Secure distributed machine learning in healthcare: a study on FAIR, compliance and cybersecurity for federated learning

Plug, R.B.F.

6 Discussion

> *Simplicity is a great virtue but it requires hard work to achieve it and education to appreciate it.*
>
> **— Edsger W. Dijkstra**
> *On the nature of Computing Science*

This dissertation consists of four incremental studies that focus on the application of secure and privacy-preserving machine learning in cross-border clinical use cases. This resulted in the development of the secure distributed machine learning architecture, which is a secure variant of federated learning [1]. The overall focus of this dissertation is on FAIR [2, 3, 4, 5, 6, 7], regulatory compliance [8, 9, 10], and cybersecurity [11, 12, 13] for epidemiological healthcare data across the four studies. **Chapter 2** introduced a conceptual framework for health FAIR (Findable, Accessible, Interoperable, Reusable) data management [2], defining key terminologies and FAIR data stewardship principles for the virus outbreak data network (VODAN) [14]. **Chapter 3** built on this foundation by designing an architecture for fully FAIR at point of care and GDPR-compliant data infrastructure, demonstrating how local data stewardship [15] and a data visiting [16] approach enable federated analytics for infectious disease outbreak monitoring in VODAN without compromising data ownership or privacy [7]. Going from the high level view of **Chapter 3** to a more granular technical design in **Chapter 4**, focus is put on technical challenges such as risks to security and privacy of patients and healthcare institutions. In this chapter the Secure Distributed Machine Learning (SDML) architecture is proposed that combines concepts from federated learning with cryptography [17] and secret sharing

[18] from Secure Multi-Party Computation (SMPC) [19] to enhance privacy and cyber-resilience in edge computing environments found in distributed healthcare use cases [20, 21] for epidemiology. Concluding, in **Chapter 5** the performance of the methodology presented in **Chapter 4** is benchmarked for distributed automated diagnostics of chest x-rays for diseases such as viral pneumonia, measuring the overhead of homomorphic encryption and secret sharing on a ResNet18 TinyML model [22] to quantify the trade-offs between security and training efficiency in medical imaging across clinics.

6.1 Main Findings

FAIR stewardship is a necessity for federated analytics in epidemiology

The first study in **Chapter 2** established the importance of standardising health data management and stewardship through a common terminology framework aligned with the FAIR principles [23, 24, 15]. It identified key challenges in current health information systems such as unclear data ownership and the problem of heterogeneous data standards across clinics. This was especially apparent during the SARS-CoV-2 pandemic and the 2013-2016 Ebola epidemic, where fragmented data slowed clinical response and observability for healthcare organisations. These challenges are addressed in **Chapter 2** with concepts of FAIR and FAIR Data Points (FDPs) [4, 25].

A notable finding is that FDPs coupled with data stewardship is a necessary combination when implementing concepts of federated analytics and data visiting. Data visiting replaces the process of data transfer, allowing only aggregate or computed outputs that do not contain any personal or sensitive data to leave a health facility [16]. This is a form of federated analytics that safeguards personal data. Data stewards are needed to guide the complex (meta)data creation, curation [26] and management process inside health facilities to

properly index and annotate data. This can be seen as an increase in overhead, but given that this increased potential for findability, interoperability and reusability of data, in the long term it likely saves both time and resources [27, 15]. Only if metadata are properly presented, can we identify and match data across FDPs for use in federated analytics, as this requires either horizontal or vertical data alignment [28, 29].

A second key insight is that the use of community-defined ontologies encourages standardisation at point of clinical use [30, 31]. This requires the expertise of data stewardship to implement. Standardised data contain rich metadata that is used to support both clinical use of data and machine actionability of data produced in a health facility. The resulting data is graph data in the form of triples [32], which are knowledge representations that enable automated relationship discovery without compromising data integrity. Additionally in **Chapter 2** it was found that implementing ontologies and shared vocabularies can greatly improve data provenance in healthcare as well, and doing so is one of the prerequisites for the deployment of a FAIR implementation network [33].

FAIR and GDPR compliance are complementary for healthcare data

Chapter 3 shows that the FAIR data principles and GDPR regulations can be complementary to each other. The concept of data stewardship introduced in **Chapter 2** demonstrated that FAIR data stewardship is one of the key building blocks for federated analytics [34, 15, 35], but is also essential in making sure data management practices for FL can align with GDPR [10]. This combination of FAIR data stewardship and GDPR-based architecture can be used as a legal compliance framework for the use and reuse of clinical health data.

By placing data controllers and processors at the point of care, full data ownership and curation is retained at the local level [26]. This

is essentially data sovereignty [36] as personal data never leaves a health facility for such analytical purposes. This type of architecture enables federated analytics through a structured process of data visiting, where only anonymised or aggregated data is shared externally [16]. This process is subject to data use agreements and audits by the data steward. In addition to FAIR data stewardship, **Chapter 3** introduces as one of the key findings a framework that includes FAIR data creation at origin and the whole chain of data controller and processor all in-residence.

The importance of well-trained FAIR data stewards was demonstrated in VODAN. A high level of knowledge is required to configure, operate and curate a federated infrastructure [26]. Within VODAN thirty data stewards were trained through a structured Training of Trainers (ToT) program. These stewards manage the FAIRification process, oversee metadata curation, and ensure that all data access requests conform to GDPR standards. Their involvement also strengthens local data governance [15, 37] and builds trust within facilities.

To facilitate this FAIR metadata templates were defined based on the WHO SARS-CoV-2 eCRF standard [38, 30] and were structured through localised instances of CEDAR [39]. The clinical data entry forms used in clinics were based on these FAIR templates, resulting in data that was being generated in FAIR format as RDF graph data [40, 41]. This ensured semantic consistency across the facilities in the participating healthcare facilities from eight different countries including Ethiopia, Kenya, Nigeria, Somalia, Tanzania, Uganda, Tunisia, and Zimbabwe. The first cross-border query, executed on 29 September 2020 between Uganda and the Netherlands, as described in **Chapter 3**, demonstrated that this FAIR and GDPR based architecture supports cross-border federated analytics without exposing medical data to external parties [25].

SMPC techniques add necessary security for applying FL in healthcare

In the SDML methodology presented by **Chapter 4** we show that integrating encryption and secret sharing inspired by SMPC [19, 42] into FL enables secure aggregation of model updates without risk of exposing individual contributions in the form of gradients. Local gradients are encrypted using homomorphic encryption [43] and split via secret sharing before transmission [18, 44]. Aggregation servers process only partial, encrypted shares, preventing access to any single client's gradient from which information could be extracted. This enforces a strict zero-trust policy [45]. No single party such as edge nodes or aggregators can observe individual intermediate computations [46].

The main insight from this study is that this design ensures zero-knowledge computation throughout the training process. By structurally decoupling local computation, secure aggregation, and global model updates, the architecture addresses specific attack vectors targetting gradient reconstruction [47, 48] and maintains confidentiality even in adversarial settings. This is fundamentally more secure when compared to plain FL training methods, which requires individual parties to trust other parties with their intermediate computations [49].

Another key takeaway from **Chapter 4** is the role of localised computation in the efficiency of performing machine learning in resource-constrained environments. Gradient computation remains entirely at the edge with computational load spread out over multiple smaller datasets, rather than one large central dataset. This avoids the need to offload most of the computationally heavy processing to central servers. This is a strategy similar to the one used by AlphaFold for protein research [50, 51]. In the African healthcare landscape of VODAN, this enables learning from sensitive data without violating sovereignty or overburdening limited and potentially unreliable local infrastructure. The decoupled architecture inspired by techniques from edge

computing [52, 53] reduces the central coordination overhead and allows asynchronous, secure participation from heterogeneous clients.

Secure FL for collaborative diagnostic support is empirically feasible

The final study in **Chapter 5** performed an empirical study through simulation of privacy-preserving techniques in a federated learning setting using medical imaging data. It benchmarked a TinyML-scale ResNet18 model [22] trained on the benchmark chest X-ray dataset of MedMNIST [54, 55] with and without encryption of gradients and secret sharing of keys.

The empirical study results presented in **Chapter 5** demonstrate that incorporating secure aggregation through homomorphic encryption and key-based secret sharing in federated learning does not impair model performance in a binary classification task on medical imaging. Across ten simulated hospital clients the secure and non-secure variants yielded indistinguishable test with ROC-AUC of 0.801 annd 0.802 respectively. In an infectious-disease triage setting this translates to less than 0.2 difference in false-negative rate per 1 000 chest x-ray examinations, which is clinically indistinguishable. Variability in these metrics across client counts K was minor and within the bounds expected from stochastic training effects. Observed instability in F_1-score at higher client counts ($K > 6$) was similar in both non-secure and secure setups and was found to be due to data limitations, as each client retains a smaller portion of the dataset when split equally among all clients. Importantly, the ROC AUC values remained stable throughout, indicating that the precision and recall ratio for a static classification threshold remained similar even when fewer samples per client were available.

The key finding from this study is that the computational overhead introduced by the secure configuration scales linearly with the number of clients K, increasing from 11.1% at $K = 1$ to 25.3% at $K = 10$, with

a mean overhead of 19.47% $\pm$ 4.93%. This scaling was consistent with the overhead of additive HE gradient aggregation. The per-client encryption cost remained stable at $\sim$14.9 seconds across all tested configurations, indicating that overhead is dependent on model size rather than data volume. The secure learning loop was implemented without simulating network latency, allowing precise attribution of overhead to encryption and secret sharing processes. This confirmed that secure FL, a simplified variant of SDML, is practically feasible with a predictable impact on computational resources, while providing additional protection in privacy-sensitive domains such healthcare, particularly when rapid action is required such as on cross-regional and cross-country diagnostic data on infectious diseases.

6.2 Future Research

The research performed in this dissertation across the four publications is an incremental exploration all the way from the basis of the FAIR data principles, to compliance and ultimately applications in cyber security. We gradually used methods from design science research [56] to design a methodology for privacy-preserving machine learning, which we empirically tested. The strength of this incremental research approach is that the research directions, and the resulting articles, gradually shifted from a very broad theoretic focus to an increasingly specific and application-oriented direction.

In retrospect, this narrowing-down approach is also the main limitation of this study. Since the scope of research was initially so broad, it was infeasible to combine and empirically validate all the different combinations of concepts and techniques that we explored. For instance in the FAIR-focused contributions **Chapter 2-3** we provide an extensive conceptual framework but only empirically validate a limited fraction of the whole framework in this dissertation. At this point, the research direction had shifted to a very practical application

that is a result from this theoretic work: federated learning and the practical challenges involved from a privacy-preservation and security standpoint. This meant that FAIR, while implicitly present in the concept of FL and the framework of **Chapter 4**, wasn't an explicit research goal for the empirical section in **Chapter 5** of this dissertation. These limitations are opportunities for future research, which are provided in this section.

FAIR integration in federated learning

A promising research direction is the integration of the FAIR principles into the secure federated learning architecture presented in **Chapter 4** and validated in **Chapter 5**. **Chapters 2-3** have shown that standardizing data through rich metadata enables machine-actionable cross-clinic data sharing. Further research can aim to demonstrate the application of secure FL methodologies on FAIR data, which was initially explored by Sinaci et al. [57]. This could be further extended to a more complex architecture including FDPs as FAIR data repositories. The research scope of this not only covers enabling FL to be able to train and perform inference on FAIR data, but also developing methods to ensure that FL model configurations, weights and inferences, are FAIR as well.

We can design ML model metadata schemas to describe the architecture, hyper-parameter configuration, training metadata and model parameters of federated models in a standardised way as artifacts [58]. This could enable automated collaboration and discovery of models. In the ideal case this means not only using FAIR data, but also making sure that data describing a model is FAIR as well. Interoperability is the key aspect: if different locales use common ontologies and data standards, then federated models can be set up for shared analyses (**Chapter 3**) with potential for transfer-learning with fine-tuned to optimise models locally, while learning globally [59].

To scale this concept, the use and standardised infrastructure in the form of FDPs is key. FL as a distributed machine learning method can only be ever as good as its infrastructure. And by combining FAIR and FL together via FDPs, this approach can provide provenance and data lineage with rich metadata describing each training round and complete dataset. It also helps ensure compliance, since FAIR metadata contains information on accessibility, such as on personal data. Ultimately this allows us to curate data and models, which leads to more reliable and trustworthy machine learning. And it is exactly this that is required for communities that participate in FL to build trust.

SOLID pods for federated learning

Chapters 2-3 demonstrated the value of keeping health data at the point of origin using the concept of data ownership and data sovereignty. We can take this concept a step further by linking it to SOLID data pods [60, 61]. This is fundamentally different from an approach that uses FDPs, as those store data at the institutional level. SOLID pods on the other hand logically organise data storage at the personal level, implementing the concept of the personal data vault [62]. Individuals retain full ownership over their personal data.

Combining the concept of SOLID with FL could allow patients to contribute to health models directly by allowing use of their personal data under specific approved conditions, without having their personal data stored at a third party. This would enable individuals to donate the use of their data for scientific research. This aligns well with the principle of data sovereignty, not just at the institutional level but at an individual level.

Many of the existing challenges of FDPs translate directly to SOLID. The new challenge specific to SOLID is the difference in scale. The SOLID implementation would need to be able to support the computational load and communication cost of FL at very large client

counts. In addition, new security measures will be required to handle authentication, data-use verification (access and control) and client coordination in the massively decentralised network of data stations. In **Chapter 3-4** we already laid some groundwork for this in terms of security, but the scope of concerns in the case of SOLID is much larger.

Emerging machine learning research

There are additional emerging topics related to machine learning that could be applied to FL. One such topic is explainable AI [63]. The use of ML in FL requires the results from such models to be explainable and interpretable for clinicians to understand why certain predictions or classifications are made by a model. Further research could investigate the use of existing techniques from machine learning and apply them to FL. For example being able to interpret a globally learned model in terms of local data patterns while preserving privacy.

Another area of research could be further investigation into scalability and efficiency for secure FL. **Chapter 5** had shown the overhead introduced by gradient encryption and secret sharing. Follow-up research may focus on optimising the cryptography step, such as more efficient additive HE instead of FHE [43] or the use of model update compression techniques [64]. Additional techniques from ML like model quantisation [65], pruning [66], or distillation [67] could be adapted to FL to create smaller and faster models that provide non-inferior performance in clinical edge computing application.

References

[1] H.B. McMahan, E. Moore, D. Ramage, S. Hampson, and B.A. Arcas. „Communication - Efficient Learning of Deep Networks from Decentralized Data“. In: *International Conference on Artificial Intelligence and Statistics*. 2016 (cit. on p. 163).

[2] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, et al. „The FAIR Guiding Principles for scientific data management and stewardship“. In: *Scientific Data* 3 (2016), p. 160018 (cit. on p. 163).

[3] Annika Jacobsen, Ricardo de Miranda Azevedo, Nick S. Juty, et al. „FAIR Principles: Interpretations and Implementation Considerations“. In: *Data Intelligence* 2 (2020), pp. 10–29 (cit. on p. 163).

[4] Barend Mons, Cameron Neylon, Jan Velterop, et al. „Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud“. In: *Inf. Serv. Use* 37 (2017), pp. 49–56 (cit. on pp. 163, 164).

[5] M.V. Reisen, M. Stokmans, M. Basajja, et al. „Towards the tipping point for FAIR implementation“. In: *Data Intelligence* 2 (2020), pp. 264–275 (cit. on p. 163).

[6] Mirjam van Reisen, Mia Stokmans, Munyaradzi Mawere, et al. „FAIR practices in Africa“. In: *Data Intelligence* 2.1–2 (2020), pp. 246–256 (cit. on p. 163).

[7] Mirjam van Reisen, Francisca Onaolapo Oladipo, Mia Stokmans, et al. „Design of a FAIR digital data health infrastructure in Africa for COVID-19 reporting and research“. In: *Advanced Genetics (Hoboken, N.j.)* 2 (2021) (cit. on p. 163).

[8] M. Hintze. „Viewing the GDPR through a de-identification lens: A tool for compliance, clarification and consistency“. In: *International Data Privacy Law* 8.1 (2018) (cit. on p. 163).

[9] Rocco de Filippis, Abdullah Al Foysal, Vincenzo Rocco, et al. „The risk perspective of AI in healthcare: GDPR and GELSI framework (Governance, Ethical, Legal and Social Implications) and the new European AI Act“. In: *Italian Journal of Psychiatry* 10.1 (2024) (cit. on p. 163).

[10] Nguyen Truong, Kai Sun, Siyao Wang, Florian Guitton, and Yike Guo. „Privacy Preservation in Federated Learning: An Insightful Survey from the GDPR Perspective“. In: *Computers & Security* 110 (2021), p. 102402 (cit. on pp. 163, 165).

[11] Pengrui Liu, Xiangrui Xu, and Wei Wang. „Threats, Attacks and Defenses to Federated Learning: Issues, Taxonomy and Perspectives“. In: *Cybersecurity* 5.4 (2022) (cit. on p. 163).

[12] P. Mohassel and Y. Zhang. „SecureML: A System for Scalable Privacy-Preserving Machine Learning“. In: *Proc. 2017 IEEE Symp. Secur. Priv.* (2017), pp. 19–38 (cit. on p. 163).

[13] P.H. Jati, M. van Reisen, E. Flikkenschild, et al. „Data Access, Control, and Privacy Protection in the VODAN-Africa Architecture“. In: *Data Intelligence* 4 (2022), pp. 938–954 (cit. on p. 163).

[14] M. van Reisen, F.O. Oladipo, M. Mpezamihigo, et al. „Incomplete COVID-19 Data: The Curation of Medical Health Data by the Virus Outbreak Data Network-Africa“. In: *Data Intelligence* 4 (2022), pp. 673–697 (cit. on p. 163).

[15] Barend Mons. *Data Stewardship for Open Science: Implementing FAIR Principles*. Chapman and Hall/CRC, 2018, p. 244. ISBN: 978-1032095707 (cit. on pp. 163–166).

[16] Samson Yohannes Amare, Araya Abrha Medhanyie, and Mirjam van Reisen. „Data Visiting in Digital Black Holes: FAIR Based Digital Health Innovation during War“. In: *Tigray. War in a Digital Black Hole. Book 3*. Final draft. Langaa RPCIG, 2024, pp. 477–508. ISBN: 9789956554188 (cit. on pp. 163, 164, 166).

[17] Abbas Acar, Hidayet Aksu, A. Selcuk Uluagac, and Mauro Conti. „A Survey on Homomorphic Encryption Schemes: Theory and Implementation“. In: *ACM Computing Surveys* 51.4 (2018), pp. 1–35 (cit. on p. 163).

[18] Arup Kumar Chattopadhyay, Sanchita Saha, Amitava Nag, and Sukumar Nandi. „Secret Sharing: A Comprehensive Survey, Taxonomy and Applications“. In: *Computer Science Review* 51 (2024), p. 100608 (cit. on pp. 164, 167).

[19] C. Zhao, S. Zhao, M. Zhao, et al. „Secure Multi-Party Computation: Theory, practice and applications“. In: *Information Sciences* 476 (2019), pp. 357–372 (cit. on pp. 164, 167).

[20] L. Li, Y. Fan, M. Tse, and K. Lin. „A Review of Applications and Challenges of Federated Learning in Healthcare“. In: *Elsevier Computers & Industrial Engineering* 24.10 (2020), pp. 350–359 (cit. on p. 164).

[21] Jie Xu, Benjamin S. Glicksberg, Chang Su, et al. „Federated Learning for Healthcare Informatics“. In: *Journal of Healthcare Informatics Research* 5 (2021), pp. 1–19 (cit. on p. 164).

[22] K. He, X. Zhang, S. Ren, and J. Sun. „Deep Residual Learning for Image Recognition“. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778 (cit. on pp. 164, 168).

[23] Lynn M. Schriml, Maria Chuvochina, Neil Davies, et al. „COVID-19 pandemic reveals the peril of ignoring metadata standards“. In: *Scientific Data* 7 (2020) (cit. on p. 164).

[24] Ching-Heng Lin, Nai-Yuan Wu, and Der-Ming Liou. „A multi-technique approach to bridge electronic case report form design and data standard adoption“. In: *Journal of biomedical informatics* 53 (2015), pp. 49–57 (cit. on p. 164).

[25] M. Basajja, M. Suchanek, G.T. Taye, et al. „Proof of Concept and Horizons on Deployment of FAIR Data Points in the COVID-19 Pandemic“. In: *Data Intelligence* 4.4 (2022), pp. 917–937 (cit. on pp. 164, 166).

[26] M. van Reisen, S.Y. Amare, R. Plug, et al. „Curation of Federated Patient Data: A Proposed Landscape for the African Health Data Space“. In: *Federated Learning for Digital Healthcare Systems*. Intelligent Data-Centric Systems. Elsevier, 2024. Chap. 3, pp. 59–80 (cit. on pp. 164–166).

[27] Alicia Martínez-García, Celia Alvarez-Romero, Esther Román-Villaran, Máximo Bernabeu-Wittel, and Carlos Luis Parra-Calderón. „FAIR principles to improve the impact on health research management outcomes“. In: *Heliyon* 9.5 (2023), e15733 (cit. on p. 165).

[28] Lan Zhang, Anran Li, Hongyi Peng, et al. „Privacy-Preserving Data Selection for Horizontal and Vertical Federated Learning“. In: *IEEE Transactions on Parallel and Distributed Systems* 35.11 (2024), pp. 2054–2068 (cit. on p. 165).

[29] Afsana Khan, Marijn ten Thij, and Anna Wilbik. „Vertical federated learning: a structured literature review“. In: *Knowledge and Information Systems* 67.2 (2025), pp. 3205–3243 (cit. on p. 165).

[30] Patricia L. Whetzel, Natasha Noy, Nigam Haresh Shah, et al. „BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications“. In: *Nucleic Acids Research* 39 (2011), W541–W545 (cit. on pp. 165, 166).

[31] M.A. Sicilia. „Metadata, semantics, and ontology: Providing meaning to information resources“. In: *International Journal of Metadata, Semantics and Ontologies* 1.1 (2006), pp. 83–86 (cit. on p. 165).

[32] P. Heim, S. Hellmann, J. Lehmann, S. Lohmann, and T. Stegemann. „RelFinder: Revealing relationships in RDF knowledge bases“. In: *Semantic Multimedia. Lecture Notes in Computer Science* 5887 (2009) (cit. on p. 165).

[33] B. Mons. „The VODAN IN: Support of a FAIR-based infrastructure for COVID-19“. In: *European Journal of Human Genetics* 28 (2020), pp. 724–727 (cit. on p. 165).

[34] M. Reisen, F. Oladipo, M. Stokmans, et al. „Design of a FAIR digital data health infrastructure in Africa for COVID-19 reporting and research“. In: *Advanced Genetics* 2.2 (2021) (cit. on p. 165).

[35] Robert Pergl, Rob W.W. Hooft, M. Suchánek, Vojtech Knaisl, and Jan Slifka. „"Data Stewardship Wizard": A Tool Bringing Together Researchers, Data Stewards, and Data Experts around Data Management Planning“. In: *Data Sci. J.* 18 (2019), p. 59 (cit. on p. 165).

[36] M. Jarke, B. Otto, and S. Ram. „Data sovereignty and data space ecosystems“. In: *Business and Information Systems Engineering* 61 (2019), pp. 549–550 (cit. on p. 166).

[37] J. Winter and E. Davidson. „Big data governance of personal health information and challenges to contextual integrity“. In: *The Information Society* 35 (2019), pp. 36–51 (cit. on p. 166).

[38] Luiz Bonino. *WHO COVID-19 Rapid Version CRF semantic data model* (cit. on p. 166).

[39] Rafael S Gonçalves, M. O'Connor, M. M. Romero, et al. „The CEDAR Workbench: An Ontology-Assisted Environment for Authoring Metadata that Describe Scientific Experiments“. In: *The semantic Web–ISWC: International Semantic Web Conference proceedings. International Semantic Web Conference* 10588 (2017), pp. 103–110 (cit. on p. 166).

[40] N. Gibbins and N. Shadbolt. „Resource Description Framework (RDF)“. In: *Intelligence, Agents, Multimedia Group, University of Southampton* (2009) (cit. on p. 166).

[41] R.S. Gonçalves, M. O'Connor, M.M. Romero, et al. „The CEDAR Workbench: An ontology-assisted environment for authoring metadata that describe scientific experiments“. In: *The Semantic Web – ISWC 2017. Lecture Notes in Computer Science* 10588 (2017), pp. 103–110 (cit. on p. 166).

[42] David Evans, Vladimir Kolesnikov, and Mike Rosulek. *A Pragmatic Introduction to Secure Multi-Party Computation*. Vol. 2. Foundations and Trends in Privacy and Security 2–3. Now Publishers, 2018, pp. 70–246 (cit. on p. 167).

[43] L. T. Phong, Y. Aono, T. Hayashi, L. Wang, and S. Moriai. „Privacy-Preserving Deep Learning via Additively Homomorphic Encryption“. In: *IEEE Transactions on Information Forensics and Security* 13.5 (2018), pp. 1333–1345 (cit. on pp. 167, 172).

[44] Elette Boyle, Niv Gilboa, and Yuval Ishai. „Function Secret Sharing“. In: *Advances in Cryptology*. Vol. 9057. Lecture Notes in Computer Science. Springer, 2015, pp. 337–367 (cit. on p. 167).

[45] Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky, and Amit Sahai. „Zero-Knowledge from Secure Multiparty Computation“. In: *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*. 2007, pp. 21–30 (cit. on p. 167).

[46] Wirawan Agahari, Hosea Ofe, and Mark de Reuver. „It is not (only) about privacy: How multi-party computation redefines control, trust, and risk in data sharing“. In: *Electronic Markets* 32 (2022), pp. 1577–1602 (cit. on p. 167).

[47] L. Zhu, Z. Liu, and S. Han. „Deep Leakage from Gradients“. In: *Advances in Neural Information Processing Systems*. Vol. 32. 2019 (cit. on p. 167).

[48] Rui Zhang, Song Guo, and Ping Li. „GradFilt: Class-wise Targeted Data Reconstruction from Gradients in Federated Learning“. In: *Proceedings of the 2024 ACM Asia Conference on Computer and Communications Security (AsiaCCS)*. 2024 (cit. on p. 167).

[49] Arjun Nitin Bhagoji, Supriyo Chakraborty, Prateek Mittal, and Seraphin Calo. „Analyzing Federated Learning through an Adversarial Lens“. In: *Proceedings of the 36th International Conference on Machine Learning*. Vol. 97. Proceedings of Machine Learning Research. PMLR, 2019, pp. 634–643 (cit. on p. 167).

[50] John Jumper, Richard Evans, Alexander Pritzel, et al. „Highly accurate protein structure prediction with AlphaFold“. In: *Nature* 596.7873 (2021), pp. 583–589 (cit. on p. 167).

[51] Hyun Park, Parth Patel, Roland Haas, and E. A. Huerta. „APACE: AlphaFold2 and advanced computing as a service for accelerated discovery in biophysics“. In: *Proceedings of the National Academy of Sciences of the United States of America* 121.27 (2024), e2311888121 (cit. on p. 167).

[52] Z. Zhou, X. Chen, E. Li, et al. „Edge Intelligence: Paving the Last Mile of Artificial Intelligence with Edge Computing“. In: *Proceedings of the IEEE* 107 (2019), pp. 1738–1762 (cit. on p. 168).

[53] Hongliang Zhou, Yifeng Zheng, and Xiaohua Jia. „Towards Robust and Privacy-Preserving Federated Learning in Edge Computing“. In: *Computer Networks* 243 (2024), p. 110321 (cit. on p. 168).

[54] Xiaosong Wang, Yifan Peng, Le Lu, et al. „ChestX-ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases“. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017, pp. 3462–3471 (cit. on p. 168).

[55] J. Yang, R. Shi, and B. Ni. „MedMNIST Classification Decathlon: A Lightweight AutoML Benchmark for Medical Image Analysis“. In: *IEEE International Symposium on Biomedical Imaging (ISBI)*. 2021, pp. 191–195 (cit. on p. 168).

[56] Jan vom Brocke, Alan Hevner, and Alexander Maedche. „Introduction to Design Science Research“. In: *Design Science Research. Cases*. Progress in IS. Springer, Cham, 2020, pp. 1–13 (cit. on p. 169).

[57] A. Anil Sinaci, Mert Gencturk, Celia Alvarez-Romero, et al. „Privacy-preserving federated machine learning on FAIR health data: A real-world application“. In: *Computational and Structural Biotechnology Journal* 24 (2024), pp. 136–145 (cit. on p. 170).

[58] Jian Qin and Bingyu Yu. „Metadata in Trustworthy AI: From Data Quality to ML Modeling“. In: *Proceedings of the International Conference on Dublin Core and Metadata Applications*. 2023 (cit. on p. 170).

[59] Wei Guo, Fuzhen Zhuang, Xiao Zhang, Yiqi Tong, and Jin Dong. „A comprehensive survey of federated transfer learning: challenges, methods and applications“. In: *Frontiers of Computer Science* 18.6 (2024), p. 186356 (cit. on p. 170).

[60] Christian Esposito, Ross Horne, Livio Robaldo, Bart Buelens, and Elfi Goesaert. „Assessing the Solid Protocol in Relation to Security and Privacy Obligations“. In: *Information* 14.7 (2023), p. 411 (cit. on p. 171).

[61] Mohamed Ragab, Yury Savateev, Helen Oliver, et al. „A Demonstration of Decentralized Search Over Solid Personal Online Datastores“. In: *Companion Proceedings of the ACM Web Conference 2024*. 2024, pp. 1055–1058 (cit. on p. 171).

[62] Sofie Verbrugge, Frederic Vannieuwenborg, Marlies Van der Wee, et al. „Towards a personal data vault society: an interplay between technological and business perspectives“. In: *2021 60th FITCE Communication Days Congress for ICT Professionals: Industrial Data – Cloud, Low Latency and Privacy (FITCE)*. IEEE, 2021, pp. 1–6 (cit. on p. 171).

[63] Sajid Ali, Tamer Abuhmed, Shaker El-Sappagh, et al. „Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence“. In: *Information Fusion* 99 (2023), p. 101805 (cit. on p. 172).

[64] Bo Chen, Ali Bakhshi, Gustavo Batista, Brian Ng, and Tat-Jun Chin. „Update Compression for Deep Neural Networks on the Edge“. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2022, pp. 3075–3085 (cit. on p. 172).

[65] Benoit Jacob, Skirmantas Kligys, Bo Chen, et al. „Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference“. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018, pp. 2704–2713 (cit. on p. 172).

[66] Andy Li, Milan Markovic, Peter Edwards, and Georgios Leontidis. „Model pruning enables localized and efficient federated learning for yield forecasting and data sharing“. In: *Expert Systems with Applications* 242 (2024), p. 122847 (cit. on p. 172).

[67] F. MohiEldeen Alabbasy, A. S. Abohamama, and Mohammed F. Alrahmawy. „Compressing medical deep neural network models for edge devices using knowledge distillation“. In: *Journal of King Saud University - Computer and Information Sciences* 35.7 (2023), p. 101616 (cit. on p. 172).