

Algorithmic Unfairness Through the Lens of EU Non-Discrimination Law

Or Why the Law is Not a Decision Tree

Hilde Weerts^{1,†}, Raphaële Xenidis^{2,†}, Fabien Tarissan³, Henrik Palmer Olsen⁴ and Mykola Pechenizkiy¹

¹*Eindhoven University of Technology, The Netherlands*

²*Sciences Po Law School, France*

³*CNRS & ENS Paris-Saclay, France*

⁴*University of Copenhagen, Denmark*

Abstract

Concerns regarding unfairness and discrimination in the context of artificial intelligence (AI) systems have recently received increased attention from both legal and computer science scholars. Yet, the degree of overlap between notions of algorithmic bias and fairness on the one hand, and legal notions of discrimination and equality on the other, is often unclear, leading to misunderstandings between computer science and law. In this paper, we aim to illustrate to what extent European Union (EU) non-discrimination law coincides with notions of algorithmic fairness proposed in computer science literature and where they differ. Ultimately, we show that metaphors depicting the law as a decision tree are misleading. We suggest moving away from asking *what* should be equal, and towards asking *why* a particular distribution of burdens and benefits is right in a given context.

Keywords

discrimination, fairness metrics, technical interventions, EU law, legal compliance

1. Introduction

Concerns regarding unfairness and discrimination in the context of artificial intelligence (AI) systems have recently received increased attention from both legal and computer science scholars. Yet, the degree of overlap between notions of algorithmic bias and fairness on the one hand, and legal notions of discrimination and equality on the other, is often unclear, leading to misunderstandings between computer science and law.

On the one hand, computer scientists have put forward various metrics and technical interventions to measure and mitigate unfairness of AI systems. However, an AI practitioner hoping

This is an extended abstract of the full paper appearing in the proceedings of 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23).

EWAF'23: European Workshop on Algorithmic Fairness, June 07–09, 2023, Winterthur, Switzerland


[†]These authors contributed equally.

✉ h.j.p.weerts@tue.nl (H. Weerts); raphael.xenidis@sciencespo.fr (R. Xenidis); fabien.tarissan@ens-paris-saclay.fr (F. Tarissan); henrik@jur.ku.dk (H. P. Olsen); m.pechenizkiy@tue.nl (M. Pechenizkiy)

🆔 0000-0002-2046-1299 (H. Weerts); 0000-0003-0556-385X (R. Xenidis); 0000-0002-7588-300X (F. Tarissan); 0000-0001-8486-8752 (H. P. Olsen); 0000-0003-4955-0743 (M. Pechenizkiy)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

for an explicit answer to the question: “what should be the value of my fairness metric for my system to be compliant with the law?” is likely to be disappointed, as most of the time the answer will amount to a variation of “it depends”.

On the other hand, challenges of algorithmic unfairness are not always properly understood by legal scholars. The technical translation of legal standards raises a range of difficult normative questions that force lawyers to question the content of overarching legal principles such as equal treatment and non-discrimination. Since courts are called on to interpret the normative content of those polysemous legal norms contextually and on a case-by-case basis, a straightforward technical translation of those norms is impossible.

As a result, computer scientists struggle to understand how legal compliance with equality law can be ensured, and legal experts and regulators struggle with figuring out how discrimination law can properly address algorithmic bias and unfairness. Additionally, we observe a tendency on both sides to overestimate the solutions and answers provided by each discipline. The legal community tends to overestimate the effectiveness and applicability of technical interventions [1]. In turn, computer scientists place perhaps too much confidence in the principle of legal certainty, and the determinacy and specificity of legal norms.

This raises several important questions. What types of bias and unfairness does the law address when it prohibits discrimination? What role can fairness metrics play in establishing legal compliance – if any? This paper aims to respond to computer scientists’ uncertainties about what is legal when it comes to discrimination, and to lawyers’ questions regarding the challenges and technical possibilities to realise equality rights and non-discrimination law obligations. To this end, we consider European Union (EU) non-discrimination law and we show to what extent non-discrimination law coincides with notions of algorithmic fairness proposed in computer science literature and where they differ. In so doing, we target a broader audience, bridging two distinct disciplines.

Existing work in this direction has primarily targeted a legal audience [2, 3, 4]. Most notably, Wachter et al. [4] set out how the contextual nature of EU non-discrimination law makes it impossible to automate non-discrimination in the context of AI systems and propose a fairness metric that aligns with the “gold standard” of the Court of Justice of the European Union (CJEU). Additionally, several works focus on US anti-discrimination law [5, 6, 7]. For example, Hellman considers the compatibility of several fairness metrics under US anti-discrimination law and touches upon the legitimacy of particular types of technical interventions [5].

The contributions of this paper are as follows. First, we analyse influential examples of algorithmic unfairness through the lens of EU non-discrimination law, drawing parallels with EU case law. Second, we set out the normative underpinnings of fairness metrics and technical interventions and compare these to the legal reasoning of the Court of Justice of the EU. Specifically, we show how normative assumptions often remain implicit in both disciplinary approaches and explain the ensuing limitations of current AI practice and non-discrimination law. We conclude with implications for AI practitioners and regulators.

2. Algorithmic Unfairness Through the Lens of Non-Discrimination Law

We analyse three influential examples of algorithmic unfairness through the lens of EU non-discrimination law, namely the Dutch childcare benefits scandal [8, 9, 10], the Amazon’s recruitment algorithm [11] and the gender shades [12, 13], drawing parallels with EU case law. The purpose is to establish a taxonomy of algorithmic discriminatory harms, assess when and how those harms fall within the scope of EU equality law, and determine how they can be redressed from a legal point of view. Relying on those examples, we show that, although EU non-discrimination law is in principle suited to deal with types of algorithmic unfairness that closely resemble human discrimination, it cannot be readily applied to all cases of disparate predictive performance. Moreover, the unintelligibility of prediction-generating mechanisms and lack of transparency regarding important design choices of AI systems make it difficult for applicants to provide *prima facie* evidence.

3. The Problem of Emptiness

We then set out the normative underpinnings of fairness metrics and technical interventions [14] and compare these to the legal reasoning of the CJEU. Specifically, we show how normative assumptions often remain implicit in both disciplinary approaches, and explain the ensuing limitations of current AI practice and non-discrimination law. To do so, we map the requirements of non-discrimination law to algorithmic fairness research relying on a case law analysis of landmark decisions of the Court of Justice of the EU. We reveal the ‘emptiness’ [15] of equality norms and fairness approaches and argue that uncovering – and reflecting upon – the normative baselines used as equality standards is key to ‘translating’ legal and technical approaches to fairness and discrimination.

4. Conclusion

Understanding when particular interventions are appropriate is especially important considering the difficulties applicants face in providing *prima facie* evidence in the context of opaque algorithmic systems. While many fairness metrics have taken inspiration from non-discrimination law, legal compliance cannot translate into a single threshold or fairness metric. In other terms, EU equality law is not a decision tree. Rather, fulfilling the requirements of non-discrimination law demands reflecting explicitly on the normative goal of legal and technical fairness interventions. We suggest that, in order to meaningfully answer the question that non-discrimination law poses, we must move beyond merely asking *what* should be equal and, instead, ask ourselves *why* a particular distribution of burdens and benefits is right in a given context.

Acknowledgments

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 898937. We

also thank the organisers of the Lorentz workshop on *Fairness in Algorithmic Decision Making: A Domain-Specific Approach* in March 2022 for bringing together a group of researchers with diverse disciplinary backgrounds as well as the participants for their valuable insights.

References

- [1] A. Balayn, S. Gürses, Beyond debiasing: Regulating ai and its inequalities, EDRI Report. https://edri.org/wp-content/uploads/2021/09/EDRI_Beyond-Debiasing-Report_Online.pdf (2021).
- [2] R. Binns, Algorithmic decision-making: A guide for lawyers, *Judicial Review* 25 (2020) 2–7.
- [3] S. Wachter, B. Mittelstadt, C. Russell, Bias preservation in machine learning: the legality of fairness metrics under eu non-discrimination law, *W. Va. L. Rev.* 123 (2020) 735.
- [4] S. Wachter, B. Mittelstadt, C. Russell, Why fairness cannot be automated: Bridging the gap between eu non-discrimination law and ai, *Computer Law & Security Review* 41 (2021) 105567.
- [5] D. Hellman, Measuring algorithmic fairness, *Virginia Law Review* 106 (2020) 811–866.
- [6] P. T. Kim, Race-aware algorithms: Fairness, nondiscrimination and affirmative action, *Cal. L. Rev.* 110 (2022) 1539.
- [7] I. E. Kumar, K. E. Hines, J. P. Dickerson, Equalizing credit opportunity in algorithms: Aligning algorithmic fairness research with us fair lending regulation, in: *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 2022, pp. 357–368.
- [8] Nederlandse Autoriteit Persoonsgegevens, Besluit tot boeteoplegging, 2021. URL: https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/boetebesluit_belastingdienst.pdf.
- [9] Nederlandse Autoriteit Persoonsgegevens, Onderzoeksrapport Belastingdienst/Toeslagen - De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag, 2021. URL: https://www.autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/onderzoek_belastingdienst_kinderopvangtoeslag.pdf.
- [10] N. Nieuws, Ouders zwartgelakte dossiers: 'Ik weet nog steeds niet wat ik fout heb gedaan', *NOS Nieuws* (2019). URL: <https://nos.nl/artikel/2314288-ouders-zwartgelakte-dossiers-ik-weet-nog-steds-niet-wat-ik-fout-heb-gedaan>.
- [11] J. Dastin, Amazon scraps secret ai recruiting tool that showed bias against women, in: *Ethics of Data and Analytics*, Auerbach Publications, 2018, pp. 296–299.
- [12] J. Buolamwini, T. Gebru, Gender shades: Intersectional accuracy disparities in commercial gender classification, in: *Conference on fairness, accountability and transparency*, PMLR, 2018, pp. 77–91.
- [13] College voor de Rechten van de Mens, Tussenoordeel. De Stichting Vrije Universiteit krijgt de gelegenheid om te bewijzen dat de door haar ingezette antispieksoftware een studente met een donkere huidskleur niet heeft gediscrimineerd., 2022. URL: <https://oordelen.mensenrechten.nl/oordeel/2022-146>, Oordeelnummer 2022-146.
- [14] H. Weerts, L. Royakkers, M. Pechenizkiy, Does the end justify the means? on the moral justification of fairness-aware machine learning, *arXiv preprint arXiv:2202.08536* (2022).

[15] P. Westen, The empty idea of equality, *Harvard Law Review* (1982) 537–596.