Lucas Secchim Ribeiro, Joachim Schultze & Stuart McLennan (2026) Does Swarm Learning Solve Federation Opacity? The Need for Governance Beyond Decentralization, The American Journal of Bioethics, 26:8, 136-139, DOI: 10.1080/15265161.2026.2690963

To link to this article: https://doi.org/10.1080/15265161.2026.2690963

Published online: 03 Aug 2026.

Pratt B, Loff B. 2012. Health research systems: promoting health equity or economic competitiveness? Bull World Health Organ. 90(1):55–62. https://doi.org/10.2471/BLT.11.092007

Sauer CM, Pucher G, Celi LA. 2024. Why federated learning will do little to overcome the deeply embedded biases in clinical medicine. Intensive Care Med. 50(8):1390–1392. https://doi.org/10.1007/s00134-024-07491-8

Tu X et al. 2022. Incentive mechanisms for federated learning: from economic and game theoretic perspective. IEEE Trans Cogn Commun Netw. 8(3):1566–1593. https://doi.org/10.1109/TCCN.2022.3177522

Zhang F et al. 2024. Recent methodological advances in federated learning for healthcare. Patterns (N Y). 5(6):101006. https://doi.org/10.1016/j.patter.2024.101006

THE AMERICAN JOURNAL OF BIOETHICS
2026, VOL. 26, NO. 8, 136–139
https://doi.org/10.1080/15265161.2026.2690963

Taylor & Francis
by informa

OPEN PEER COMMENTARIES

Does Swarm Learning Solve Federation Opacity? The Need for Governance Beyond Decentralization

Marieke Bak[a,b], Stefanie Warnat-Herresthal[c], Theresa Willem[a,d], Bettina Zimmermann[a,e], Jonas Schulte-Schrepping[c], Lucas Secchim Ribeiro[c], Joachim Schultze[c] and Stuart McLennan[a,e]

[a]Technical University of Munich; [b]Amsterdam UMC; [c]DZNE; [d]Helmholtz AI; [e]University Hospital Basel

Hatherley and colleagues (2026) analyze a key problem in federated learning (FL) that they call "federation opacity". They argue that FL creates a distinctive epistemic and ethical challenge because no stakeholder can directly access, analyze, or curate the full dataset on which the model has been trained. The target article explains that while this opacity is the key selling point of FL for healthcare organizations because it helps overcome data protection hurdles, the same feature can obstruct accountability, make data quality problems difficult to trace, and increase vulnerability to poisoning attacks. Some of us previously made these points in a commentary for Nature Machine Intelligence, where we criticized the tendency to treat FL as a techno-fix for a data ethics problem (Bak et al. 2024). Because data do not move to a central repository, FL is too easily described as "privacy-preserving," and its use is sometimes assumed to reduce or even remove ethical and legal concerns. We noted the problems arising from the fact that "each node sees only its own data", and discussed issues of contribution unfairness, data poisoning, data heterogeneity, difficulties with bias detection, and the fact that FL still requires justifications for data collection and sharing, such as patient consent. Hatherley and colleagues describe these issues under the umbrella of federation opacity. This concept clarifies why certain ethical issues in FL are not merely familiar machine learning problems repeated in a distributed setting: according to Hatherley et al., they are shaped by the inherent fragmentation of training data which is inaccessible for direct comparison or curation across institutions.

This argument invites a further question. If we assume that federation opacity arises from FL's distributed architecture, what happens when machine learning is further decentralized? The field is changing rapidly, resulting in ongoing efforts to develop distributed approaches that are more decentralized than FL, which still relies on a central server to aggregate contributions into a global model. For example, Swarm Learning (SL) moves beyond FL by removing the central aggregation server and replacing it with peer-to-peer coordination, edge computing, and a blockchain-based coordination layer. Warnat-Herresthal et al. (2021) demonstrated the promise and feasibility of this approach by collaboratively training disease classifiers for COVID-19, tuberculosis, and leukemia at multiple clinical institutions. Their resulting classifiers outperformed those trained individually at each site while upholding local privacy regulations. As there is no central server, no single trusted party, no single point of failure, and a tamper-resistant record of participation, could the shift to more decentralized

CONTACT Marieke Bak marieke.bak@tum.de Institute of History and Ethics in Medicine, Technical University of Munich, Munich, Germany.

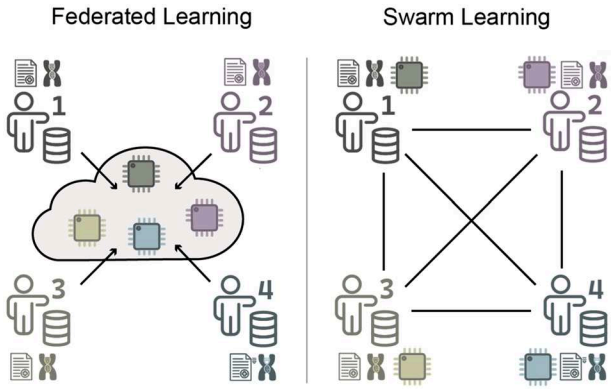

Figure 1. The structure of Federated Learning which keeps data with the contributor, performs computation locally, and coordinates model parameters through a central server; compared to Swarm Learning where both data and parameters are kept at edge without a central custodian. Adapted from Warnat-Herresthal et al. (2021).

machine learning be an improvement over FL in terms of ethics? Federation opacity may offer a useful lens through which to examine not only FL, but also newer approaches such as SL (Figure 1).

FROM FEDERATION TO SWARM: ARE THE ETHICS DIFFERENT?

SL differs from classic FL in an important architectural respect. In FL, local nodes train a model on their own data and send updates to a central aggregation server. In SL, the central server is removed. Nodes train locally and exchange model parameters or gradients through a peer-to-peer network. Coordination is handled through a distributed ledger, such as a blockchain. At defined synchronization intervals, the blockchain's architecture (e.g. predefined rule sets and/or smart contracts by SL network participants) triggers the election of a temporary leader node. This leader aggregates the received model updates and produces a merged global model. The updated model is broadcast back to all nodes for optimization of further training rounds. The cycle then repeats: local training on private data, exchange of learned parameters, aggregation, redistribution. The blockchain's immutable ledger ensures all participants follow the same protocol and provides an auditable record of the training history, but no raw patient data is exposed. In principle, this design reduces dependence on a single trusted coordinator and provides a more auditable record of the learning process. This is a technical difference that may matter ethically, because a central aggregation server can be a point of power, trust, and failure. While SL platforms might still be provided and organized by one partner, all participating institutions in an SL network need to agree to a clearly defined data type and standard before the training of the machine learning model can be executed, which reduces major risks of FL.

Hatherley and colleagues argue that federation opacity can make it difficult to identify which node is responsible for errors or performance problems: SL's ledger helps address this by recording the provenance of updates (e.g. who participated, when updates were submitted, and whether the agreed protocol was followed). Yet knowing that a node has submitted an update does not, by itself, reveal the quality of the updated model or underlying data, nor does it determine who should investigate harms, or exclude a node from the network. These issues thus require not merely technical traceability, but governance arrangements created within the SL network before any learning, re-learning, updating, or continuous learning occurs, with clear criteria regarding data quality standards, quality control procedures, legal liabilities, and other governance responsibilities.

The SL approach also addresses certain issues of institutional unfairness (or "free-riding") by using the blockchain to track resource use and contributions of nodes in the swarm (Cila et al. 2020) and potentially even provide organizations with personalized models based on their contributions (Tajabadi and Heider 2024). However, especially in healthcare, contributions should not be reduced to a logic in which more data or more computational power automatically earns greater benefit: smaller centers may represent populations that are clinically important or otherwise underrepresented. This does not mean that all participants should necessarily receive identical benefits regardless of contribution. It means that the fairness question is normative before it is computational. What counts as a valuable contribution (e.g. volume, quality, inclusivity, diversity, infrastructure, expertise) will differ across projects. Technical mechanisms for contribution tracking can support fair governance, but they cannot define fairness on their own. Moreover, formal tracking should be designed in a way that does not have negative effects on existing relationships based on trust and reciprocity (Bak et al. 2023).

Finally, SL can be described as improving data privacy because removing the central server may reduce some risks associated with a malicious or compromised server. While risks of data poisoning or inference attacks are not entirely eliminated, SL reduces these risks beyond central and federated approaches. In fact, by minimizing these threats, SL could open discussions about necessary safeguards to

protect against downstream harm of AI applications and misuses of data beyond reidentification concerns, to which Hatherley and colleagues rightly call attention.

In their target article, the core issue is that no stakeholder can inspect the full training dataset. In SL, blockchain coordination indeed does not make participants' data visible to one another, nor should it, if the data protection purpose of the architecture is to be preserved. However, data quality can still be assessed locally through agreed criteria before data are injected into model learning. Such criteria may include clinical relevance, representativeness, labelling quality, metadata granularity, inclusion and exclusion criteria, and quality control procedures. Thus, SL does not eliminate data opacity through technical traceability alone, but it can reduce opacity through shared governance, clear responsibilities, and continued involvement of all data providers throughout the model and data life cycle.

DECENTRALIZATION IS NOT A GOVERNANCE ENDPOINT

Medicine is inherently decentralized, and we need technical solutions that make better use of health data within these existing structures. The enthusiasm for decentralization in health data science is therefore not surprising and becomes apparent in developments such as the European Health Data Space (Kogut-Czarkowska and Shabani 2025). We find that just as with FL, the promise of SL should be taken seriously. SL supports collaborative medical machine learning in settings where data pooling is not feasible and central coordination is undesirable. However, as we noted in our earlier comment on FL (2024), the most relevant ethical question is not whether data are accessed centrally or locally, but whether any use of data is governed by shared quality criteria, appropriate checks, and renewed assessment for the specific machine learning task. Noise, bias, model shift, and data drift can undermine both centralized and distributed approaches if they are not detected. Distributed techniques, including FL and SL, may even have advantages here, because local data holders remain involved in assessing data quality, heterogeneity, and the conditions for reuse.

In our view, the ethical task is not to be focused only on the technical realizations (central, FL, or SL). Rather, the value and ethical robustness of these technical solutions depend on the surrounding governance arrangements: rules for participation, standards for data quality, security requirements, audit procedures, benefit-sharing agreements, mechanisms for patient and public involvement, and clear lines of responsibility. While Hatherley and colleagues suggest to distinguish between the technical and ethical aspects of the problem, we find that these are often intertwined. FL and SL networks should be evaluated as sociotechnical systems that bring trade-offs and responsibility questions, requiring ongoing interdisciplinary work. As with all AI in healthcare, decentralized machine learning, too, needs to be subject to case-by-case reflection that helps networks to remain accountable to the patients, publics, clinicians, and institutions who entrusted them with their data.

AUTHOR CONTRIBUTIONS

CRediT: **Marieke Bak**: Conceptualization, Project administration, Writing – original draft, Writing – review & editing; **Stefanie Warnat-Herresthal**: Writing – review & editing; **Theresa Willem**: Formal analysis, Writing – review & editing; **Bettina Zimmermann**: Formal analysis, Visualization, Writing – review & editing; **Jonas Schulte-Schrepping**: Writing – review & editing; **Lucas Secchim Ribeiro**: Writing – review & editing; **Joachim Schultze**: Formal analysis, Supervision, Validation, Writing – review & editing; **Stuart McLennan**: Conceptualization, Supervision, Writing – review & editing.

FUNDING

All authors of this commentary (except SWH) participate in the 'Swarm Learning for Pandemic Preparedness' project which is supported by the Volkswagen Stiftung (#8000354).

ORCID

Marieke Bak http://orcid.org/0000-0003-0655-0743
Stefanie Warnat-Herresthal http://orcid.org/0000-0002-0890-5774
Theresa Willem http://orcid.org/0000-0001-7643-8816
Bettina Zimmermann http://orcid.org/0000-0001-7047-4496
Jonas Schulte-Schrepping http://orcid.org/0000-0001-7333-8924
Lucas Secchim Ribeiro http://orcid.org/0000-0001-8876-3827
Joachim Schultze http://orcid.org/0000-0003-2812-9853
Stuart McLennan http://orcid.org/0000-0002-2019-6253

REFERENCES

Bak MAR et al. 2024. Federated learning is not a cure-all for data ethics. Nat Mach Intell. 6:370–372. https://doi.org/10.1038/s42256-024-00813-x

Bak MAR, Ploem MC, Tan HL, Blom MT, Willems DL. 2023. Towards trust-based governance of health data research. Med Health Care Philos. 26(2):185. https://doi.org/10.1007/s11019-022-10134-8

Cila N, Ferri G, de Waal M, Gloerich I, Karpinski T. 2020. The blockchain and the commons: Dilemmas in the design of local platforms. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery. p. 1–14.

Hatherley J, Søgaard A, Ballantyne A, Pauwels R. 2026. Federation opacity and the promise of federated learning in healthcare. Am J Bioethics. 26(8):119–132. https://doi.org/10.1080/15265161.2026.2637093

Kogut-Czarkowska M, Shabani M. 2025. Federated networks and secondary uses of health data: challenges in ensuring appropriate safeguards for sharing health data under the GDPR and EHDS. In The European health data space. Routledge. p 229–254.

Tajabadi M, Heider D. 2024. Fair swarm learning: improving incentives for collaboration by a fair reward mechanism. Knowledge-Based Syst. 304:112451. https://doi.org/10.1016/j.knosys.2024.112451

Warnat-Herresthal S et al. 2021. Swarm learning for decentralized and confidential clinical machine learning. Nature 594: 265–270. https://doi.org/10.1038/s41586-021-03583-3

THE AMERICAN JOURNAL OF BIOETHICS
2026, VOL. 26, NO. 8, 139–142
https://doi.org/10.1080/15265161.2026.2690949

OPEN PEER COMMENTARIES

Whose Federation? Federation Opacity Through a Global Health Lens

A. M. Viens[a], Elham Dolatabadi[b,c] and Jude D. Kong[d]

[a]School of Global Health, York University; [b]School of Health Policy & Management, York University; [c]Vector Institute; [d]Dalla Lana School of Public Health, University of Toronto

Federated learning (FL) has emerged as a promising approach to addressing the longstanding tension between data access and patient privacy in healthcare (Antunes et al. 2022). By enabling institutions to collaboratively train models without sharing patient-level data, FL is often presented as a means of building more representative datasets, improving model performance, and expanding participation in artificial intelligence (AI) development (Yoo et al. 2022). Yet these potential benefits come with important limitations.

Hatherley et al. (2026) offer a compelling analysis of federation opacity (FO)—the condition in which no individual can access, analyze, or curate the full dataset on which a FL model has been trained. Their account persuasively outlines how FO can undermine institutional accountability, complicate fairness assessments, and leave FL models vulnerable to data poisoning. As FL expands from collaborations among well-resourced hospitals to networks that span countries, health systems, and populations with markedly different health priorities, FO becomes intertwined with questions of global health equity.

The challenge is no longer only whether participants can scrutinize the data that shaped a model, but whether they can understand whose data, priorities, and disease burdens are most strongly reflected in it. Federated systems are often framed as mechanisms for broadening participation in AI development. Yet the opacity that enables such collaboration can also obscure differences in representational influence and governance capacity across participating institutions in high-income country (HIC) versus low- and middle-income country (LMIC) settings.

Viewed through a global health lens, FO raises a series of interconnected questions. Whose disease burdens are most strongly represented in the models that federations produce? Which institutions possess the resources and governance capacity to shape their development? Who can determine whether participation in a federation translates into meaningful benefit? And, ultimately, what happens when these asymmetries become difficult to observe or contest? We suggest that answering these questions reveals FO not simply as a problem of transparency, but as a potential mechanism

CONTACT A. M. Viens amviens@yorku.ca School of Global Health, York University Faculty of Health, Toronto, Canada.