

Diskriminierung aus Sicht von Philosophie und Informatik

Hauptseminar im Sommersemester 2021 – Themen und Einstiegsliteratur

1. Limitations of common fairness notions

- *Fairness, Equality, and Power in Algorithmic Decision-Making*. Maximilian Kasy, Rediet Abebe. 2021.
<https://dl.acm.org/doi/pdf/10.1145/3442188.3445919>
- *On the Moral Justification of Statistical Parity*. Corinna Hertweck, Christoph Heitz, Michele Loi. 2021.
<https://dl.acm.org/doi/pdf/10.1145/3442188.3445936>

2. Practical trade-offs of fairness notions

- *On the Applicability of ML Fairness Notions*. Karima Makhoulf, Sami Zhioua, Catuscia Palamidessi. 2020.
<https://arxiv.org/pdf/2006.16745>
- *Dropping Standardized Testing for Admissions: Differential Variance and Access*. Nikhil Garg, Hannah Li, Faidra Monachou. 2020.
<https://arxiv.org/pdf/2010.04396>

3. Unveiling unfairness from datasets

- *Missing the missing values: The ugly duckling of fairness in machine learning*. Martínez-Plumed Fernando, Ferri Cèsar, Nieves David, Hernández-Orallo José. 2021.
<https://onlinelibrary.wiley.com/doi/epdf/10.1002/int.22415>
- *"Ignorance and Prejudice" in Software Fairness*. Jie M. Zhang, Mark Harman. 2021.
https://discovery.ucl.ac.uk/id/eprint/10123814/1/ICSE2021_Ignorance_Prejudice%20%285%29.pdf
- *fairmodels: A Flexible Tool For Bias Detection, Visualization, And Mitigation*. Jakub Wiśniewski, Przemysław Biecek. 2021.
<https://arxiv.org/pdf/2104.00507>

4. **Can we make algorithms neutral?**

- *Are Algorithms Value-Free? Feminist Theoretical Virtues in Machine Learning.* Gabrielle M. Johnson. 2021. <https://philpapers.org/archive/JOHAAV.pdf>
- *The Structure of Bias.* Gabrielle M. Johnson. 2020. <https://academic.oup.com/mind/article-pdf/129/516/1193/36658767/fzaa011.pdf>

5. **The relevance of false positive/negative rates**

- *Fairness in machine learning: Against false positive rate equality as a measure of fairness.* Robert Long. 2020. <https://arxiv.org/pdf/2007.02890>
- *Measuring Algorithmic Fairness.* Deborah Hellman. 2020. <https://www.torontosri.ca/s/Measuring-Algorithmic-Fairness-print.pdf>

6. **The proxy problem and intransparent discrimination**

- *Algorithmic bias: on the implicit biases of social technology.* Gabrielle M. Johnson. 2020. <https://link.springer.com/content/pdf/10.1007/s11229-020-02696-y.pdf>
- *Intransparente Diskriminierung durch maschinelles Lernen.* Heiner Koch. 2020. https://www.praktische-philosophie.org/uploads/8/0/5/2/80520134/zfpp_1_2020_koch.pdf