References:

Friedler, Sorelle A.; Scheidegger, Carlos; Venkatasubramanian, Suresh; Choudhary, Sonam; Hamilton, Evan P.; Roth, Derek, A comparative study of fairness-enhancing interventions in machine learning, eprint arXiv:1802.04422, 2018

Indre Zliobaite. 2017. Measuring discrimination in algorithmic decision making. ˇ Data Mining and Knowledge Discovery 31, 4 (July 2017), 1060–1089.

Danielle Ensign, Sorelle A. Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian. 2018. Runaway Feedback Loops in Predictive Policing. In 1st Conference on Fairness, Accountability and Transparency in Computer Science (FAT*).

Shahin Jabbari, Matthew Joseph, Michael Kearns, Jamie Morgenstern, and Aaron Roth. 2017. Fairness in Reinforcement Learning. In PMLR. 1617–1626.

Matthew Joseph, Michael Kearns, Jamie Morgenstern, Seth Neel, and Aaron Roth. 2016. Fair Algorithms for Infinite and Contextual Bandits. arXiv:1610.09559 [cs] (2016).

Michael Feldman, Sorelle A. Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. 2015. Certifying and removing disparate impact. Proc. 21st ACM KDD (2015), 259–268.

Toon Calders and Sicco Verwer. 2010. Three Naive Bayes Approaches for Discrimination-Free Classification. Data Mining journal; special issue with selected papers from ECML/PKDD (2010).

Flavio Calmon, Dennis Wei, Bhanukiran Vinzamuri, Karthikeyan Natesan Ramamurthy, and Kush R Varshney. 2017. Optimized PreProcessing for Discrimination Prevention. In Advances in Neural Information Processing Systems 30, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 3995–4004.

F. Kamiran and T. Calders. 2012. Data preprocessing techniques for classification without discrimination. Knowledge and Information Systems 33 (2012), 1–33.

Toshihiro Kamishima, Shotaro Akaho, Hideki Asoh, and Jun Sakuma. 2012. Fairness-aware Classifier with Prejudice Remover Regularizer. Machine Learning and Knowledge Discovery in Databases (2012), 35–50.

Blake Woodworth, Suriya Gunasekar, Mesrob I Ohannessian, and Nathan Srebro. 2017. Learning non-discriminatory predictors. arXiv preprint arXiv:1702.06081 (2017).

Moritz Hardt, Eric Price, Nati Srebro, et al. 2016. Equality of opportunity in supervised learning. In Advances in neural information processing systems. 3315–3323.

Marc Böhlen, Varun Chandola, Amol Salunkhe 2017. Server server in the cloud, who is the fairest in the crowd? https://arxiv.org/abs/1711.08801v1