Distributed Machine Learning and the Semblance of Trust

Dmitrii Usynin [1][2][3], **Alexander Ziller** [2][3], **Daniel Rueckert** [1][2], **Jonathan Passerat-Palmbach** [1],
Georgios Kaissis [1][2][3]

[1]Department of Computing, Imperial College London
[2]Institute for Artificial Intelligence and Informatics in Medicine, Technical University of Munich
[3]Institute of Diagnostic and Interventional Radiology, Technical University of Munich

Abstract

The utilisation of large and diverse datasets for machine learning (ML) at scale is required to promote scientific insight into many meaningful problems. However, due to data governance regulations such as GDPR as well as ethical concerns, the aggregation of personal and sensitive data is problematic, which prompted the development of alternative strategies such as distributed ML (DML). Techniques such as Federated Learning (FL) allow the data owner to maintain data governance and perform model training locally without having to share their data. FL and related techniques are often described as *privacy-preserving*. We explain why this term is not appropriate and outline the risks associated with over-reliance on protocols that were not designed with formal definitions of privacy in mind. We further provide recommendations and examples on how such algorithms can be augmented to provide guarantees of governance, security, privacy and verifiability for a general ML audience without prior exposure to formal privacy techniques.

Introduction

Machine learning (ML) has shown promise in solving a number of important problems such as disease survival prediction (Vanneschi et al. 2011; Yan et al. 2020) or early-stage cancer discovery (Kourou et al. 2015). However, in order to train ML models capable of solving these tasks effectively, as well as fulfilling requirements such as fairness and generalisation, large, high-quality and unbiased datasets are required. Procuring these datasets has been problematic in contexts that rely on scarce, sensitive data especially in the context of healthcare. So far, this issue has been alleviated through data centralisation, where these datasets are aggregated at a single location, and model training is performed. However, the introduction of stricter data protection and governance regulations (e.g. the general data protection regulation (Radley-Gardner, Beale, and Zimmermann 2016)), and an increased societal awareness of privacy issues resulting in public resistance against aggressive data collection, centralisation of data is increasingly becoming an unsustainable solution.

As a result, the ML community has witnessed a surge in interest for scalable, privacy-preserving technologies allow-

ing large-scale training of ML models on distributed data across geographically distant institutions. The common factor of this family of protocols is their reliance on sharing the algorithms or the model updates instead of the datasets, thus remaining in line with existing data governance and data protection regulations. This allows the researchers to conduct collaborative machine learning at scale while minimising the data transmitted. One such mechanism, namely *federated learning* (FL) (Konečný et al. 2016), has received particular appraisal from the research community and has seen widespread adoption in a large number of studies (Bonawitz et al. 2019; Nasirigerdeh et al. 2020; Brisimi et al. 2018; Roy et al. 2019; Nasirigerdeh et al. 2021; Zhao et al. 2018; Brisimi et al. 2018). However, while FL –by design– only avoids transmitting data and thus preserves data governance, it is often wrongly referred to as *privacy-preserving*. In fact, collaboratively trained models can unintentionally leak information about the sensitive properties of the training population. Attacks on the models trained and transmitted between institutions have, on numerous occasions, been shown to be able to reveal sensitive information (Usynin et al. 2021; He, Zhang, and Lee 2019; Geiping et al. 2020; Shokri et al. 2017; Zhang et al. 2020; Ziller et al. 2021). Thus, FL and other governance mechanisms alone are insufficient means of privacy protection and require augmentations in the forms of formal privacy enhancing technologies (PETs).

Beyond privacy, a number of additional considerations arise when data is processed, independent of whether such processing occurs centrally or –as in FL– in a decentralised manner. The requirements towards such systems are outlined in a theoretical framework named *structured transparency* (ST) (Trask et al. 2020; Kaissis et al. 2020), whose terminology we will employ in the current work. These include:

1. The ability to train a model collaboratively without revealing the input data to other contributors (**input privacy**);

2. The protection of private information which can be learned from the results (i.e. the output) of the computation (**output privacy**);

3. The ability to verify the origin of the computation's result i.e. that the model update is not submitted by an unauthorised party (**input verification**);

4. The capability to guarantee the correctness of the output

and that the processing of inputs is honest (**output verification**)

5. The ability to control data ownership and exercise governance over it (**flow governance**).

Adhering to these principles (along with others, such as explainability, that lie outside of scope of this work) is fundamental to frameworks that claim to be trustworthy, giving sound guarantees of privacy to the individuals whose sensitive data is utilised, but also of reliability of the training protocol in general. Beyond FL, (and in part due to its vulnerabilities), a number of other DML frameworks have been introduced, for instance Swarm Learning (SL) (Warnat-Herresthal et al. 2021), Split Learning (Vepakomma et al. 2018) or gossip learning (Hegedűs, Danner, and Jelasity 2019)). These attempt to address some of the aforementioned requirements, but often fail to offer full ST guarantees either. In this article, motivated by the increasing popularity of DML, we aim to equip ML practitioners interested in applying DML techniques or reading related literature with an overview of relevant techniques, so that they can avoid misunderstandings and common pitfalls.

Our contributions can be summarised as follows:

- We provide definitions for the commonly used (and misunderstood) terms *governance, privacy, secrecy, security, accountability* and *verification*.

- We contextualise these techniques within the structured transparency framework.

- Finally, we discuss mechanisms by which the governance-preserving attributes of DML frameworks can be supplemented to render them fully privacy-preserving, verifiable, and secure.

Governance

In its core, data governance is a framework that defines the way data should be handled from the perspectives of *access*, *ownership* and *auditability* (Eryurek et al. 2021). For the purposes of distributed algorithmic processing, we consider auditability (that is, the ability of being inspected for e.g. quality) as a process that is either performed locally by the data owner on their own data or, alternatively, by the central aggregator on the updates submitted by the members of the consortium. We discuss auditability in the *Accountability* Section.

Central to exercising data governance is the ability to control data *locality* and data *accessibility*. In frameworks such as FL, control over data locality is maintained through local model training and subsequent update sharing, meaning that the data does not leave its owner owner and only the results of the algorithmic processing of this data are shared with the rest of the federation.

Nevertheless, DML allows (indirectly) accessing the data (or, more precisely, computational results derived from it), to the members of the federation. To prevent unwarranted or unauthorised access, security measures must be put in place, which are discussed below. In this regard, governance over data is maintained in DML paradigms such as FL, provided a suitable framework for auditing the input data is in place.

However, even though appropriately audited data remains local and is only accessible through the collaborative learning protocol, the protection of privacy and secrecy are contingent on additional techniques being utilised.

Privacy and secrecy

In this section we consider the fundamental differences between the concepts of *secrecy* and *privacy*, which are often conflated.

Privacy is the ability to control how much can be learned from the data about an individual. In the context of ST, there exist two separate privacy-related concepts: input and output privacy. Input privacy controls the extent to which the input data is visible and accessible to other actors, whereas output privacy is concerned with how much can be learnt from the data itself. The former can be achieved through strategies that maintain the secrecy of the computation, whereby secrecy implies that the sensitive data itself cannot be seen by anyone other than the data owner, concealing it from other parties. While decentralised protocols such as FL or SL allow data owners to enforce governance over their data and prevent unwarranted access to the training data directly, they do not provide the contributors with any mechanisms to control what can be *inferred* about that data when it is used for model training (output privacy). Input and output privacy pursue two complementary goals: secrecy prevents data from being usable when accessed inappropriately while privacy reduces the amount of information that can be extracted from the data about the individual while not wholly preventing drawing insights from the data otherwise.

The secrecy of datasets or of the computational process itself is often maintained through the application of formal cryptographic protocols. For example, individual contributions can be encrypted to conceal them from other participants during training. One such solution, namely homomorphic encryption (HE) (Gilad-Bachrach et al. 2016; Hesamifard, Takabi, and Ghasemi 2017) allows the federation to run the entire training procedure on encrypted data and only decrypting the end result of the computation. Another solution, namely secure multi-party computation (SMPC) (Rouhani, Riazi, and Koushanfar 2018; Mohassel and Zhang 2017) offers the federation an option of masking their contributions through splitting them into a number of encrypted *shares*, which, unless aggregated with the agreement of a quorum of participants (sometimes all participants), do not allow another client to learn anything about the model update. This fact also renders SMPC an attractive option for distributed data governance, where only a given number of parties are able to reveal the data that is shared between them. Most DML algorithms can be augmented with either of these mechanisms to provide the data owners with guarantees of secrecy, and, as a consequence, of input privacy. However, it is important to outline that typically such mechanisms usually come with a performance overhead that can affect the training time and/or the utility of the final model (Kaissis et al. 2021; Mohassel and Zhang 2017).

In contrast, enforcing output privacy requires limiting the information that can be derived about an individual from the data used for model training. While DML can guarantee data

governance, it does not in itself offer *any* meaningful privacy guarantees to the data owners. Therefore, the over-reliance on DML protocols alone, and their further adoption without the inclusion of formal privacy-preserving mechanisms must be viewed critically. Differential privacy (DP) (Dwork et al. 2006), which allows to objectively quantify and bound the amount of information that can be inferred from the training data, has established itself as the gold standard of formal privacy protection. DP allows data owners to perform model training while offering the individuals whose data is used to train the model, a quantifiable privacy guarantee in the form of a *privacy budget*. This privacy budget can be thought of as *fungible*, because it can be expended through model training. After an individual's privacy budget is exhausted, no further interaction with their data is permitted. The combination of DML with DP and cryptographic tools can equip the federation with the means of ascertaining governance as well as input and output privacy.

Security

In addition to the concepts of privacy and secrecy outlined above, we additionally need to consider *security*. In general, security is a property of a protocol or system where the protocol or system *as-a-whole* cannot be threatened by an adversarial actor. Security is thus not just complementary to privacy and secrecy, but a prerequisite for the design of any private and secure system. For instance, the physical security of the buildings in which computational equipment is housed is paramount to ensuring that these systems cannot be tampered with or destroyed. Moreover, a series of software and hardware measures are employed to safeguard the security of the learning protocol. For example, technologies such as transport layer security (TLS) are deployed to encrypt the data *in transit* and protect it against adversaries who could intercept data packets and obtain potentially sensitive information (so-called *man-in-the-middle* attacks). In general, the security of the system concerns *all* factors that can be exploited by adversaries that are covered in the ST framework and beyond. Issues such as physical unwarranted access to the data owner cannot be mitigated regardless of the privacy measures deployed, as these are covered under the notion of physical security. Similarly, issues such as attacks that are not exploiting/targeting the DML models directly (e.g. malicious hackers that attempt to steal the dataset through exploitation of the site itself rather than the trained model) do not lie within the scope of the ST framework.

Verification

In order to verify that the model is trained well on the underlying learning task, the federation needs to establish a method to ascertain that the data of the federation was used faithfully and the training protocol was not subverted by any party. Currently, no published DML implementation support this notion of output verification, as it comes in direct conflict with the notions of privacy described above, as such verification could entail inspection of the training protocol and –as a result– of sensitive data supplied by individual participants. One solution that can be employed for this task

can build upon techniques from the verifiable computing (VC) domain to enable the actors of the federation to certify the results they compute. VC encompasses methods that can be used to attest that a given computation's result was produced by a specific instance of an algorithm, matching its dependencies to the very bit, as pre-agreed upon between the *prover*, who runs the computation, and the *verifier*, who seeks to obtain guarantees of integrity of the computation, i.e. that the result of the computation was no malformed and represents an honest processing of inputs by the algorithm.

There exist two main strategies of implementations of VC in the context of ML: one involves special hardware embodied as trusted execution environment (TEE) (Chen et al. 2020; Lee et al. 2020a) and relies on the hardware properties in conjunction with a mechanism termed remote attestation (RA) (Ali, Nauman, and Jan 2018) to fulfil the verifiable trait of the execution. We note that TEEs (also termed *secure enclaves*) can also be used for end-to-end encrypted computations, thus fulfilling the role of an input privacy mechanism. The second form of VC that has gained traction over the past few years is entirely software-based and relies on cryptographic proofs. A popular form of cryptographic proof, namely *zero knowledge succinct non-interactive arguments of knowledge* (zk-SNARKs) (Bitansky et al. 2014), gained significant interest from the research community and has been previously leveraged in the context of neural network verification (Lee et al. 2020b; Weng et al. 2021). Augmented with these mechanisms, DML is able to guarantee that the computation was performed faithfully.

We note that –although verification of correctness is possible with these techniques– the verification of data *quality* is much more complex. For instance, a number of measures may be employed to gauge data quality, such as the reduction in model uncertainty obtained through each individual data sample. Moreover, members of the federation may agree upon metrics other than data quality (e.g. the speed with which a computational result is returned), to determine participant reimbursement.

Accountability

In a setting where several parties collaborate to solve a learning task, one cannot easily anticipate the intentions of each individual participant, as some actors might be actively attempting to subvert the training protocol. As a result, attacks on the utility of the resulting model are possible in decentralised learning, such as model poisoning (Fang et al. 2020; Yang et al. 2017) or backdoor attacks (Bagdasaryan et al. 2020; Bagdasaryan and Shmatikov 2020). Thus, paradigms such as FL require substantial augmentations to mitigate potential adversarial influence from being incorporated into the jointly trained model. This influence can take multiple forms ranging from colluding with other participants to submitting malformed updates or not submitting model updates at all, causing the training procedure to halt indefinitely. The ability to track the source of malicious interference is therefore a required component of trustworthy DML protocols. Moreover, as described above, it may also be required to track properties of the data pertaining to its quality, or about the

Term	Definition
Governance	Framework that defines the way data should be handled from the perspectives of access, ownership and auditability.
Privacy	The ability to control how much can be learned from the data about an individual.
Secrecy	Trait of the computation that implies that the sensitive data cannot be seen by anyone other than the data owner.
Security	Property of a protocol or system where the protocol or system as a whole cannot be threatened by an adversarial actor.
Accountability	The ability to track the source of the computation's results.
Verifiability	The capability to guarantee an honest processing of the input data and the integrity of the computation.

Table 1: Summary of the relevant definitions

computation, such as its speed or the result of a verification workflow.

Thus, certain DML implementations (Warnat-Herresthal et al. 2021), propose the utilisation of a permissioned blockchain to track the contributions of each individual data owner, discouraging them from submitting intentionally malformed model updates. Concretely, model aggregation and the selection of the aggregation server for the round (so-called *leader*) occur through the execution of blockchain-backed smart contracts.

In general, the reliance on blockchains in DML can thus allow the federation to obtain a decentralised, immutable transcript of individual contributions to the training protocol. This can include the information about how each individual's contribution influences the resulting model, the time it took them to produce each contribution etc. Such information is essential in the identification of malicious actors whose contributions (or deliberate lack thereof, which can result in protocol halting), affect the utility of the jointly trained model. The immutability of blockchains additionally prevents such actors from concealing their contributions. Moreover, DML protocols combining the notions of data governance and accountability within the same distributed learning system have been developed (Passerat-Palmbach et al. 2020).

While accountability approaches undoubtedly increase the overall trust level, their contribution must be weighed against the cost of developing and deploying such complex solutions. A potential solution for future generations of truly trustless (i.e. those that can function when 50% of all parties are assumed to be malicious) DML systems would be to benefit from leveraging a public blockchain network such as the Ethereum main network. On such truly decentralised public infrastructure, no participant in the federation can tamper with the data stored nor with the execution of the smart contracts.

Conclusion

In this work, we study the vulnerabilities associated with the naïve utilisation of DML. We deduce that most DML protocols cannot be relied upon unless accompanied by additional mechanisms enhancing trust between participants. Unfortunately, much published literature implies that FL and other DML protocols provide privacy protections. It is paramount to note that all that is offered by DML itself is –at best– a *semblance* of privacy. As a result, promoting the utilisa-

tion of DML without any formal measures of input or output privacy protection in place can lead to the disclosure of sensitive information, potentially causing irreparable damage. As evident from the comparison between the components of the ST framework and the current abilities of published DML systems, most frameworks can currently only satisfy a subset of ST requirements. Moreover, we contend that the terms privacy, secrecy and security should not be used interchangeably, and that all are required for ensuring the trustworthiness of DML systems. As noted by (Carlini et al. 2021) "There is no room for error in privacy", showing that the misinterpretation or entanglement of concepts can result in the violation of trust between parties of DML protocols. We summarise these (along with other) definitions in Table 1.

Finally, we propose the following recommendations that arise from our work:

1. Future research in the field of trustworthy AI (particularly in the field of private ML) should agree upon and adhere to terminological guidelines. For instance, neither systems without formal privacy guarantees (e.g. FL) nor systems only offering input privacy (e.g. encryption) should be haphazardly termed *privacy-preserving* (or similar).

2. Only systems adhering to all aforementioned principles should be termed *trustworthy*, to avoid negative consequences associated with over-reliance on protocols that were not designed with fundamental privacy requirements in mind. Such systems should undergo external auditing in order to verify their correctness e.g. through the means of formal network certification (Lecuyer et al. 2019),

3. Most existing DML solutions already offer promising foundations for emergence of private reliable ML systems, but they must be augmented with additional mechanisms in order to be able to guarantee both the privacy of the participants and the robustness of the jointly trained models. We emphasise the importance of education pertaining to these systems, both for experts and laypeople, in this regard.

References

Ali, T.; Nauman, M.; and Jan, S. 2018. Trust in IoT: dynamic remote attestation through efficient behavior capture. *Cluster Computing*, 21(1): 409–421.

Bagdasaryan, E.; and Shmatikov, V. 2020. Blind backdoors in deep learning models. *arXiv preprint arXiv:2005.03823*.

Bagdasaryan, E.; Veit, A.; Hua, Y.; Estrin, D.; and Shmatikov, V. 2020. How to backdoor federated learning. In *International Conference on Artificial Intelligence and Statistics*, 2938–2948. PMLR.

Bitansky, N.; Canetti, R.; Chiesa, A.; Goldwasser, S.; Lin, H.; Rubinstein, A.; and Tromer, E. 2014. The Hunting of the SNARK. Cryptology ePrint Archive, Report 2014/580. https://ia.cr/2014/580.

Bonawitz, K.; Eichner, H.; Grieskamp, W.; Huba, D.; Ingerman, A.; Ivanov, V.; Kiddon, C.; Konečný, J.; Mazzocchi, S.; McMahan, H. B.; et al. 2019. Towards federated learning at scale: System design. *arXiv preprint arXiv:1902.01046*.

Brisimi, T. S.; Chen, R.; Mela, T.; Olshevsky, A.; Paschalidis, I. C.; and Shi, W. 2018. Federated learning of predictive models from federated electronic health records. *International journal of medical informatics*, 112: 59–67.

Carlini, N.; Garg, S.; Jha, S.; Mahloujifar, S.; Mahmoody, M.; and Tramer, F. 2021. NeuraCrypt is not private. *arXiv preprint arXiv:2108.07256*.

Chen, Y.; Luo, F.; Li, T.; Xiang, T.; Liu, Z.; and Li, J. 2020. A training-integrity privacy-preserving federated learning scheme with trusted execution environment. *Information Sciences*, 522: 69–79.

Dwork, C.; McSherry, F.; Nissim, K.; and Smith, A. 2006. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, 265–284. Springer.

Eryurek, E.; Gilad, U.; Lakshmanan, V.; Kibunguchy, A.; Ashdown, J.; and Safari, a. O. M. C. 2021. *Data Governance: The Definitive Guide*. O'Reilly Media, Incorporated. ISBN 9781492063483.

Fang, M.; Cao, X.; Jia, J.; and Gong, N. 2020. Local model poisoning attacks to byzantine-robust federated learning. In *29th {USENIX} Security Symposium ({USENIX} Security 20)*, 1605–1622.

Geiping, J.; Bauermeister, H.; Dröge, H.; and Moeller, M. 2020. Inverting Gradients–How easy is it to break privacy in federated learning? *arXiv preprint arXiv:2003.14053*.

Gilad-Bachrach, R.; Dowlin, N.; Laine, K.; Lauter, K.; Naehrig, M.; and Wernsing, J. 2016. Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy. In *International conference on machine learning*, 201–210. PMLR.

He, Z.; Zhang, T.; and Lee, R. B. 2019. Model inversion attacks against collaborative inference. In *Proceedings of the 35th Annual Computer Security Applications Conference*, 148–162.

Hegedűs, I.; Danner, G.; and Jelasity, M. 2019. Gossip learning as a decentralized alternative to federated learning. In *IFIP International Conference on Distributed Applications and Interoperable Systems*, 74–90. Springer.

Hesamifard, E.; Takabi, H.; and Ghasemi, M. 2017. Cryptodl: Deep neural networks over encrypted data. *arXiv preprint arXiv:1711.05189*.

Kaissis, G.; Ziller, A.; Passerat-Palmbach, J.; Ryffel, T.; Usynin, D.; Trask, A.; Lima, I.; Mancuso, J.; Jungmann, F.; Steinborn, M.-M.; et al. 2021. End-to-end privacy preserving deep learning on multi-institutional medical imaging. *Nature Machine Intelligence*, 3(6): 473–484.

Kaissis, G. A.; Makowski, M. R.; Rückert, D.; and Braren, R. F. 2020. Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, 2(6): 305–311.

Konečný, J.; McMahan, H. B.; Yu, F. X.; Richtárik, P.; Suresh, A. T.; and Bacon, D. 2016. Federated learning: Strategies for improving communication efficiency. *arXiv preprint arXiv:1610.05492*.

Kourou, K.; Exarchos, T. P.; Exarchos, K. P.; Karamouzis, M. V.; and Fotiadis, D. I. 2015. Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal*, 13: 8–17.

Lecuyer, M.; Atlidakis, V.; Geambasu, R.; Hsu, D.; and Jana, S. 2019. Certified robustness to adversarial examples with differential privacy. In *2019 IEEE Symposium on Security and Privacy (SP)*, 656–672. IEEE.

Lee, D.; Kohlbrenner, D.; Shinde, S.; Asanović, K.; and Song, D. 2020a. Keystone: An open framework for architecting trusted execution environments. In *Proceedings of the Fifteenth European Conference on Computer Systems*, 1–16.

Lee, S.; Ko, H.; Kim, J.; and Oh, H. 2020b. vCNN: Verifiable Convolutional Neural Network based on zk-SNARKs. Technical report, Cryptology ePrint Archive, Report 2020/584. https://eprint. iacr. org/2020/584.

Mohassel, P.; and Zhang, Y. 2017. Secureml: A system for scalable privacy-preserving machine learning. In *2017 IEEE symposium on security and privacy (SP)*, 19–38. IEEE.

Nasirigerdeh, R.; Torkzadehmahani, R.; Matschinske, J.; Baumbach, J.; Rueckert, D.; and Kaissis, G. 2021. HyFed: A Hybrid Federated Framework for Privacy-preserving Machine Learning. *arXiv preprint arXiv:2105.10545*.

Nasirigerdeh, R.; Torkzadehmahani, R.; Matschinske, J.; Frisch, T.; List, M.; Späth, J.; Weiß, S.; Völker, U.; Heider, D.; Wenke, N. K.; et al. 2020. sPLINK: a federated, privacy-preserving tool as a robust alternative to meta-analysis in genome-wide association studies. *BioRxiv*.

Passerat-Palmbach, J.; Farnan, T.; McCoy, M.; Harris, J. D.; Manion, S. T.; Flannery, H. L.; and Gleim, B. 2020. Blockchain-orchestrated machine learning for privacy preserving federated learning in electronic health data. In *2020 IEEE International Conference on Blockchain (Blockchain)*, 550–555. IEEE.

Radley-Gardner, O.; Beale, H.; and Zimmermann, R., eds. 2016. *Fundamental Texts On European Private Law*. Hart Publishing. ISBN 978-1-78225-864-3 978-1-78225-865-0 978-1-78225-866-7 978-1-78225-867-4.

Rouhani, B. D.; Riazi, M. S.; and Koushanfar, F. 2018. Deepsecure: Scalable provably-secure deep learning. In *Proceedings of the 55th Annual Design Automation Conference*, 1–6.

Roy, A. G.; Siddiqui, S.; Pölsterl, S.; Navab, N.; and Wachinger, C. 2019. Braintorrent: A peer-to-peer environment for decentralized federated learning. *arXiv preprint arXiv:1905.06731*.

Shokri, R.; Stronati, M.; Song, C.; and Shmatikov, V. 2017. Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy (SP)*, 3–18. IEEE.

Trask, A.; Bluemke, E.; Garfinkel, B.; Cuervas-Mons, C. G.; and Dafoe, A. 2020. Beyond Privacy Trade-offs with Structured Transparency. *arXiv preprint arXiv:2012.08347*.

Usynin, D.; Ziller, A.; Makowski, M.; Braren, R.; Rueckert, D.; Glocker, B.; Kaissis, G.; and Passerat-Palmbach, J. 2021. Adversarial interference and its mitigations in privacy-preserving collaborative machine learning. *Nature Machine Intelligence*, 3(9): 749–758.

Vanneschi, L.; Farinaccio, A.; Mauri, G.; Antoniotti, M.; Provero, P.; and Giacobini, M. 2011. A comparison of machine learning techniques for survival prediction in breast cancer. *BioData mining*, 4(1): 1–13.

Vepakomma, P.; Gupta, O.; Swedish, T.; and Raskar, R. 2018. Split learning for health: Distributed deep learning without sharing raw patient data. *arXiv preprint arXiv:1812.00564*.

Warnat-Herresthal, S.; Schultze, H.; Shastry, K. L.; Manamohan, S.; Mukherjee, S.; Garg, V.; Sarveswara, R.; Händler, K.; Pickkers, P.; Aziz, N. A.; et al. 2021. Swarm Learning for decentralized and confidential clinical machine learning. *Nature*, 594(7862): 265–270.

Weng, C.; Yang, K.; Xie, X.; Katz, J.; and Wang, X. 2021. Mystique: Efficient Conversions for Zero-Knowledge Proofs with Applications to Machine Learning. In *30th {USENIX} Security Symposium ({USENIX} Security 21)*, 501–518.

Yan, L.; Zhang, H.; Goncalves, J.; Xiao, Y.; Wang, M.; Guo, Y.; Sun, C.; Tang, X.; Jin, L.; Zhang, M.; et al. 2020. A machine learning-based model for survival prediction in patients with severe COVID-19 infection.

Yang, C.; Wu, Q.; Li, H.; and Chen, Y. 2017. Generative poisoning attack method against neural networks. *arXiv preprint arXiv:1703.01340*.

Zhang, Y.; Jia, R.; Pei, H.; Wang, W.; Li, B.; and Song, D. 2020. The secret revealer: Generative model-inversion attacks against deep neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 253–261.

Zhao, Y.; Li, M.; Lai, L.; Suda, N.; Civin, D.; and Chandra, V. 2018. Federated learning with non-iid data. *arXiv preprint arXiv:1806.00582*.

Ziller, A.; Usynin, D.; Remerscheid, N.; Knolle, M.; Makowski, M.; Braren, R.; Rueckert, D.; and Kaissis, G. 2021. Differentially private federated deep learning for multi-site medical image segmentation. *arXiv preprint arXiv:2107.02586*.