Navigating Risks and Rewards of Generative Model-based Synthetic Datasets: A Regulatory Perspective

Debalina Padariya [1] **Isabel Wagner** [2] **Aboozar Taherkhani** [1] **Eerke Boiten** [1]

Abstract

The rapid development of generative AI models has gained groundbreaking attention. This article looks into the practical challenges of generative model-based synthetic datasets with an intersection of ethical considerations inherent to this field. These challenges include privacy attacks and limitations in existing privacy-preserving approaches. We also highlight future research directions that foster fair and responsible use of synthetic data while ensuring ethical oversight in the landscape of generative AI.

1. Introduction

The field of generative AI has emerged with a transformative leap in scientific exploration and commercial technologies, such as image recognition, natural language processing, Drug Discovery, music/video generation, product design, and many more (Feuerriegel et al., 2024). While Big tech giants, such as Apple, Microsoft, Google, Meta, and OpenAI, compete to accelerate generative AI to a central position (Khanal et al., 2024), several privacy breaches undermine trust in these advancements (Golda et al., 2024).

Synthetic data generation (SDG) is one of the emerging use cases of generative AI and has made significant progress as a privacy-enhancing technology (Bellovin et al., 2019). SDG aims to closely resemble real-world data, maintaining data privacy while preserving sufficient usefulness for future purposes. There are various methods for creating synthetic data with machine learning-based models, such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), statistical-based Gaussian Copula, transformer-based and agent-based models, or other ML-based methods (Lu et al., 2024). In this article, we primarily focus on the most popular deep generative models, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), which show remarkable performance in producing high-quality realistic synthetic samples by learning complex data distributions in high dimensions (Mendes et al., 2023).

An important SDG consideration is its reliability in the current regulatory landscape. The potential privacy attacks associated with generative models emerge as critical issues, i.e., re-identification risks of synthetic data (Stadler et al., 2022; Yoon et al., 2020). From a legal perspective, re-identification is crucial in determining how data protection laws, such as the European Union's GDPR, are applied (Rupp & von Grafenstein, 2024). A popular and formal mathematical approach is differential privacy (DP), which holds great promise for disclosure control and quantifying the privacy risk of synthetic data (Wood et al., 2018). However, DP has limitations, and a stronger privacy guarantee can negatively impact task utility (Stadler et al., 2022).

This article addresses two main questions: 1. What are the implications of using generative model-based synthetic datasets regarding regulatory compliance? 2. What are the potential gaps in state-of-the-art privacy metrics of generative models? The first section examines the current regulatory landscapes associated with synthetic data from the EU, the UK, and the US perspectives. The second section demonstrates the state-of-the-art privacy metrics in generative models and the limitations of existing privacy-preserving approaches. The third section highlights potential future research directions that can promote fair and responsible synthetic data innovations with regulatory compliance.

2. Current Regulatory Perspectives

This section explores current regulations and guidelines, illustrating the worldwide efforts to establish an ethical standard in the development of generative AI. Generally,

[1] School of Computer Science and Informatics, De Montfort University, Leicester, United Kingdom [2] Department of Mathematics and Computer Science, University of Basel, Switzerland. Correspondence to: Debalina Padariya <p2723446@my365.dmu.ac.uk>, Isabel Wagner <isabel.wagner@unibas.ch>, Aboozar Taherkhani <aboozar.taherkhani@dmu.ac.uk>, Eerke Boiten <eerke.boiten@dmu.ac.uk>.

Accepted to the 2^{st} *Workshop on Generative AI and Law*, co-located with the *International Conference on Machine Learning*, Vienna, Austria. 2024.

an anonymization process modifies a dataset by removing or altering personal identifiers (PII) to prevent individuals from being linked to the information. Worldwide, regulatory efforts aim to address this while focusing on privacy and data protection regulations. According to Europe's GDPR guidelines, it is crucial to protect sensitive personal information, where the data is only considered anonymous if individuals cannot be re-identified, either directly or indirectly (Council, 2016). While Article 9(4) of the GDPR recognizes that some data is susceptible (Council, 2016), the Italian data protection issued Legislative Decree no. 101, which reflects the GDPR's principles with stricter requirements for processing biometric, genetic, and health-related data (Olivi, 2018). The CPRA (California Privacy Rights Act, an update to CCPA) refers to this anonymization as de-de-identification, which addresses data that is considered de-identified if it cannot be reasonably linked to a particular consumer or a small group of individuals, while the business using de-identification must ensure to prevent re-identification risks (Blesch, 2023). Moreover, according to the U.S. Health Insurance Portability and Accountability Act of 1996 (HIPAA), privacy is protected if a database lacks specific identifiers, such as names, social security numbers, geographic indicators, and any other elements deemed 'unique identifiers' (HIPAA, 1996).

The Financial Conduct Authority (FCA), Information Commissioner's Office (ICO) in the UK, and the Alan Turing Institute have investigated the standard framework and regulatory guidelines for using synthetic data (FCA, 2023; ICO, 2022; Jordon et al., 2022a). According to the findings in the report by Royal Society and the Alan Turing Institute, the synthetic data produced by machine learning models demonstrated the ability to memorize their training inputs, making them susceptible to inference attacks (Jordon et al., 2022a). The ICO UK also advised that organizations could leverage the benefits of synthetic data while ensuring the information is handled ethically and responsibly (ICO, 2022). Moreover, ICO UK has analyzed three key risk indicators: singling out, linkability, and inferences, which must be reduced to ensure effective anonymization (ICO, 2021). Therefore, the SDG must be aligned with relevant laws, especially data protection regulations.

Challenges - Privacy Attacks in Synthetic Data

While synthetic data generation is a promising solution for privacy-preserving data publishing, it is not the case of "Fake it till you make it" (Stadler et al., 2022). The synthetic data drawn from generative models are susceptible to various privacy or inference attacks, aiming to gain information not intended to be shared. Generally, privacy attacks try to infer sensitive information about the target generative model at different levels, such as training data (Chen et al., 2020b), attributes (Stadler et al., 2022), models (Hu & Pang, 2021), and identification-based (Croft et al., 2022). Existing research has focused on determining under which conditions a model is vulnerable to various privacy-related attacks.

At the training data level, an adversary can compromise the sensitive information in several ways, such as target individuals, full data samples, or macro-level information of the training samples. Various training data level attacks include membership inference (Chen et al., 2020b), property inference (Zhou et al., 2022), Set membership or co-membership (Liu et al., 2019), membership collision (Hu et al., 2021), and reconstruction attacks (Li et al., 2019). At the attribute level, the attacker tries to infer sensitive features of the attributes of training data, such as attribute inference (Stadler et al., 2022) and attribute disclosure (Goncalves et al., 2020). Moreover, at the model level, the adversary tries to steal information from the target model and potentially replicate the model, i.e., model extraction attacks (Hu & Pang, 2021). Further, the attacker attempts to recognize an individual's identity from different aspects, such as an identity recognition attack, identifying individuals based on their patterns, features, or characteristics (Croft et al., 2022) or re-identify individuals from an anonymized dataset, compromising privacy in re-identification attacks (Yoon et al., 2020).

3. Privacy Metrics in Generative Models

Despite the widespread success of generative models in various applications, several privacy threats have recently emerged as a significant concern. This has led researchers to focus on protecting privacy in generative models. This section demonstrates various privacy measures, highlighting this field's possible range of privacy guarantees.

A generic approach is attack-based privacy metrics, primarily focusing on data or model-level privacy by measuring the adversarial success rate (Chen et al., 2020b; Hu & Pang, 2021). Alternatively, some researchers have pointed out the poor generalization properties of generative models, where the proportion of overfitting can be a factor that measures information leakage (Chen et al., 2021). Besides, the widely accepted robust differential privacy-based metric (Dwork, 2008) attracts the most attention to protect generative models, which provide a theoretical privacy guarantee to protect individuals in training samples (Xie et al., 2018; Chen et al., 2018). Generally, a parameter epsilon (ϵ), known as privacy budget, controls the privacy level in differential privacy. The most common approach is to train the model using DPSGD (Differentially Private Stochastic Gradient Descent), adding Gaussian noise to the gradients during training (Abadi et al., 2016), or the PATE (Private Aggregation of Teacher Ensembles) mechanism, training distributed teacher models to transfer knowledge to generators (Jordon et al., 2022b). Since its introduction, several researchers have used DPSGD directly or extended this in various situations to train GANs and VAEs, such as different noise-adding mechanisms or improved optimization strategies (Xie et al., 2018; Chen et al., 2020a).

Challenges in State-of-the-art Privacy Metrics

This section identifies potential challenges in the current privacy-preserving approaches to synthetic data. Generally, the attack-based metrics use either classification-based or distance/similarity-based metrics (Chen et al., 2020b; Stadler et al., 2022). These metrics primarily focus on data or model-level privacy by measuring the adversarial success rate, and they do not offer formal guarantees about the level of model privacy protection. In classification-based metrics, there is no silver bullet since the choice of each metric depends on a problem's particular requirements. For instance, single-score classification-based metrics, e.g., accuracy, are most common due to their effectiveness for balanced datasets; however, they might overlook outliers and are inadequate for imbalanced and multi-class classification problems (Hyeong et al., 2022). Additionally, selecting privacy metrics that reflect the average and worst case is often recommended (Wagner & Eckhoff, 2018). However, distance/similarity-based metrics often focus on average-case performance and may not effectively address worst-case possibilities (Ganev & De Cristofaro, 2023). Besides, the generalization-based metric improves model generalizability that can address overfitting problems to protect privacy in generative models (Chen et al., 2021); however, they may not fully address the privacy concerns regarding model-level privacy protection.

The robust differential privacy (DP) approach gained attention after the well-known issues with the Netflix Prize contest (Aitsam, 2022). DP ensures a measurable privacy guarantee; however, differential privacy is unlike having a "rich, calorie-free cake." The implications of DP in generative models are wide-ranging, and researchers should consider these aspects in their works. First, in DP, adding more noise improves privacy and reduces task accuracy (Stadler et al., 2022). Second, determining the appropriate privacy budget is complicated, and researchers have investigated the optimal selection of privacy budget to protect their models (Ganev et al., 2023). Third, the applicability of DP may be challenging in healthcare settings, which often deal with finite training samples (Yoon et al., 2020). Since DP performs well with many training samples, it can not be directly computed with finite training samples. Finally, DP unfairly increases the influence of majority subgroups, which becomes more significant with downstream predictions due to highly imbalanced datasets (Cheng et al., 2021).

The ICO UK has issued guidance on privacy-enhancing technologies, offering additional insights about using synthetic data (ICO, 2022). The guidance is developed with the support of the Financial Conduct Authority (FCA) and the Alan Turing Institute, which address the trade-offs between privacy and utility for various use cases. While DP offers rigorous statistical assurance to counter privacy attacks, several challenges in DP require careful consideration in future research.

4. Conclusions and Future Directions

In this work, we address the possible gaps in current privacy-preserving approaches of synthetic datasets through the lens of privacy protection regulations. This section highlights potential future research directions that can promote fair and responsible innovation in synthetic data while ensuring ethical practice in generative AI:

Bias Mitigation and Fairness: Despite the groundbreaking innovations of generative AI, bias and fairness remain a significant challenge. Generative model-based synthetic data can inherit biases during their development, which are introduced through the algorithms used to learn from the training samples of real-world data (Chen et al., 2024). Fairness in synthetic data involves recognizing and rectifying biases in training data and algorithms to avoid discriminating against certain groups. Fairness in synthetic data can be interpreted in some ways, such as debiasing techniques (Draghi et al., 2024), fairness metrics (Zhou et al., 2024), or counterfactual fairness (Abroshan et al., 2022).

Transparency and Explainability: Transparency assesses the reliability of the synthetic data generation process, which can be accelerated by generating high-fidelity synthetic data (Smith et al., 2022). Moreover, transparency enables the stakeholders to understand the decision-making process, enabling the development of explainability methods for synthetic data. Explainability is the best practice for building trust while effectively assessing potential biases incorporated with fairness.

Privacy and Regulatory Compliance: Differential Privacy has emerged as a de facto standard for privacy-preserving synthetic data; however, the trade-offs between privacy and utility are complex. While advanced differential privacy mechanisms provide robust privacy guarantees (Ma et al., 2023), carefully calibrating DP parameters is crucial to balance the trade-offs. Since the creation of generative AI has sparked significant ethical issues regarding misinformation and consent (Kwok & Koh, 2021), organizations should proactively ensure compliance with data protection regulations regarding responsible synthetic data innovation.

Acknowledgements

We thank the anonymous reviewers for their insightful feedback and suggestions. This work was supported by the Alan Turing Institute under the Turing/Accenture strategic partnership grant R-AST-040.

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L. Deep Learning with Differential Privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 308–318, Vienna Austria, October 2016. ACM. ISBN 9781450341394. doi: 10.1145/2976749.2978318. URL https://dl.acm.org/doi/10.1145/2976749.2978318.

Abroshan, M., Khalili, M. M., and Elliott, A. Counterfactual Fairness in Synthetic Data Generation. October 2022. URL https://openreview.net/forum?id=tge5NiX4CZo.

Aitsam, M. Differential Privacy Made Easy. In *2022 International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE)*, pp. 1–7, December 2022. doi: 10.1109/ETECTE55893.2022.10007322. URL https://ieeexplore.ieee.org/document/10007322/;jsessionid=C3BDF4356E8E2B829ABC4A27566AA031.

Bellovin, S. M., Dutta, P. K., and Reitinger, N. Privacy and synthetic datasets. *Stan. Tech. L. Rev.*, 22:1, 2019.

Blesch, W. The GDPR's Anonymization versus CCPA/CPRA's De-identification, August 2023.

Chen, D., Orekondy, T., and Fritz, M. GS-WGAN: a gradient-sanitized approach for learning differentially private generators. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, pp. 12673–12684, Red Hook, NY, USA, December 2020a. Curran Associates Inc. ISBN 9781713829546.

Chen, D., Yu, N., Zhang, Y., and Fritz, M. GAN-Leaks: A Taxonomy of Membership Inference Attacks against Generative Models. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, pp. 343–362, Virtual Event USA, October 2020b. ACM. ISBN 9781450370899. doi: 10.1145/3372297.3417238. URL https://dl.acm.org/doi/10.1145/3372297.3417238.

Chen, J., Wang, W. H., Gao, H., and Shi, X. PAR-GAN: Improving the Generalization of Generative Adversarial Networks Against Membership Inference Attacks. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, pp. 127–137, New York, NY, USA, August 2021. Association for Computing Machinery. ISBN 9781450383325. doi: 10.1145/3447548.3467445. URL https://doi.org/10.1145/3447548.3467445.

Chen, Q., Xiang, C., Xue, M., Li, B., Borisov, N., Kaarfar, D., and Zhu, H. Differentially private data generative models. *arXiv preprint arXiv:1812.02274*, 2018.

Chen, T., Hirota, Y., Otani, M., Garcia, N., and Nakashima, Y. Would Deep Generative Models Amplify Bias in Future Models?, April 2024. URL http://arxiv.org/abs/2404.03242. arXiv:2404.03242 [cs].

Cheng, V., Suriyakumar, V. M., Dullerud, N., Joshi, S., and Ghassemi, M. Can You Fake It Until You Make It?: Impacts of Differentially Private Synthetic Data on Downstream Classification Fairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 149–160, Virtual Event Canada, March 2021. ACM. ISBN 9781450383097. doi: 10.1145/3442188.3445879. URL https://dl.acm.org/doi/10.1145/3442188.3445879.

Council, E. a. Art. 9 GDPR – Processing of special categories of personal data, 2016. URL https://gdpr-info.eu/art-9-gdpr/.

Croft, W. L., Sack, J.-R., and Shi, W. Differentially private facial obfuscation via generative adversarial networks. *Future Generation Computer Systems*, 129:358–379, April 2022. ISSN 0167-739X. doi: 10.1016/j.future.2021.11.032. URL https://www.sciencedirect.com/science/article/pii/S0167739X21004763.

Draghi, B., Wang, Z., Myles, P., and Tucker, A. Identifying and handling data bias within primary healthcare data using synthetic data generators. *Heliyon*, 10(2):e24164, January 2024. ISSN 2405-8440. doi: 10.1016/j.heliyon.2024.e24164. URL https://www.sciencedirect.com/science/article/pii/S2405844024001956.

Dwork, C. Differential Privacy: A Survey of Results. In Agrawal, M., Du, D., Duan, Z., and Li, A. (eds.), *Theory and Applications of Models of Computation*, Lecture Notes in Computer Science, pp. 1–19, Berlin, Heidelberg, 2008. Springer. ISBN 9783540792284. doi: 10.1007/978-3-540-79228-4_1.

FCA, U. Synthetic Data Call for Input feedback Statement. Technical report, 2023. URL https://www.fca.org.uk/publication/feedback/fs23-1.pdf.

Feuerriegel, S., Hartmann, J., Janiesch, C., and Zschech, P. Generative AI. *Business & Information Systems Engineering*, 66(1):111–126, February 2024. ISSN 1867-

0202. doi: 10.1007/s12599-023-00834-7. URL https://doi.org/10.1007/s12599-023-00834-7.

Ganev, G. and De Cristofaro, E. On the Inadequacy of Similarity-based Privacy Metrics: Reconstruction Attacks against "Truly Anonymous Synthetic Data", December 2023. URL http://arxiv.org/abs/2312.05114. arXiv:2312.05114 [cs].

Ganev, G., Xu, K., and De Cristofaro, E. Understanding how Differentially Private Generative Models Spend their Privacy Budget, May 2023. URL http://arxiv.org/abs/2305.10994. arXiv:2305.10994 [cs].

Golda, A., Mekonen, K., Pandey, A., Singh, A., Hassija, V., Chamola, V., and Sikdar, B. Privacy and Security Concerns in Generative AI: A Comprehensive Survey. *IEEE Access*, 12:48126–48144, 2024. ISSN 2169-3536. doi: 10.1109/ACCESS.2024.3381611. URL https://ieeexplore.ieee.org/abstract/document/10478883/.

Goncalves, A., Ray, P., Soper, B., Stevens, J., Coyle, L., and Sales, A. P. Generation and evaluation of synthetic patient data. *BMC Medical Research Methodology*, 20 (1):108, May 2020. ISSN 1471-2288. doi: 10.1186/s12874-020-00977-1. URL https://doi.org/10.1186/s12874-020-00977-1.

HIPAA. HIPAA Basics Overview | Health Insurance Portability and Accountability Act (HIPAA), 1996. URL https://uwm.edu/hipaa/overview/hipaa-basics-overview/.

Hu, A., Xie, R., Lu, Z., Hu, A., and Xue, M. TableGAN-MCA: Evaluating Membership Collisions of GAN-Synthesized Tabular Data Releasing. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, CCS '21, pp. 2096–2112, New York, NY, USA, November 2021. Association for Computing Machinery. ISBN 9781450384544. doi: 10.1145/3460120.3485251. URL https://doi.org/10.1145/3460120.3485251.

Hu, H. and Pang, J. Model extraction and defenses on generative adversarial networks. *arXiv preprint arXiv:2101.02069*, 2021.

Hyeong, J., Kim, J., Park, N., and Jajodia, S. An Empirical Study on the Membership Inference Attack against Tabular Data Synthesis Models. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, CIKM '22, pp. 4064–4068, Atlanta, GA, USA, October 2022. Association for Computing Machinery. ISBN 978-1-4503-9236-5. doi: 10.1145/3511808.3557546. URL https://doi.org/10.1145/3511808.3557546.

ICO, U. Chapter 2: How do we ensure anonymisation is effective?, 2021. URL https://ico.org.uk/media/about-the-ico/documents/4018606/chapter-2-anonymisation-draft.pdf.

ICO, U. Chapter 5: privacy-enhancing technologies (PETs). Technical report, 2022. URL https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf.

Jordon, J., Szpruch, L., Houssiau, F., Bottarelli, M., Cherubin, G., Maple, C., Cohen, S. N., and Weller, A. Synthetic Data – what, why and how?, May 2022a. URL http://arxiv.org/abs/2205.03257. arXiv:2205.03257 [cs].

Jordon, J., Yoon, J., and Schaar, M. v. d. PATE-GAN: Generating Synthetic Data with Differential Privacy Guarantees. In *ICLR 2019*, New Orleans, LA, USA, February 2022b. URL https://openreview.net/forum?id=S1zk9iRqF7.

Khanal, S., Zhang, H., and Taeihagh, A. Why and how is the power of big tech increasing in the policy process? the case of generative ai. *Policy and Society*, pp. puae012, 2024.

Kwok, A. O. J. and Koh, S. G. M. Deepfake: a social construction of technology perspective. *Current Issues in Tourism*, 24(13):1798–1802, July 2021. ISSN 1368-3500, 1747-7603. doi: 10.1080/13683500.2020.1738357. URL https://www.tandfonline.com/doi/full/10.1080/13683500.2020.1738357.

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Li, Y., Wang, Y., and Li, D. Privacy-preserving lightweight face recognition. *Neurocomputing*, 363(C):212–222, October 2019. ISSN 0925-2312. doi: 10.1016/j.neucom.2019.07.039. URL https://doi.org/10.1016/j.neucom.2019.07.039.

Liu, K. S., Xiao, C., Li, B., and Gao, J. Performing Co-membership Attacks Against Deep Generative Models. In *2019 IEEE International Conference on Data Mining (ICDM)*, pp. 459–467, Beijing, China, november 2019. IEEE. ISBN 9781728146041. doi: 10.1109/ICDM.2019.00056. URL https://ieeexplore.ieee.org/document/8970995/.

Lu, Y., Shen, M., Wang, H., Wang, X., van Rechem, C., Fu, T., and Wei, W. Machine Learning for Synthetic

Data Generation: A Review, May 2024. URL http://arxiv.org/abs/2302.04062. arXiv:2302.04062 [cs].

Ma, C., Li, J., Ding, M., Liu, B., Wei, K., Weng, J., and Poor, H. V. RDP-GAN: A rényi-differential privacy based generative adversarial network. *IEEE Transactions on Dependable and Secure Computing*, pp. 1–15, 2023. ISSN 1941-0018. doi: 10.1109/TDSC.2022.3233580. Conference Name: IEEE Transactions on Dependable and Secure Computing.

Mendes, J., Pereira, T., Silva, F., Frade, J., Morgado, J., Freitas, C., Negrão, E., de Lima, B. F., da Silva, M. C., Madureira, A. J., Ramos, I., Costa, J. L., Hespanhol, V., Cunha, A., and Oliveira, H. P. Lung CT image synthesis using GANs. *Expert Systems with Applications*, 215:119350, April 2023. ISSN 0957-4174. doi: 10.1016/j.eswa.2022.119350. URL https://www.sciencedirect.com/science/article/pii/S0957417422023685.

Olivi, G. Italian data protection code reformed to enact GDPR. What is new?, September 2018.

Rupp, V. and von Grafenstein, M. Clarifying "personal data" and the role of anonymisation in data protection law: Including and excluding data from the scope of the GDPR (more clearly) through refining the concept of data protection. *Computer Law & Security Review*, 52:105932, April 2024. ISSN 0267-3649. doi: 10.1016/j.clsr.2023.105932. URL https://www.sciencedirect.com/science/article/pii/S0267364923001425.

Smith, A., Lambert, P. C., and Rutherford, M. J. Generating high-fidelity synthetic time-to-event datasets to improve data transparency and accessibility. *BMC Medical Research Methodology*, 22(1):176, June 2022. ISSN 1471-2288. doi: 10.1186/s12874-022-01654-1. URL https://doi.org/10.1186/s12874-022-01654-1.

Stadler, T., Oprisanu, B., and Troncoso, C. Synthetic Data – Anonymisation Groundhog Day. pp. 1451–1468, 2022. ISBN 9781939133311. URL https://www.usenix.org/conference/usenixsecurity22/presentation/stadler.

Wagner, I. and Eckhoff, D. Technical Privacy Metrics: A Systematic Survey. *ACM Comput. Surv.*, 51 (3):57:1–57:38, June 2018. ISSN 0360-0300. doi: 10.1145/3168389. URL https://doi.org/10.1145/3168389.

Wood, A., Altman, M., Bembenek, A., Bun, M., Gaboardi, M., Honaker, J., Nissim, K., O'Brien, D. R., Steinke, T., and Vadhan, S. Differential privacy: A primer for a non-technical audience. *Vand. J. Ent. & Tech. L.*, 21:209, 2018.

Xie, L., Lin, K., Wang, S., Wang, F., and Zhou, J. Differentially Private Generative Adversarial Network, 2018. URL https://arxiv.org/abs/1802.06739.

Yoon, J., Drumright, L. N., and van der Schaar, M. Anonymization Through Data Synthesis Using Generative Adversarial Networks (ADS-GAN). *IEEE Journal of Biomedical and Health Informatics*, 24(8):2378–2388, August 2020. ISSN 2168-2194, 2168-2208. doi: 10.1109/JBHI.2020.2980262. URL https://ieeexplore.ieee.org/document/9034117/.

Zhou, J., Chen, Y., Shen, C., and Zhang, Y. Property Inference Attacks Against GANs. In *NDSS 2022*, San Diego, California, April 2022. URL https://publications.cispa.saarland/3636/.

Zhou, M., Abhishek, V., Derdenger, T., Kim, J., and Srinivasan, K. Bias in Generative AI, March 2024. URL http://arxiv.org/abs/2403.02726. arXiv:2403.02726 [cs, econ, q-fin].