Fairness of Machine Learning in Search Engines

Yi Fang*
yfang@scu.edu
Santa Clara University
Santa Clara, CA, USA

Hongfu Liu
hongfuliu@brandeis.edu
Brandeis University
Waltham, MA, USA

Zhiqiang Tao
zhiqiang.tao@rit.edu
Rochester Institute of Technology
Rochester, NY, USA

Mikhail Yurochkin
mikhail.yurochkin@ibm.com
IBM Research and MIT-IBM Watson AI lab
Cambridge, MA, USA

ABSTRACT

Fairness has gained increasing importance in a variety of AI and machine learning contexts. As one of the most ubiquitous applications of machine learning, search engines mediate much of the information experiences of members of society. Consequently, understanding and mitigating potential algorithmic unfairness in search have become crucial for both users and systems. In this tutorial, we will introduce the fundamentals of fairness in machine learning, for both supervised learning such as classification and ranking, and unsupervised learning such as clustering. We will then present the existing work on fairness in search engines, including the fairness definitions, evaluation metrics, and taxonomies of methodologies. This tutorial will help orient information retrieval researchers to algorithmic fairness, provide an introduction to the growing literature on this topic, and gathering researchers and practitioners interested in this research direction.

CCS CONCEPTS

• **Information systems → Information retrieval**; • **Computing methodologies → Machine learning**.

KEYWORDS

Fairness; Search Engines; Machine Learning

ACM Reference Format:

Yi Fang, Hongfu Liu, Zhiqiang Tao, and Mikhail Yurochkin. 2022. Fairness of Machine Learning in Search Engines. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22), October 17–21, 2022, Atlanta, GA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3511808.3557501

1 MOTIVATION

With the widespread use of machine learning in our society, accounting for fairness has gained significant importance in designing and engineering of such systems. Specifically, search engines have been playing a crucial role on assisting human information access in our everyday lives. Addressing fairness concerns around search systems becomes necessary for increasing the trust among end users [18]. As highly data driven systems, search engines could be significantly affected by data and algorithmic bias, thus yielding unfair results. The tutorial will cover the fundamental concepts, theories, and methods to address the issue from the perspective of machine learning.

2 OBJECTIVES

This tutorial aims to help participants achieve the following goals:

- Gain the basic knowledge of fairness research in both supervised and unsupervised learning settings.
- Be familiar with the taxonomies of fairness methods in search engines.
- Understand the existing metrics and evaluation protocols for fairness-aware systems.
- Identify challenges and new research questions in the fairness field.

3 TUTORIAL OUTLINE

This is a half-day tutorial in lecture format with a comprehensive survey about fairness of machine learning in search engines. In particular, this tutorial covers the following topics.

3.1 Background and Introduction

We will start with an introduction to the general area of research and then discuss the motivation of fairness in search engines. This will be contextualized within the larger area of study of bias and fairness in machine learning and AI ethics.

3.2 Fairness in Supervised Learning

Algorithmic fairness studies methods for understanding, detecting, and mitigating biases in ML/AI systems. The field of algorithmic fairness was initially developed in the context of supervised learning. At a high level, there are two fairness perspectives: group fairness [13] and individual fairness [17]. The former formalizes the idea that ML systems should treat certain groups of individuals similarly, e.g., requiring the average loan approval rate for applicants of different ethnicities be similar. The latter asks for similar treatment of similar individuals, e.g., the same outcome for applicants with resumes that differ only in names. In this tutorial, we

*The authors are listed in alphabetic order of the surnames.

CIKM '22, October 17–21, 2022, Atlanta, GA, USA

ACM ISBN 978-1-4503-9236-5/22/10…$15.00
https://doi.org/10.1145/3511808.3557501

will cover the key concepts of algorithmic fairness in supervised learning supporting them with data examples and demonstrations. In practice, with an existing trained model, the first step is to audit this model for fairness violations. Then we describe methods for training fair models. Finally, we discuss fairness post-processing approaches that can be used on any pre-trained model.

Group Fairness. Identifying group fairness violations in practice is fairly simple when one has access to the protected attributes of interest such as gender or race. The challenge, however, is to identify an appropriate definition of group fairness for the application, as some of them are known to be conflicting [12, 25]. We will demonstrate examples of identifying group fairness violations using popular Demographic Parity and Equalized Odds [22] definitions. We will also interpret group fairness in the context of the 80% rule published by the US Equal Employment Opportunity Commission and highlight the conflicting nature of the considered definitions. To train group fair predictors we will use AIF360 software [3] and demonstrate approaches to interpreting group fairness as an adversarial objective [61] and as a constrained optimization problem [1]. Finally, we discuss several post-processing approaches [22, 49].

Individual Fairness. The key challenge in building a practical individual fairness pipeline is the selection of an appropriate fair metric. Thus, we discuss several approaches for learning an individual fairness metric from data using various types of supervision [23, 35]. Equipped with the fair metric, individual fairness can be interpreted as an adversarial robustness problem [55]. Using this perspective, we discuss methods for auditing models for violations of individual fairness [33, 52] and training individually fair predictors [55, 56]. For post-processing, we discuss a recent method based on the graph signal processing formalism [38]. We interpret individual fairness in the context of the celebrated study of racial discrimination in the US labor market [4] and discuss the complementary and conflicting nature of the group and individual fairness perspectives. In our demonstrations of the individual fairness methods, we will use the inFairness[1] software.

3.3 Fairness in Unsupervised Learning

Fair clustering aims to hide sensitive attributes during data partition by balancing the distribution of protected subgroups in each cluster. Chierichettiss et al. [11] propose a pioneering fairlet method by employing a pre-processing technique to partition original data into chunks, followed by a k-center-based algorithm, which encourages clusters with balanced demographic groups. In light of the expensive computation involved in fairlet, Backurs et al. [2] provide a scalable fair clustering algorithm with approximate fairlet decomposition that runs in nearly linear time. Recently, some in-processing methods have been proposed to jointly learn the representation and achieve fair clustering in a deep fashion. Wang et al. [47] propose learning a fair embedding by forcing the cluster centers to be equidistant from group centers. Li et al. [31] make a step forward to explore fair clustering on visual data by achieving fairness through adversarial training. Zhang et al. [62] generate fair pseudo cluster assignments and boost the clustering performance via contrastive learning.

Fair outlier detection is another crucial topic in unsupervised fairness learning. FairLOF [16] is the first paper to address the fair outlier detection problem, which incorporates a corrective term on the baseline LOF algorithm, in regards to local sensitive subgroup diversity and global outlier alignment with the baseline. FairOD [42] targets an equal outlier rate on the majority and minority-sensitive subgroups. Specifically, it employs an autoencoder as the base outlier detector and performs subgroup debiasing with statistical parity fairness constraint, while maintaining fidelity to within-group rankings with respect to the baseline. Recently, DCFOD [46] adopts representation learning and fairness-adversarial training, with a novel dynamic weight in the regulation of negative impacts from outlier points, to obtain a downstream task-favorable representation while simultaneously ensuring improvement in fairness.

Beyond introducing and summarizing the existing studies in unsupervised fairness learning, we will extract the common techniques to achieve fairness including reweigh, kernel, adversarial training, balancing constraints, and analyze their strengths and weaknesses for practical use.

3.4 Fairness in Search

Search engines are concerned with ranking documents given a query. Fairness in ranking has so far received less attention than fairness in other machine learning settings such as classification [26]. One popular fairness notion in ranking is fairness of exposure [6, 34, 39, 43, 44, 53, 58]. It assumes that the exposure received by a group of items (or an individual item) should be in proportion to its utility. Some other works focused on achieving sufficient representation of documents or items from different groups in the top-k positions of a ranking [9, 21, 57]. In addition to group fairness, individual fairness has been investigated in the context of ranking [8, 54].

The existing work can also be categorized into pre-processing, in-processing, and post-processing methods [60]. Pre-processing approaches such as iFair [27] tackle biases in the training data. In-processing fair ranking methods such as DELTR [58], Fair-PG-Rank [44], and Pairwise Ranking Fairness [5] extend the objective function of a learning-to-rank algorithm by a fairness term. Post-processing algorithms assume that a ranking model has already been trained. A predicted ranking is handed to the algorithm, which re-orders items to improve fairness. The representative methods include FA*IR [57], Equity of Attention [6], Fair Ranking at LinkedIn [21], and others [43, 59]. Most algorithms operate on a notion of group membership, where certain groups are denoted as protected, while one group is denoted as non-protected. We will present the fairness definitions, formulations, and algorithms of different approaches, and discuss their respective advantages and disadvantages. We will also introduce common evaluation metrics and datasets. In addition to the measures for assessing relevance, diversity, and novelty, we will focus on the metrics for quantifying bias and fairness.

3.5 Opportunities and Future Directions

In accordance with the rapid development of AutoML libraries in the industry [28, 36], the proposed tutorial will particularly focus on recent studies that incorporate fairness metrics into a series

[1]https://github.com/IBM/inFairness

of automated research problems, including pipeline design, hyperparameter optimization (HPO), and neural architecture search (NAS). Among existing "AutoFair" works, the HPO-based methods stand out as hyperparameters naturally play a key role in mitigating data bias [48], balancing fairness and utility terms [15], and enabling multi-objective selection [40]. Besides, some other interesting works include fair AutoML [50], fair Bayesian optimization [37], and FairNAS [14]. The proposed tutorial will discuss the above methods from the perspective of generalized hyperparameter choices, which automatically combine data engineering, model selection, and post-processing in sequence to render fair results.

Meta learning (*learning to learn*) falls in the same vein of AutoML, and has received significant research attention in recent years, such as fair meta learner [64], fairness-aware online meta learner [63], Fair-MAML [45], etc. Spotlight to search relevant problems, the AutoDebias [10] method leverages a set of uniformly-distributed data to optimize the meta-learner under a unified risk discrepancy, accounting for various biases in a recommender system. For another example, the meta-weight network [48] realizes equal exposures oriented to protected attributes through dynamically tuning loss weights under the guidance of a fair meta set. The tutorial will also extensively discuss the potential challenges and solutions of applying meta learning to achieve fair rankings.

4 RELATION TO PREVIOUS TUTORIALS

We note that there have been similar tutorials related to investigating fairness issues in other venues including the following: Addressing Bias and Fairness in Search Systems at SIGIR 2021 [20], Fairness of Machine Learning in Recommender Systems at CIKM 2021 [32], Fair Graph Mining at CIKM 2021 [24], and Gender Fairness in Information Retrieval Systems at SIGIR 2022 [7]. The previous tutorials can be considered complementary and synergistic to the theme of our proposed tutorial. The major difference between this tutorial and the previous ones is that they either only focus on recommendation systems or consider fairness mainly from user study and evaluation perspectives. This tutorial focuses on fairness in search tasks from the machine learning perspective.

5 INTENDED AUDIENCE AND RELEVANCE

The tutorial is intended for graduate students, researchers, and practitioners in information retrieval and data mining. The tutorial will also attract researchers who work in broader ML/AI communities especially AI Ethics, as participants will learn the fundamentals of fairness in machine learning in general and in search engines in particular. Moreover, the tutorial will attract industry researchers and practitioners from different areas, since fairness has become an important concern in many real-world applications. This tutorial will be closely connected to the existing fairness work at the major IR, DM, ML, and AI conferences.

6 PRESENTERS

Yi Fang[2] is an Associate Professor in the Department of Computer Science and Engineering at Santa Clara University. His research interests broadly lie in information retrieval and machine learning. His recent research on fairness include fair learning to rank [48], achieving outcome fairness in machine learning [19], and advancing diversity, equity, and inclusion (DEI) in news media [41]. He has served as Senior PC members in various IR and AI-related conferences such as SIGIR, CIKM, WSDM, and AAAI, and he is serving as the Chair of the Steering Committee for ACM ICTIR. He also served as the Tutorial Chair for IEEE BigData 2016. He received the Outstanding Service Award at CIKM 2013.

Hongfu Liu[3] is an Assistant Professor of Computer Science at Brandeis University. His research interests lie in core machine learning, especially fairness learning, visual semantic learning and graph learning, and AI-assisted applications. He has published over 70 papers (e.g., KDD, NeurIPS, ICLR, ICML, IJCAI, AAAI, CIKM, CVPR, ICCV, TPAMI, and TKDE). He has also won several awards including the First Place Award in MS-Celel-1M Grand Challenge in ICCV 2017, the NVIDIA CCS Best Student Paper Award in FG 2021, the 2021 INNS Aharon Katzir Young Investigator Award, the top reviewer in UAI 2022, the highlighted Area Chair in ICLR 2022, and the 2022 Global Top-25 Chinese Young Scholars in AI (Data Mining Area) by Baidu Scholar. He has served as an Associate Editor of IEEE CIM and as an Area Chair of ICLR and NeurIPS. His recent work on fairness includes deep fair clustering [31], dyadic fairness on link prediction [30], fair outlier detection [46], fair feature selection [51] and studies on the trade-off between utility and fairness [29].

Zhiqiang Tao[4] is an Assistant Professor in the School of Information at Rochester Institute of Technology (RIT). Prior joining RIT, he was an Assistant Professor in the Department of Computer Science and Engineering at Santa Clara University. His research interests are machine learning, data mining, and computer vision, with a particular interest in AutoML, uncertainty estimation, and hyperparameter optimization. He has published around 40 peer-reviewed papers in leading journals and conferences, including TPAMI, TNNLS, TIP, TCYB, TKDD, NeurIPS, ICLR, KDD, SIGIR, CVPR, ICCV, ECCV, AAAI, IJCAI, CIKM, ICDM, SDM, etc. He serves as the Associate Editor of Neurocomputing, and he also has served as reviewers and PC members for prestige journals and international conferences. He won the 3rd place award in KDD Cup AutoML track in 2019.

Mikhail Yurochkin[5] is a Research Staff Member at IBM Research and MIT-IBM Watson AI Lab in Cambridge, Massachusetts. He is interested in developing the methodology for algorithmic fairness and other topics pertaining to the safe and inclusive adoption of AI and ML in practice. Before joining IBM, he completed a Ph.D. in Statistics at the University of Michigan, where he worked with Long Nguyen. Mikhail received his bachelor's degree in applied mathematics and physics from the Moscow Institute of Physics and Technology.

7 AVAILABILITY OF MATERIALS

We will make a Github repository publicly available before the conference so that the participants of the tutorial can familiarize themselves with the content. The repository will include a comprehensive slide deck and links to relevant resources.

[2]https://www.cse.scu.edu/~yfang/

[3]http://hongfuliu.com/

[4]http://ztao.cc/

[5]https://moonfolk.github.io/

8 ACKNOWLEDGEMENT

This work is supported in part by the Ciocca center research award at Santa Clara University. Mikhail thanks Mayank Agarwal, Onkar Bhardwaj, and Aldo Pareja for their help in preparing demonstrations for this tutorial.

REFERENCES

[1] Alekh Agarwal, Alina Beygelzimer, Miroslav Dudík, John Langford, and Hanna Wallach. 2018. A reductions approach to fair classification. In *ICML*.

[2] Arturs Backurs, Piotr Indyk, Krzysztof Onak, Baruch Schieber, Ali Vakilian, and Tal Wagner. 2019. Scalable fair clustering. In *ICML*.

[3] Rachel KE Bellamy, Kuntal Dey, Michael Hind, Samuel C Hoffman, Stephanie Houde, Kalapriya Kannan, Pranay Lohia, Jacquelyn Martino, Sameep Mehta, Aleksandra Mojsilović, et al. 2019. AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development* 63, 4/5 (2019), 4–1.

[4] Marianne Bertrand and Sendhil Mullainathan. 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review* 94, 4 (2004), 991–1013.

[5] Alex Beutel, Jilin Chen, Tulsee Doshi, Hai Qian, Li Wei, Yi Wu, Lukasz Heldt, Zhe Zhao, Lichan Hong, Ed H Chi, et al. 2019. Fairness in recommendation ranking through pairwise comparisons. In *SIGKDD*.

[6] Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. 2018. Equity of attention: Amortizing individual fairness in rankings. In *SIGIR*.

[7] Amin Bigdeli, Negar Arabzadeh, Shirin SeyedSalehi, Morteza Zihayat, and Ebrahim Bagheri. 2022. Gender Fairness in Information Retrieval Systems. In *SIGIR*.

[8] Amanda Bower, Hamid Eftekhari, Mikhail Yurochkin, and Yuekai Sun. 2021. Individually Fair Ranking. In *ICLR*.

[9] L Elisa Celis, Damian Straszak, and Nisheeth K Vishnoi. 2017. Ranking with fairness constraints. *arXiv preprint arXiv:1704.06840* (2017).

[10] Jiawei Chen, Hande Dong, Yang Qiu, Xiangnan He, Xin Xin, Liang Chen, Guli Lin, and Keping Yang. 2021. AutoDebias: Learning to Debias for Recommendation. In *SIGIR*.

[11] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. 2017. Fair clustering through fairlets. In *NeurIPS*.

[12] Alexandra Chouldechova. 2017. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data* 5, 2 (2017), 153–163.

[13] Alexandra Chouldechova and Aaron Roth. 2020. A snapshot of the frontiers of fairness in machine learning. *Commun. ACM* 63, 5 (2020), 82–89.

[14] Xiangxiang Chu, Bo Zhang, and Ruijun Xu. 2021. FairNAS: Rethinking Evaluation Fairness of Weight Sharing Neural Architecture Search. In *ICCV*.

[15] André F. Cruz, Pedro Saleiro, Catarina Belém, Carlos Soares, and Pedro Bizarro. 2021. Promoting Fairness through Hyperparameter Optimization. In *ICDM*.

[16] P Deepak and Savitha Sam Abraham. 2020. Fair Outlier Detection. *arXiv preprint arXiv:2005.09900*.

[17] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *ITCS*.

[18] Michael D Ekstrand, Anubrata Das, Robin Burke, and Fernando Diaz. 2022. Fairness in Information Access Systems. *Foundations and Trends® in Information Retrieval* 16, 1-2 (2022), 1–177.

[19] Boli Fang, Miao Jiang, Pei-yi Cheng, Jerry Shen, and Yi Fang. 2020. Achieving Outcome Fairness in Machine Learning Models for Social Decision Problems.. In *IJCAI*.

[20] Ruoyuan Gao and Chirag Shah. 2021. Addressing bias and fairness in search systems. In *SIGIR*.

[21] Sahin Cem Geyik, Stuart Ambler, and Krishnaram Kenthapadi. 2019. Fairness-aware ranking in search and recommendation systems with application to Linkedin talent search. In *SIGKDD*.

[22] Moritz Hardt, Eric Price, and Nati Srebro. 2016. Equality of opportunity in supervised learning. In *NeurIPS*.

[23] Christina Ilvento. 2019. Metric learning for individual fairness. *arXiv preprint arXiv:1906.00250* (2019).

[24] Jian Kang and Hanghang Tong. 2021. Fair graph mining. In *CIKM*.

[25] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. 2016. Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807* (2016).

[26] Nikola Konstantinov and Christoph H Lampert. 2021. Fairness through regularization for learning to rank. *arXiv preprint arXiv:2102.05996* (2021).

[27] Preethi Lahoti, Krishna P Gummadi, and Gerhard Weikum. 2019. ifair: Learning individually fair data representations for algorithmic decision making. In *ICDE*.

[28] Erin LeDell and Sebastien Poirier. 2020. H2o automl: Scalable automatic machine learning. In *AutoML Workshop at ICML*.

[29] Peizhao Li and Hongfu Liu. 2022. Achieving Fairness at No Utility Cost via Data Reweighing. *arXiv preprint arXiv:2202.00787* (2022).

[30] Peizhao Li, Yifei Wang, Han Zhao, Pengyu Hong, and Hongfu Liu. 2020. On dyadic fairness: Exploring and mitigating bias in graph connections. In *ICLR*.

[31] Peizhao Li, Han Zhao, and Hongfu Liu. 2020. Deep fair clustering for visual learning. In *CVPR*.

[32] Yunqi Li, Yingqiang Ge, and Yongfeng Zhang. 2021. Tutorial on fairness of machine learning in recommender systems. In *SIGIR*.

[33] Subha Maity, Songkai Xue, Mikhail Yurochkin, and Yuekai Sun. 2021. Statistical inference for individual fairness. *arXiv preprint arXiv:2103.16714* (2021).

[34] Marco Morik, Ashudeep Singh, Jessica Hong, and Thorsten Joachims. 2020. Controlling fairness and bias in dynamic learning-to-rank. In *SIGIR*.

[35] Debarghya Mukherjee, Mikhail Yurochkin, Moulinath Banerjee, and Yuekai Sun. 2020. Two simple ways to learn individual fairness metrics from data. In *ICML*.

[36] Deepak Mukunthu, Parashar Shah, and Wee Hyong Tok. 2019. *Practical Automated Machine Learning on Azure: Using Azure Machine Learning to Quickly Build AI Solutions*. O'Reilly Media.

[37] Valerio Perrone, Michele Donini, Krishnaram Kenthapadi, and Cédric Archambeau. 2020. Fair Bayesian optimization. In *AutoML at ICML*.

[38] Felix Petersen, Debarghya Mukherjee, Yuekai Sun, and Mikhail Yurochkin. 2021. Post-processing for individual fairness. *NeurIPS* (2021).

[39] Piotr Sapiezynski, Wesley Zeng, Ronald E Robertson, Alan Mislove, and Christo Wilson. 2019. Quantifying the impact of user attentionon fair group representation in ranked lists. In *WWW*.

[40] Robin Schmucker, Michele Donini, Valerio Perrone, and Cédric Archambeau. 2020. Multi-objective multi-fidelity hyperparameter optimization with application to fairness. In *Meta-learning Workshop at NeurIPS*.

[41] Xiaoxiao Shang, Zhiyuan Peng, Qiming Yuan, Sabiq Khan, Lauren Xie, Yi Fang, and Subramaniam Vincent. 2022. DIANES: A DEI Audit Toolkit for News Sources. In *SIGIR (Demo Track)*.

[42] Shubhranshu Shekhar, Neil Shah, and Leman Akoglu. 2021. Fairod: Fairness-aware outlier detection. In *AIES*.

[43] Ashudeep Singh and Thorsten Joachims. 2018. Fairness of exposure in rankings. In *SIGKDD*.

[44] Ashudeep Singh and Thorsten Joachims. 2019. Policy learning for fairness in ranking. In *NeurIPS*.

[45] Dylan Slack, Sorelle A Friedler, and Emile Givental. 2020. Fairness warnings and Fair-MAML: learning fairly with minimal data. In *FAccT*.

[46] Hanyu Song, Peizhao Li, and Hongfu Liu. 2021. Deep Clustering based Fair Outlier Detection. In *SIGKDD*.

[47] Bokun Wang and Ian Davidson. 2019. Towards fair deep clustering with multi-state protected variables. *arXiv preprint arXiv:1901.10053* (2019).

[48] Yuan Wang, Zhiqiang Tao, and Yi Fang. 2022. A Meta-learning Approach to Fair Ranking. In *SIGIR*.

[49] Dennis Wei, Karthikeyan Natesan Ramamurthy, and Flavio P Calmon. 2021. Optimized Score Transformation for Consistent Fair Classification. *Journal of Machine Learning Research* 22 (2021), 258–1.

[50] Qingyun Wu and Chi Wang. 2021. Fair AutoML. *CoRR* abs/2111.06495 (2021). arXiv:2111.06495 https://arxiv.org/abs/2111.06495

[51] Xiaoying Xing, Hongfu Liu, Chen Chen, and Jundong Li. 2021. Fairness-Aware Unsupervised Feature Selection. *arXiv preprint arXiv:2106.02216* (2021).

[52] Songkai Xue, Mikhail Yurochkin, and Yuekai Sun. 2020. Auditing ml models for individual bias and unfairness. In *AI&STAT*.

[53] Himank Yadav, Zhengxiao Du, and Thorsten Joachims. 2020. Fair learning-to-rank from implicit feedback. In *SIGIR*.

[54] Ke Yang, Vasilis Gkatzelis, and Julia Stoyanovich. 2019. Balanced ranking with diversity constraints. In *IJCAI*.

[55] Mikhail Yurochkin, Amanda Bower, and Yuekai Sun. 2019. Training individually fair ML models with sensitive subspace robustness. *arXiv preprint arXiv:1907.00020* (2019).

[56] Mikhail Yurochkin and Yuekai Sun. 2020. Sensei: Sensitive set invariance for enforcing individual fairness. *arXiv preprint arXiv:2006.14168* (2020).

[57] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. 2017. Fa*ir: A fair top-k ranking algorithm. In *CIKM*.

[58] Meike Zehlike and Carlos Castillo. 2020. Reducing disparate exposure in ranking: A learning to rank approach. In *WWW*.

[59] Meike Zehlike, Philipp Hacker, and Emil Wiedemann. 2020. Matching code and law: achieving algorithmic fairness with optimal transport. *Data Mining and Knowledge Discovery* 34, 1 (2020), 163–200.

[60] Meike Zehlike, Ke Yang, and Julia Stoyanovich. 2021. Fairness in ranking: A survey. *arXiv preprint arXiv:2103.14000* (2021).

[61] Brian Hu Zhang, Blake Lemoine, and Margaret Mitchell. 2018. Mitigating unwanted biases with adversarial learning. In *AIES*.

[62] Hongjing Zhang and Ian Davidson. 2021. Deep Fair Discriminative Clustering. *arXiv preprint arXiv:2105.14146* (2021).

[63] Chen Zhao, Feng Chen, and Bhavani Thuraisingham. 2021. Fairness-Aware Online Meta-Learning. In *SIGKDD*.

[64] Chen Zhao, Feng Chen, Zhuoyi Wang, and Latifur Khan. 2020. A primal-dual subgradient approach for fair meta learning. In *ICDM*.