Proxy Fairness under the GDPR and the AI ACT: A perspective of sensitivity and necessity*

An Extended Abstract

Ioanna Papageorgiou[1]

[1]Leibniz University Hannover, Institute for Legal Informatics, Germany

Introduction The increasing adoption of AI systems at high stake areas of public life along with extensive studies on the discriminatory potential of AI [1] have prompted a proliferation of algorithmic methods that study and pursue fairness in AI systems (Fair-AI) ([2, 3, 4]. These methods are centered on the detection, mitigation and evaluation of bias across legally protected groups, and almost invariably require access to sensitive attributes, like demographics, that determine group membership. However, this often implies the processing of personal sensitive data, which is in principle prohibited or extensively protected according to the EU data protection law, posing challenges to the feasibility of Fair-AI approaches. In response to this challenge, a growing line of AI research [5, 6, 7, 8, 9, 10, 11] has studied computational methods that enable fairness operationalization in the absence of demographic data, notably through the use of proxy variables and inferential techniques (*Proxy Fairness*).

However, scant attention has been given thus far to the interaction of these methods with existing data protection regulations, posing significant legal uncertainty regarding their legitimacy. This uncertainty intensifies in the face of ongoing regulatory developments. Particularly, the upcoming AI Act has also addressed the challenge of data scarcity in the context of Fairness, by enabling, on grounds of public interest, the processing of personal sensitive data for the purposes of bias detection and correction in high-risk AI systems. Precisely, according to the Article 10 (5) AI Act, the processing of personal sensitive data is permitted only "to the extent that it is *strictly necessary* for the purposes of ensuring bias detection and correction in relation to the high-risk AI systems..[emphasis added]". While the enabling provision appears to be method-agnostic, meaning that it's not restricted to a particular fairness approach, the stipulated necessity requirement significantly influences the choice of fairness methods, and to a greater extent, the scope of Proxy Fairness.

By utilizing the legal notions of data- *Sensitivity* and processing- *Necessity*, the paper examines the legal implications of Proxy Fairness under the General Data Protection Regulation and the AI Act, providing a normative foundation to this line of Fair-AI approaches. Precisely, the paper scrutinizes the nature of data involved in Proxy Fairness approaches- including proxy variables and data inferences- demonstrating that inferential methods are in principle not exempt from the reach of the GDPR and its extensive regime for sensitive data. Subsequently, the paper

EWAF'24: European Workshop on Algorithmic Fairness, July 01–03, 2024, Mainz, Germany

*This is an extended abstract. Forthcoming paper in 7th AAAI/ACM Conference on AI, Ethics, and Society (AIES-24).

examines the lawfulness of processing sensitive data for Proxy Fairness under article 10 (5) of the AI Act through a comparative assessment of proxy fairness approaches versus default alternatives along the necessity axes of *intrusiveness*, *effectiveness*, and *reasonableness*.

Proxy Fairness under the GDPR: a sensitivity perspective In order to assess Proxy fairness under article 10 (5) of the AI Act, it is necessary to first investigate the extent to which it involves the processing of *sensitive* data under the meaning of the GDPR. For this purpose, the paper distinguishes between two main data-pillars involved in Proxy Fairness, namely *Proxy* and *Inferred* data, and assesses them under the legal notion of sensitivity. Particularly, through on a grammatical and systematic interpretation of article 9 (1) GDPR, which defines sensitive personal data, and by consulting the jurisprudence of the European Court of Justice [12, 13], guidelines from the Article 29 Working Party [14, 15, 16, 17] and a substantial corpus of legal scholarship [18, 19, 20, 19, 18, 21, 22, 23, 24, 25], the paper supports that both proxy and inferred data used in the context of Proxy Fairness may be considered sensitive within the meaning of the GDPR.

Proxy Fairness under the AI Act: a necessity perspective As mentioned above, according to article 10 (5) AI Act, the processing of sensitive data is permitted only "to the extent that it is strictly *necessary* for the purposes of ensuring negative bias detection and correction in relation to the high-risk AI systems [emphasis added]", i.e. only under the requirement of legal necessity. The necessity principle, which has been a recurrent condition to the processing of personal data, essentially dictates that data processing is permissible only to the extent that there is not a *less intrusive* but *similarly effective* alternative available, which can *reasonably* achieve the objective at hand [26, 27]. AI providers seeking to rely on the exception of the AI Act and process sensitive personal data for bias detection and correction must thus conduct a necessity test, which involves comparing available alternatives based on their levels of a) *intrusiveness*, b) *effectiveness* and c) *reasonableness*. The paper examines proxy fairness approaches under the necessity requirement, particularly by comparing them with default approaches that directly collect and use real sensitive attributes, along the necessity axes.

a. intrusiveness Core criteria for assessing the intrusiveness of a data processing operation — i.e. the severity of the interference with the right to data protection— include the *volume* and *type* of data processed and the associated risks of *data misuse* [27]. Examining these criteria, the paper argues that Proxy Fairness not only *de facto* involves a larger volume of personal data compared to default approaches, but also a larger volume of *de jure* sensitive data, thereby being more intrusive under the first two criteria. Subsequently, the paper discusses the lack of data subjects' control over their personal data and the risk of discrimination as relevant instances of data misuse in the cases of Proxy and Default Fairness respectively, highlighting the complexity of comparing different methods in terms of data misuse risks.

b. effectiveness Compliance with the requirement of necessity does not require prioritizing any kind of milder alternative, but only those milder alternatives that can attain the pursued objective in a comparably effective manner. In a second step, AI providers must thus compare the identified alternatives with respect to their effectiveness in detecting and correcting bias, by relying on theoretical and/or empirical evidence regarding the utility and limitations of the fairness methods under consideration. This includes qualitative and quantitative arguments about the way relevant demographic groups would be better served by the planned intervention, such as performance and fairness metrics, accuracy of fairness and associated trade-offs. Accordingly, the paper conducted a high-level effectiveness- comparison between Default and Proxy Fairness approaches based on evidence discussed in the Fairness literature.

c. reasonableness According to the last element of the necessity, AI providers are required to prioritize milder effective alternatives only if those are reasonable in terms of *financial*, *legal*, and *operational feasibility*. Particularly, nothing prohibitively costly, practically impossible or illegal shall be demanded. The paper argues that this step provides space not only for a utility-based calculus but also for *ethical* considerations, demonstrating how current research on critical ethics can gain normative relevance in the context of the GDPR and the AI Act.

Conclusion In the face of the increasing popularity of proxy fairness approaches and the lack of a thorough corresponding legal framework, this paper explored aspects of Proxy Fairness under the General Data Protection Regulation and the AI Act. By shedding light on the regulatory nuances involved in Proxy Fairness and providing interpretational tools for a lawful processing of sensitive data in this context, the paper aims to assist AI providers in regulatory compliance and safeguard the data protection rights of data subjects, while laying the groundwork for further research at the intersection of data protection law, ethics, and Fair-AI.

Acknowledgments This work has received funding from the European Union's Horizon 2020 research and innovation program under Marie Sklodowska-Curie Actions (Grant Agreement Number 860630) for the project "NoBIAS—Artificial Intelligence without Bias". This work reflects only the author's views and the European Research Executive Agency (REA) is not responsible for any use that may be made.

References

[1] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, A. Galstyan, A survey on bias and fairness in machine learning, ACM Comput. Surv. 54 (2021) 115:1–115:35.

[2] E. Ntoutsi, et al., Bias in data-driven artificial intelligence systems - an introductory survey, WIREs Data Mining Knowl. Discov. 10 (2020).

[3] R. Schwartz, A. Vassilev, K. Greene, L. Perine, A. Burt, P. Hall, Towards a Standard for Identifying and Managing Bias in Artificial Intelligence, Technical Report 1270, NIST Special Publication, 2022.

[4] S. Mitchell, E. Potash, S. Barocas, A. D'Amour, K. Lum, Algorithmic fairness: Choices, assumptions, and definitions, Annual Review of Statistics and Its Application 8 (2021) 141–163. URL: https://ssrn.com/abstract=3800687. doi:doi: 10.1146/annurev-statistics-042720-125902.

[5] C. Ashurst, A. Weller, Fairness without demographic data: A survey of approaches, in: Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, Association for Computing Machinery, New York, NY, USA, 2023. URL: https://doi.org/10.1145/3617694.3623234. doi:doi: 10.1145/3617694.3623234.
[6] Centre for Data Ethics and Innovation and Department for Science, Innovation and Technology, Enabling responsible access to demographic data to make ai systems fairer, Research and analysis report, 2023. URL: https://www.gov.uk/government/publications/enabling-responsible-access-to-demographic-data-to-make-ai-systems-fairer/report-enabling-responsible-access-to-demographic-data-to-make-ai-systems-fairer, published on 14 June 2023.
[7] R. Awasthi, A. Beutel, M. Kleindessner, J. Morgenstern, X. Wang, Evaluating fairness of machine learning models under uncertain and incomplete information, in: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 2021, pp. 206–214.
[8] J. Chen, N. Kallus, X. Mao, G. Svacha, M. Udell, Fairness under unawareness: Assessing disparity when protected class is unobserved, in: Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 339–348. URL: https://doi.org/10.1145/3287560.3287594. doi:doi: 10.1145/3287560.3287594.
[9] S. Yan, H. te Kao, E. Ferrara, Fair class balancing: Enhancing model fairness without observing sensitive attributes, in: Proceedings of the 29th ACM, 2020.
[10] Z. Zhu, Y. Yao, J. Sun, H. Li, Y. Liu, Weak proxies are sufficient and preferable for fairness with missing sensitive attributes, in: Proceedings of the 40th International Conference on Machine Learning, ICML'23, JMLR.org, 2023.
[11] M. R. Gupta, A. Cotter, M. M. Fard, S. L. Wang, Proxy fairness, CoRR abs/1806.11212 (2018). URL: http://arxiv.org/abs/1806.11212. arXiv:1806.11212.
[12] Nowak, Judgment of the court (second chamber) of 20 december 2017, court of justice of the european union c-434/16, 2017.
[13] K. Egan, M. H. v European Parliament, Egan and hackett v parliament, ECLI:EU:C:2019:1064, 2012. Judgment of the General Court (Fifth Chamber) of 28 March 2012.
[14] Article 29 Data Protection Working Party, Opinion on the concept of personal data, 2007.
[15] A. . W. P. Art29WP, Article 29 data protection working party opinion 3/2012 on developments in biometric technologies, 2012. URL: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2012/wp193_en.pdf.
[16] Article 29 Data Protection Working Party, Guidelines on the right to data portability, Data Protection Working Party (2016) 9–11. URL: https://ec.europa.eu/newsroom/document.cfm?doc_id=44099, on file with the Columbia Business Law Review.
[17] Article 29 Data Protection Working Party, Guidelines on automated individual decision-making and profiling for the purposes of regulation 2016/679, 2017. URL: https://ec.europa.eu/newsroom/article29/items/612053, document No. 17/EN, WP251rev.01.
[18] S. Wachter, B. Mittelstadt, A right to reasonable inferences: Re-thinking data protection law in the age of big data and ai, Columbia Business Law Review 2019 (2018). URL: https://ssrn.com/abstract=3248829, october 5, 2018.
[19] P. Quinn, G. Malgieri, The difficulty of defining sensitive data – the concept of sensitive

data in the eu data protection framework, German Law Journal (2020). URL: https://ssrn.com/abstract=3713134. doi:doi: 10.2139/ssrn.3713134, (Forthcoming).
[20] G. Malgieri, G. Comandè, Sensitive-by-distance: quasi-health data in the algorithmic era, Information & Communications Technology Law 26 (2017) 229–249. URL: https://doi.org/10.1080/13600834.2017.1335468. doi:doi: 10.1080/13600834.2017.1335468.
[21] D. J. Solove, Data is what data does: Regulating based on harm and risk instead of sensitive data, Northwestern University Law Review 118 (2024) 1081. URL: https://ssrn.com/abstract=4322198. doi:doi: 10.2139/ssrn.4322198, gWU Legal Studies Research Paper No. 2023-22, GWU Law School Public Law Research Paper No. 2023-22.
[22] A. Schiff, Ehmann, Selmayr, Datenschutz-Grundverordnung DS-GVO Kommentar, 3. auflage ed., C.H.BECK, 2017.
[23] P. Gola, D. Heckmann, Datenschutz-Grundverordnung, Bundesdatenschutzgesetz: DS-GVO / BDSG, 3 ed., C.H. Beck, 2022.
[24] M. Finck, Hidden Personal Insights and Entangled in the Algorithmic Model: The Limits of the GDPR in the Personalisation Context, Cambridge University Press, 2021, pp. 95–107.
[25] D. Hallinan, F. Zuiderveen Borgesius, Opinions can be incorrect! in our opinion. on the accuracy principle in data protection law, International Data Privacy Law ipz025 (2020). doi:doi: 10.1093/idpl/ipz025.
[26] European Data Protection Supervisor, Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A toolkit, 2023. URL: https://www.edps.europa.eu/data-protection/our-work/publications/papers/necessity-toolkit_en.
[27] P. Schantz, H. A. Wolff, Das neue Datenschutzrecht: Datenschutz-Grundverordnung und Bundesdatenschutzgesetz in der Praxis, C.H.BECK., 2017.