Beyond Silos: An Interdisciplinary Analysis of Intersectional Discrimination from an EU Perspective

Stephan Wolters [1]

[1] *Universidad Complutense de Madrid, Pl. Menéndez Pelayo, 4, 28040 Madrid, Spain*

Abstract

This article examines intersectional fairness from socio-legal and socio-technical perspectives, focusing on the complexities arising at the intersection of multiple social identities. Intersectionality, initially conceptualized to address overlapping discrimination based on race and gender, is explored in the context of EU legislation and AI systems. The European Union's legal framework, while comprehensive in its approach to anti-discrimination, often falls short in addressing the nuances of intersectional discrimination. Similarly, AI technologies exhibit biases that disproportionately affect marginalized groups, highlighting the limitations of current fairness metrics in addressing intersectional biases. The article discusses various approaches to defining and achieving intersectional fairness in machine learning, emphasizing the challenges of fairness gerrymandering and data sparsity. It advocates for an interdisciplinary approach, calling for inclusive subgroup definitions, strategies to address data gaps, and a focus on equity beyond mere parity. The article underscores the importance of ongoing research and collaboration in understanding and mitigating intersectional discrimination.

Keywords

Law, Artificial Intelligence, EU, Intersectional Discrimination, Fairness Gerrymandering

1. Introduction

Intersectionality, coined by Kimberlé Crenshaw in 1989, is a vital concept in understanding the complexities of discrimination, particularly concerning race and sex [1]. Crenshaw used the DeGraffenreid vs. General Motors[2] case to illustrate the specific challenges black women face, with discrimination at the race and sex intersection often overlooked. This framework acknowledges how various oppressions, like racism and sexism, intersect and compound disadvantages.

Sandra Fredman's analysis of EU discrimination law categorizes experiences as sequential, additive, or intersectional [2]. Sequential discrimination involves separate instances of different discrimination types, while additive discrimination sees multiple types concurrently but independently. Intersectional discrimination, the most complex, involves inseparable, concurrent discriminations creating unique challenges. This intricacy is exemplified in the EU's approach under Article 21 of the EU Charter of Fundamental Rights, which lists 17 protected attributes. The potentially vast number of intersectional combinations[3] reveal the challenges in addressing such complexities legally and technically.

In AI fairness, intersectional biases are significant and stem from systemic societal biases reflected in training data, among many other reasons [3]. These biases often get replicated or magnified in AI systems, as shown in facial recognition technologies where errors are higher for women or those with

EWAF'24: Third European Workshop on Algorithmic Fairness, July 01–03, 2024, Mainz, Germany
EMAIL: swolters@ucm.es

[2] The DeGraffenreid case refers to a landmark lawsuit filed against General Motors (GM) in 1976 by twelve African American women. These women sued GM for gender **and** race discrimination, alleging that the company's seniority system disproportionately disadvantaged black women, effectively excluding them from better-paying jobs. The case highlighted the intersectionality of race and gender discrimination in employment practices. Ultimately, the Supreme Court ruled in favor of GM in 1979, stating that the company's seniority system did not discriminate against the women based on their race **or** gender, not acknowledging the compound effect of intersectional discrimination.

[3] 17 protected attributes from article 21 of the EU Charter of Fundamental Rights in any combination from 2 to 15 attributes could amount to 131,036 combinations, where ethnic or social origin, religion or belief, political or any other opinion, count as different protected attributes. Admittedly, it seems unlikely to have more than 3 or 4 protected intersecting attributes in a particular situation, however, the problem persists even in a reduced context of few protected attributes.

darker skin tones, particularly dark-skinned women [4]. Traditional fairness metrics in AI often overlook the compounded impact of intersecting attributes, thus failing to capture the full scope of discrimination. For example, AI in recruitment processes has been found to disproportionately disadvantage women of color, favoring resumes from predominantly white, male-dominated fields [5].

This article aims to bridge socio-legal and socio-technical perspectives, seeking cross-disciplinary insights while acknowledging its preliminary status in this complex field.

2. EU Legislation and Jurisdiction on Intersectional Discrimination

The evolution of European legal frameworks to address discrimination, especially intersectional discrimination, is marked by foundational documents and evolving treaties. The Universal Declaration of Human Rights (1948) sets the stage with its emphasis on universal human equality and dignity, as outlined in Articles 1 and 2. This is bolstered by the European Convention on Human Rights (1950), particularly through Article 14, which explicitly prohibits discrimination in securing Convention rights, a scope further extended by Protocol 12 in 2005.

Central to the European Union's legal landscape is the Charter of Fundamental Rights (2009), notably Article 21, which explicitly prohibits discrimination on various grounds. This is reinforced by EU Directives like the Racial Equality Directive (2000/43/EC) and the Employment Equality Directive (2000/78/EC), focusing on employment discrimination. The Treaty on European Union (TEU, 1993) and the Treaty on the Functioning of the European Union (TFEU, 2009) further underscore the EU's commitment to combating discrimination (Articles 2, 3, and 10 of the TEU; Articles 18 and 19 of the TFEU).

Despite these frameworks, EU legislation generally treats each ground of discrimination separately, not in an intersectional context [6]. This gap was addressed in the European Parliament's resolution of 6 July 2022, focusing on intersectional discrimination, particularly regarding women of diverse racial backgrounds[4]. It calls for a holistic approach in policy-making and comprehensive impact assessments of legislation. The "Report on the Situation of Fundamental Rights in the European Union" (27 November 2023)[5] reinforces the need to address intersectional discrimination. The report highlights the Council's inaction, indicating the necessity for an urgent enhancement of current EU anti-discrimination laws to include intersectional considerations.

The European Court of Justice (ECJ) and the European Court of Human Rights (ECtHR) have adjudicated numerous discrimination cases[6]. However, they face criticism for their limited acknowledgment of intersectionality, often focusing more on 'multiple discrimination' [7]. This approach has been seen as insufficient to fully grasp the nuances of intersectional discrimination.

In response to these challenges, D. Schiek suggests a reformation of anti-discrimination laws around key 'nodes' like race, gender, and disability [8]. This restructuring aims to recognize and address the overlapping nature of discrimination more effectively, offering a nuanced legal mechanism to deal with intersecting discrimination factors. Such an approach would streamline legal processes and provide a more comprehensive framework for addressing discrimination. Concurrently, I. Solanke [7] introduces the concept of stigma as an essential factor in understanding discrimination in the EU. She argues that traditional legal approaches focusing on individual characteristics like race and sex are limited. Solanke advocates for an anti-stigma principle that

[4] European Parliament resolution of 6 July 2022 on intersectional discrimination in the European Union https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022IP0289

[5] No. 29, 36, and 44 of the report: https://www.europarl.europa.eu/doceo/document/A-9-2023-0376_EN.html

[6] Parris v Trinity College Dublin and Others (C-433/15): The ECJ ruled that a pension scheme's refusal to grant a survivor's pension to the same-sex partner of a scheme member, where the partnership was registered after the member had reached a specified age limit, was not discriminatory. This case was seen as lacking in intersectional consideration, particularly regarding age and sexual orientation.

Achbita v G4S Secure Solutions NV (C-157/15): In this case, the ECJ found that a company policy banning visible signs of political, philosophical, or religious beliefs in the workplace did not constitute direct discrimination. This involved a Muslim woman who was dismissed for wearing a hijab. The ruling was criticized for not adequately addressing the intersection of religion and gender. Cf. Bougnaoui and ADDH v Micropole SA (C-188/15): Similar to Achbita.

B.S. v. Spain (no. 47159/08): The ECHR found a violation of Article 3 (prohibition of inhuman or degrading treatment) and of Article 14 in conjunction with Article 3 of the European Convention on Human Rights in this case. It involved a woman of Nigerian origin who worked as a prostitute and alleged that she was racially abused by authorities on several occasions. The Court found that the Spanish courts had failed to effectively investigate the complaints, particularly not considering the applicant's vulnerability as an African woman working as a prostitute. This case was noteworthy for pointing to possible intersectional discrimination based on race/ethnicity, gender, and profession.

considers the social meanings and synergistic effects of various attributes, potentially providing a more effective legal tool against intersectional discrimination. This principle would account for the socio-cultural power dynamics that underlie discriminatory practices, offering a deeper insight into the complex nature of discrimination.

Both Schiek and Solanke's propositions indicate a need for a legislative paradigm shift to effectively address intersectional discrimination. Their approaches suggest that a more focused legal framework, incorporating concepts like stigma, could more accurately capture the experiences of those facing discrimination on multiple, intersecting grounds. These proposals not only aim to enhance existing legal frameworks but also contribute to the broader debate on AI fairness [9]. In this context, understanding and addressing intersectional considerations are increasingly recognized as critical for developing equitable and just technology.

In summary, the EU's legal evolution demonstrates a growing awareness and response to the complexities of intersectional discrimination. While foundational treaties and directives lay the groundwork for anti-discrimination measures, recent resolutions and reports highlight the need for more nuanced approaches[7][8]. The critiques and suggestions by legal scholars point towards an emerging consensus on the necessity to reshape legal frameworks to better acknowledge and address the interwoven nature of discrimination experiences [10]. These developments are not only pivotal for legislative reforms but also have significant implications for sectors like AI, where fairness and non-discrimination are paramount concerns.

3. The Weak Link between AI Intersectional Fairness and EU Law

Limitations in AI systems are well-known and range from data-related issues like sampling and measurement biases [3][11]; to modeling limitations such as misclassification and overfitting [12]; to user-related issues such as confirmation biases and lack of trust [13]; to design and usability issues like lack of transparency and explainability [14]. These technical challenges intersect with some legal mandates, such as those in the EU AI Act, particularly Article 10 on data quality. However, from an intersectional fairness perspective, many of these limitations are exacerbated and not adequately addressed in EU legislation.

Despite numerous fairness metrics, a generally accepted definition of algorithmic fairness has not been agreed upon [15][16]. Kearns et al. [17] illustrated fairness gerrymandering, where apparent fairness across a group masks discrimination against specific subgroups. This issue parallels the limitations of EU legislation in addressing intersectional discrimination discussed in section 2.

Recent attempts to define intersectional fairness in machine learning include various techniques [18]. Most extend existing concepts of group fairness to multiple subgroups. **Subgroup fairness** adapts group fairness for structured subgroups, allowing some leniency from strict statistical parity [17], but may not fully protect smaller, highly affected subgroups.

Calibration-based fairness in binary prediction tasks emphasizes predictor confidence accuracy. Multicalibration [19] ensures well-calibrated outcomes across subgroups, with multiaccuracy [20] providing a more lenient approach. A hierarchy of multicalibration methods [21] balances fairness strength with computational complexity. **Metric-based fairness** offers solutions for fairness in multiple intersectional groups, allowing minor fairness errors for practical application [22]. This approach ensures similar treatment for subgroups based on inter-individual distances.

Differential fairness [9], inspired by U.S. anti-discrimination laws and differential privacy principles [23], aligns closely with EU anti-discrimination legislation, providing a comprehensive assessment of fairness across all groups but not specifically intersectional subgroups. **Max-Min fairness** applies the Rawlsian principle [24] to maximize utility for the least advantaged groups, though its effectiveness decreases with higher dimensions of intersectionality [25]. **Probabilistic fairness** addresses data gaps in intersectional groups using a differential fairness approach [26], recognizing limitations in detecting high-risk groups not present in the data.

This overview highlights common problems: fairness gerrymandering, which may obscure disparities [27], and the marginalization of minority groups due to disproportionate weighting. These issues underscore the necessity for inclusive stakeholder involvement in developing fairness frameworks. Comprehensive approaches like Max-Min and Differential Fairness face challenges with data sparsity, particularly in highly intersectional or underrepresented groups [18]. This predicament often renders probabilistic methods more pragmatic but does not guarantee fairness for all subgroups.

Nonetheless, this short discussion illustrates that no homogeneous technical framework exists to automate intersectional fairness nor (algorithmic) fairness in general as demonstrated by different researchers [28]. However, the legal sphere might not even be aware of intersectional discrimination without technical support, which reflects the need for ongoing legal and technical collaboration to ensure comprehensive protection against intersectional discrimination.

4. Conclusions and Future Research

The perspectives of both legal and technical domains regarding intersectional fairness encapsulate significant notions pertinent to the pursuit of a more equitable society for marginalized minorities embodying intersecting identities. Nonetheless, a harmonization of these viewpoints appears elusive. Despite disparate approaches, the underlying congruence in the issues underscores the potential for synergy in addressing them, more precisely:

1. **Subgroup Selection and Stakeholder Involvement**: Both AI and legal systems tend to focus on larger, more easily identifiable groups, often overlooking smaller, intersectional identities. This oversight can perpetuate biases and discrimination against these marginalized groups. Involving diverse stakeholders, including representatives from various intersectional groups, legal experts, AI developers, sociologists, and ethicists, could increase visibility representation and protection [29].
2. **Addressing Data and Legal Representation Gaps**: AI and legal frameworks both face significant challenges in adequately representing marginalized subgroups. In AI, data scarcity for these groups leads to biases and inaccuracies in algorithmic outcomes [4]. In the legal realm, limited recognition of intersectional discrimination hinders the development of comprehensive anti-discrimination policies [2]. Joint efforts between these fields can address these gaps by creating more inclusive datasets and legal frameworks that fully recognize and protect intersectional identities.
3. **Achieving True Equity in Fairness and Law**: Both AI and legal systems often aim for parity or equal treatment but can neglect the specific needs of intersectional groups. To achieve true equity, these systems must look beyond simple distributive measures and ensure genuine equity and representation. Interdisciplinary efforts are crucial in developing nuanced approaches that address the unique challenges faced by individuals with intersecting identities [30].
4. **Flexibility in Mitigation and Adaptation**: Both AI and legal systems must be flexible to adapt to the evolving nature of intersectional discrimination. AI faces the challenge of creating algorithms that can adapt to various tasks and contexts (e.g. via transfer learning [31]), while the legal field requires frameworks that can address a broad range of discrimination cases [32]. Versatile and adaptable approaches in both fields are essential for effectively addressing intersectional fairness.
5. **Understanding Bias and Causality**: Understanding how bias propagates through AI systems and legal frameworks is crucial for developing effective strategies to address intersectional fairness. In AI, this involves tracing bias through the machine learning lifecycle [33], while in law, it entails understanding the causal pathways of discrimination [34]. Investigating causal approaches in both disciplines can lead to deeper insights and more robust solutions.
6. **Evaluating Fairness and Legal Measures**: Evaluating fairness in AI via auditing frameworks [35] and assessing the effectiveness of legal anti-discrimination measures are parallel processes that benefit from continuous refinement and the inclusion of diverse perspectives. Both fields must adopt practical evaluation methods to ensure that fairness and anti-discrimination efforts are effective and adaptable to changing societal contexts.

7. **Test Cases and Legal Precedents as Benchmarks**: Testing AI systems for (intersectional) biases [36] and establishing legal precedents or regulatory testbeds for such discrimination are crucial steps in creating benchmarks that guide future progress in both AI and legal sectors.[7]

The primary aim of this article is not to offer a comprehensive exposition on intersectional fairness, but rather to elucidate how forthcoming interdisciplinary endeavors may propel inquiries on both fronts.

5. References

[1] Crenshaw, Kimberle. "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics." University of Chicago Legal Forum 1989, no. 1 (December 7, 2015). https://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8.

[2] Directorate-General for Justice and Consumers (European Commission), European network of legal experts in gender equality and non-discrimination, and Sandra Fredman. Intersectional discrimination in EU gender equality and non-discrimination law. Publications Office of the European Union, 2016. https://data.europa.eu/doi/10.2838/241520.

[3] Suresh, Harini, and John Guttag. "A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle." Equity and Access in Algorithms, Mechanisms, and Optimization, October 5, 2021, 1–9. https://doi.org/10.1145/3465416.3483305.

[4] Buolamwini, Joy, and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," 2018. https://www.semanticscholar.org/paper/Gender-Shades%3A-Intersectional-Accuracy-Disparities-Buolamwini-Gebru/18858cc936947fc96b5c06bbe3c6c2faa5614540.

[5] Kazim, Emre, Adriano Soares Koshiyama, Airlie Hilliard, and Roseline Polle. "Systematizing Audit in Algorithmic Recruitment." Journal of Intelligence 9, no. 3 (September 17, 2021): 46. https://doi.org/10.3390/jintelligence9030046.

[6] May, Vivian M. "'Speaking into the Void'? Intersectionality Critiques and Epistemic Backlash." Hypatia 29, no. 1 (2014): 94–112. https://doi.org/10.1111/hypa.12060.

[7] Solanke, Iyiola. "The EU Approach to Intersectional Discrimination in Law." edited by Gabriele Abels, Andrea Krizsán, Heather MacRae, and Anna Van Der Vleuten, 1st ed., 93–104. Abingdon, Oxon; New York, NY: Routledge, 2021. | Series: Routledge international handbooks: Routledge, 2021. https://doi.org/10.4324/9781351049955-9.

[8] Schiek, Dagmar. "On Uses, Mis-Uses and Non-Uses of Intersectionality before the Court of Justice (EU)." International Journal of Discrimination and the Law 18, no. 2–3 (June 2018): 82–103. https://doi.org/10.1177/1358229118799232.

[9] Foulds, James R., Rashidul Islam, Kamrun Naher Keya, and Shimei Pan. "An Intersectional Definition of Fairness." 2020 IEEE 36th International Conference on Data Engineering (ICDE), April 2020, 1918–21. https://doi.org/10.1109/ICDE48307.2020.00203.

[10] Schiek, D., and A. Lawson. "European Union Non-Discrimination Law and Intersectionality: Investigating the Triangle of Racial, Gender and Disability Discrimination," 2011. https://www.semanticscholar.org/paper/European-Union-Non-Discrimination-Law-and-the-of-Schiek-Lawson/eb7de6864d00acbf576f5a1b27d7e0065a58339f.

[11] Selbst, Andrew D., Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. "Fairness and Abstraction in Sociotechnical Systems." Proceedings of the Conference on Fairness, Accountability, and Transparency, January 29, 2019, 59–68. https://doi.org/10.1145/3287560.3287598.

[12] Mehrabi, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. "A Survey on Bias and Fairness in Machine Learning." ACM Computing Surveys 54, no. 6 (July 31, 2022): 1–35. https://doi.org/10.1145/3457607.

[7] Recently, legal auditing frameworks related to AI and legal testbeds as sandboxes have found their way through the EU AI Act (cf. annex VII Conformity Assessment or Article 57 AI Regulatory Sandboxes); it remains to be seen if they can be applied to intersectional fairness.

[13] Lee, Min Kyung. "Understanding Perception of Algorithmic Decisions: Fairness, Trust, and Emotion in Response to Algorithmic Management." Big Data & Society 5, no. 1 (January 2018): 205395171875668. https://doi.org/10.1177/2053951718756684.
[14] Hohman, Fred, Andrew Head, Rich Caruana, Robert DeLine, and Steven M. Drucker. "Gamut: A Design Probe to Understand How Data Scientists Understand Machine Learning Models." Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, May 2, 2019, 1–13. https://doi.org/10.1145/3290605.3300809.
[15] Pagano, Tiago P., Rafael B. Loureiro, Fernanda V. N. Lisboa, Rodrigo M. Peixoto, Guilherme A. S. Guimarães, Gustavo O. R. Cruz, Maira M. Araujo, et al. "Bias and Unfairness in Machine Learning Models: A Systematic Review on Datasets, Tools, Fairness Metrics, and Identification and Mitigation Methods." Big Data and Cognitive Computing 7, no. 1 (January 13, 2023): 15. https://doi.org/10.3390/bdcc7010015.
[16] Barocas, Solon, Moritz Hardt, and Arvind Narayanan. "Fairness and Machine Learning Limitations and Opportunities," 2018. https://www.semanticscholar.org/paper/Fairness-and-Machine-Learning-Limitations-and-Barocas-Hardt/bae7f0b3448a3eac77886f2a683c0cf9256bb8bf.
[17] Kearns, Michael, Seth Neel, Aaron Roth, and Zhiwei Steven Wu. "Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness." ArXiv, November 14, 2017. https://www.semanticscholar.org/paper/Preventing-Fairness-Gerrymandering%3A-Auditing-and-Kearns-Neel/19930147204c97be4d0964e166e8fe72ac1d6c3d.
[18] Gohar, Usman, and Lu Cheng. "A Survey on Intersectional Fairness in Machine Learning: Notions, Mitigation, and Challenges." In Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, 6619–27. Macau, SAR China: International Joint Conferences on Artificial Intelligence Organization, 2023. https://doi.org/10.24963/ijcai.2023/742.
[19] Hébert-Johnson, Úrsula, Michael P. Kim, Omer Reingold, and G. Rothblum. "Multicalibration: Calibration for the (Computationally-Identifiable) Masses," 2018. https://www.semanticscholar.org/paper/Multicalibration%3A-Calibration-for-the-Masses-H%C3%A9bert-Johnson-Kim/916c816f16e4934e41f09a3ff81a10e5fc4bb459.
[20] Kim, Savina, Stefan Lessmann, Galina Andreeva, and Michael Rovatsos. "Fair Models in Credit: Intersectional Discrimination and the Amplification of Inequity," 2023. https://doi.org/10.48550/ARXIV.2308.02680.
[21] Gopalan, Parikshit, Michael P. Kim, Mihir Singhal, and Shengjia Zhao. "Low-Degree Multicalibration." [object Object], 2022. https://doi.org/10.48550/ARXIV.2203.01255.
[22] Rothblum, G., and G. Yona. "Probably Approximately Metric-Fair Learning," 2018. https://www.semanticscholar.org/paper/Probably-Approximately-Metric-Fair-Learning-Rothblum-Yona/9f00d66392d1a7ed388ee55d53eebe5b3381e36e.
[23] Dwork, Cynthia. "Differential Privacy." edited by Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, 4052:1–12. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. https://doi.org/10.1007/11787006_1.
[24] Rawls, John. Justice as Fairness: A Restatement. Harvard University Press, 2001.
[25] Ghosh, A., Lea Genuit, and Mary Reagan. "Characterizing Intersectional Group Fairness with Worst-Case Comparisons," 2021. https://www.semanticscholar.org/paper/Characterizing-Intersectional-Group-Fairness-with-Ghosh-Genuit/2f93f745625e1e66d1a8d16465c4bf239977f235.
[26] Morina, Giulio, V. Oliinyk, J. Waton, Ines Marusic, and K. Georgatzis. "Auditing and Achieving Intersectional Fairness in Classification Problems." ArXiv, November 4, 2019. https://www.semanticscholar.org/paper/Auditing-and-Achieving-Intersectional-Fairness-in-Morina-Oliinyk/d47a311297ff0ebf44d8206a7cfc9b482cdeac97.
[27] Kong, Youjin. "Are 'Intersectionally Fair' AI Algorithms Really Fair to Women of Color? A Philosophical Analysis." In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 485–94. FAccT '22. New York, NY, USA: Association for Computing Machinery, 2022. https://doi.org/10.1145/3531146.3533114.
[28] Wachter, S., Mittelstadt, B., & Russell, C. (2020). Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3547922

[29] Young, M., Magassa, L., & Friedman, B. (2019). Toward inclusive tech policy design: A method for underrepresented voices to strengthen tech policy documents. *Ethics and Information Technology*, *21*(2), 89–103. https://doi.org/10.1007/s10676-019-09497-z

[30] Barocas, S., Hardt, M., & Narayanan, A. (2023). Fairness and Machine Learning: Limitations and Opportunities. MIT Press.

[31] Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A Comprehensive Survey on Transfer Learning. Proceedings of the IEEE, 109(1), 43–76. https://doi.org/10.1109/JPROC.2020.3004555

[32] Boddie, E. C. (2015). Adaptive Discrimination (SSRN Scholarly Paper 2803452). https://papers.ssrn.com/abstract=2803452

[33] Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*. Wiley.

[34] Delgado, R., Stefancic, J., & Harris, A. (2017). *Critical Race Theory*. NYU Press.

[35] Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 33–44. https://doi.org/10.1145/3351095.3372873

[36] Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S., Mojsilovic, A., Nagar, S., Ramamurthy, K. N., Richards, J., Saha, D., Sattigeri, P., Singh, M., Varshney, K. R., & Zhang, Y. (2018). *AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias* (arXiv:1810.01943). arXiv. https://doi.org/10.48550/arXiv.1810.01943