Journal of
Machine Learning, Data Science
and Artificial Intelligence

**P-ISSN:** xxxx-xxxx
**E-ISSN:** xxxx-xxxx
JMLDSAI 2024; 1(1): 06-09
[www.datasciencejournal.net](http://www.datasciencejournal.net)
Received: 12-05-2024
Accepted: 18-06-2024

**Mohammad Kamal Hossain**
Department of Computer Science & Engineering, Shahjalal University of Science and Technology (SUST), Sylhet, Bangladesh

**Arifa Akhter**
Department of Computer Science & Engineering, Shahjalal University of Science and Technology (SUST), Sylhet, Bangladesh

**Corresponding Author:**
**Mohammad Kamal Hossain**
Department of Computer Science & Engineering, Shahjalal University of Science and Technology (SUST), Sylhet, Bangladesh

# Assessing fairness in large-scale recommendation systems through intersectional metrics

## Mohammad Kamal Hossain and Arifa Akhter

### Abstract
In recent years, large-scale recommendation systems have become integral to digital platforms such as e-commerce, social media, and content streaming. While these systems have demonstrated remarkable capabilities in personalizing user experiences, concerns about fairness and bias have become increasingly prominent. Traditional fairness metrics often fail to capture complex social realities, especially those affecting users at the intersection of multiple marginalized identities. This review examines the emerging research on assessing fairness in recommendation systems through intersectional metrics. We explore theoretical foundations, algorithmic frameworks, empirical evaluations and ongoing challenges. The paper highlights how intersectionality offers a richer lens for understanding systemic biases in recommendations, ultimately pushing the field toward more equitable and inclusive systems.

**Keywords:** Assessing fairness, multiple marginalized identities, ongoing challenges, algorithmic frameworks, empirical evaluations, inclusive systems

### Introduction
Recommendation systems power much of the digital economy, influencing what users buy, read, watch, and consume. Companies such as Amazon, Netflix, YouTube, Facebook, and Spotify deploy large-scale recommendation engines that analyze massive datasets to personalize content and maximize user engagement. While these systems enhance user satisfaction and drive profitability, they also introduce significant ethical and social challenges. Increasingly, researchers, regulators, and civil society organizations have raised concerns about how these algorithms may propagate, amplify, or even create biases that disproportionately affect vulnerable populations. Fairness in recommendation systems has thus emerged as a critical area of inquiry within machine learning, artificial intelligence, and information systems. Much of the initial work in algorithmic fairness focused on supervised learning tasks such as classification and regression. In contrast, recommendation systems introduce unique fairness challenges due to their personalized, dynamic, and often opaque nature. Traditional fairness metrics such as demographic parity, equalized odds, or disparate impact provide some guidance but are often insufficient in the recommendation context. In particular, they tend to treat sensitive attributes independently, failing to account for the ways in which multiple social identities intersect to produce unique forms of disadvantage.

Intersectionality, originally coined by Kimberlé Crenshaw in the context of legal studies, offers a more nuanced framework. It recognizes that individuals often experience discrimination and disadvantage as a result of overlapping identities, such as gender, race, disability, and socioeconomic status. In recommendation systems, failure to incorporate intersectional perspectives can lead to algorithmic outcomes that systematically marginalize users who belong to multiple disadvantaged groups, even if no single group appears to be underrepresented when analyzed in isolation.

The incorporation of intersectional fairness metrics into recommendation systems thus represents a frontier in both theory and practice. Several recent studies have proposed novel metrics, fairness-aware algorithms, and empirical methodologies to capture intersectional disparities. This review synthesizes the state of research on this emerging topic, identifying key contributions, debates, and areas requiring further investigation.

### Objectives
The primary objective of this paper is to comprehensively examine the role of intersectional

metrics in assessing fairness within large-scale recommendation systems. It seeks to analyze how traditional fairness approaches often overlook the compounded disadvantages faced by individuals belonging to multiple marginalized groups. The paper aims to explore the theoretical underpinnings of intersectionality, its relevance in machine learning fairness discourse, and its applicability to real-world recommendation scenarios. Additionally, the paper endeavors to review and synthesize existing algorithmic strategies that have been proposed to operationalize intersectional fairness, evaluating their strengths and limitations within the context of recommendation engines. Through critical evaluation of empirical studies and datasets used in this domain, the paper aspires to highlight both the potential and the ongoing challenges of incorporating intersectional perspectives into fairness assessments. Ultimately, the objective is to advance scholarly understanding of how intersectional metrics can lead to more equitable, inclusive, and ethically sound recommendation systems that serve diverse user populations more justly.

**The challenge of fairness in recommendation systems**

Recommendation algorithms typically optimize for user engagement, click-through rates, or purchase likelihoods. These objectives, however, often rely on historical data that may reflect existing societal biases. For example, historical viewing patterns on a video streaming platform may reflect gendered differences in content exposure or cultural preferences shaped by systemic inequalities. If these patterns are uncritically learned by the recommendation engine, they risk reinforcing the very disparities they aim to mitigate.

Moreover, recommendation systems are highly dynamic, with feedback loops that can exacerbate biases over time. An underrepresented creator or seller may receive limited exposure, resulting in fewer interactions and, consequently, lower future recommendations. This form of popularity bias can lead to algorithmic homophily, where dominant groups receive more visibility while minority groups are increasingly marginalized.

Traditional fairness metrics applied in supervised learning contexts often fail to address these complexities. For instance, demographic parity seeks to equalize outcomes across sensitive groups but struggles with personalization objectives inherent in recommendations. Equal opportunity and disparate impact metrics, while useful, generally focus on single attributes such as race or gender in isolation, thereby missing compounded disadvantages experienced at the intersection of multiple attributes.

**Intersectional Fairness: Theoretical Foundations**

Intersectionality emerged from feminist legal theory as a way to describe the multiple burdens faced by Black women in legal discrimination cases. In the context of machine learning and recommendation systems, intersectionality demands that fairness assessments consider not just isolated sensitive attributes but combinations thereof. A user may be simultaneously disadvantaged due to being, for example, both a woman and a member of a racial minority, a perspective often missed by aggregate group-based fairness assessments.

Intersectional fairness metrics thus evaluate algorithmic outcomes across multi-attribute groupings. For instance, one may measure recommendation relevance, exposure, or satisfaction across all combinations of gender, race, and disability status. This enables the identification of disparities that remain hidden when sensitive attributes are considered separately. The challenge, however, lies in the combinatorial explosion of possible groupings, which grows exponentially with the number of intersecting attributes. This leads to data sparsity, making statistical estimation difficult, especially for smaller intersectional subgroups.

Researchers have proposed various mathematical formalizations of intersectional fairness. One approach is *subgroup fairness*, which ensures that algorithmic performance meets fairness criteria for all statistically significant subgroups. Another is *multicalibration*, which enforces calibration conditions across multiple overlapping subgroups. Both methods aim to provide robust fairness guarantees even in the presence of complex intersectional structures.

**Table 1:** Comparison of fairness metrics in recommendation systems

| Fairness Metric | Definition | Strengths | Limitations | Applicability to Intersectionality |
|---|---|---|---|---|
| Demographic Parity | Equal proportion of positive outcomes across groups | Simple to compute, widely used | Ignores personalization and relevance | Low |
| Equal Opportunity | Equal true positive rates across groups | Considers accuracy trade-offs | Does not account for exposure | Low |
| Exposure Fairness | Equal visibility of items/users across groups | Targets visibility imbalance | May affect user satisfaction | Medium |
| Calibration | Predicted scores match observed outcomes for each group | Aligns predictions with true likelihood | Hard to achieve with sparse data | Medium |
| Subgroup Fairness | Ensures fairness across all statistically significant subgroups | Directly supports intersectionality | Computationally intensive with many groups | High |
| Multicalibration | Extends calibration across overlapping subgroups | Handles multiple intersections | Requires complex auditing | High |

**Algorithmic Approaches to Intersectional Fairness**

To operationalize intersectional fairness in recommendation systems, several algorithmic strategies have been developed. One line of work involves pre-processing methods that reweight or augment training data to balance representation across intersectional groups. For example, synthetic oversampling techniques may generate additional data for underrepresented subgroups to mitigate data sparsity.

Another approach focuses on in-processing methods that incorporate fairness constraints directly into the recommendation model’s objective function. Multi-objective optimization frameworks can jointly optimize for accuracy and fairness across intersectional groups, although defining the appropriate trade-offs remains an open question.

Post-processing methods also exist, where model outputs are adjusted to correct for observed disparities. For instance, exposure constraints can be applied to limit the gap in visibility between advantaged and disadvantaged intersectional groups. However, these adjustments may compromise personalization goals or introduce unintended side effects if not carefully calibrated.

Graph-based recommendation models, commonly used in modern platforms, offer unique challenges and opportunities for fairness interventions. Some studies have explored fairness-aware graph neural networks that propagate fairness constraints through the recommendation graph, helping to correct systemic imbalances at multiple levels of the recommendation pipeline.

**Empirical Evaluations and Datasets**

Empirical evaluations of intersectional fairness require datasets that contain rich, multi-attribute demographic information, which is often unavailable due to privacy concerns. Some studies have utilized synthetic datasets to simulate intersectional disparities, while others have used real-world datasets such as MovieLens or Amazon Reviews, often annotated with inferred demographic information.

Results across studies consistently show that intersectional fairness metrics reveal disparities not captured by single-attribute assessments. For example, exposure gaps for women of color may remain substantial even when overall gender or racial fairness appears acceptable. Moreover, fairness-aware algorithms that target intersectional metrics often outperform single-attribute fairness interventions in reducing compounded disadvantages.

However, challenges remain. Data sparsity for some intersectional groups makes it difficult to draw statistically reliable conclusions. Furthermore, the choice of fairness metric can significantly influence perceived disparities, underscoring the importance of carefully defining fairness objectives that align with the ethical priorities of the application domain.

**Open Challenges and Future Directions**

Despite substantial progress, assessing fairness through intersectional metrics in large-scale recommendation systems remains a complex and evolving research frontier. One major challenge is scalability. As the number of sensitive attributes increases, the computational cost of evaluating and enforcing fairness constraints grows exponentially. Efficient algorithms capable of handling high-dimensional intersectional groups are urgently needed.

Another critical issue is the trade-off between personalization and fairness. Excessive fairness constraints may degrade recommendation relevance for certain users, raising questions about user satisfaction and business objectives. Developing algorithms that achieve acceptable trade-offs without disproportionately harming either personalization or fairness remains an open problem.

Privacy concerns also complicate intersectional fairness assessments. Collecting detailed demographic information needed for intersectional evaluations risks exposing users to privacy breaches or surveillance, especially in sensitive contexts such as healthcare or finance. Privacy-preserving machine learning techniques, including federated learning and differential privacy, may offer partial solutions.

Finally, there is a need for richer theoretical frameworks that move beyond purely statistical definitions of fairness. Ethical, legal, and cultural considerations must inform the development of intersectional fairness metrics, ensuring that technical interventions reflect real-world justice concerns rather than merely statistical parity.

**Conclusion**

As recommendation systems increasingly shape human experiences in the digital world, ensuring their fairness is not merely a technical problem but a profound ethical imperative. Traditional fairness metrics, while valuable, fall short in capturing the complex realities of intersectional disadvantage. Intersectional fairness metrics provide a more nuanced lens to evaluate and correct systemic biases in large-scale recommendation systems, revealing disparities that might otherwise remain hidden.

Significant research progress has been made in developing intersectional fairness metrics, algorithms, and empirical evaluation methods. However, challenges related to data sparsity, computational scalability, privacy, and ethical alignment persist. Addressing these challenges will require continued interdisciplinary collaboration across computer science, social sciences, ethics, and policy domains. By embracing the complexities of intersectionality, researchers and practitioners can design recommendation systems that promote not only personalization and engagement but also equity and justice for all users.

**References**


1. Burke R. Multisided fairness for recommendation. Proceedings of the 2017 Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML), 2017, p. 1-5.
2. Ekstrand MD, Tian M, Azpiazu I, Ekstrand J, McNeill D, Anuyah O, *et al.* All the cool kids, how do they fit in? Popularity and demographic biases in recommender evaluation and effectiveness. Conference on Fairness, Accountability, and Transparency (FAT), 2018, p. 172-186.
3. Crenshaw K. 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;1989(1):139-167.
4. Yao S, Huang B. Beyond parity: Fairness objectives for collaborative filtering. Advances in Neural Information Processing Systems (NeurIPS). 2017;30:2921-2930.
5. Beutel A, Chen J, Zhao Z, Chi E. Data decisions and theoretical implications when adversarially learning fair representations. Proceedings of the 2017 Fairness, Accountability, and Transparency in Machine Learning (FAT/ML), 2017, p. 1-5.
6. Dastin J. Amazon scrapped secret AI recruiting tool that showed bias against women. Reuters., 2018. https://www.reuters.com.
7. Kearns M, Roth A. The ethical algorithm: The science of socially aware algorithm design. Oxford University Press, 2019.
8. Kearns M, Neel S, Roth A, Wu ZS. Preventing fairness gerrymandering: Auditing and learning for subgroup fairness. International Conference on Machine Learning (ICML), 2018, p. 2569-2577.

9. Kim P, Scott J. Towards fair recommendations: Incorporating intersectional fairness in recommender systems. Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society, 2020, p. 117-123.
10. Holstein K, Wortman Vaughan J, Daumé III H, Dudik M, Wallach H. Improving fairness in machine learning systems: What do industry practitioners need? Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, p. 1-16.
11. Singh A, Joachims T. Fairness of exposure in rankings. Proceedings of the 24t[h] ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, p. 2219-2228.
12. Patro G, Biswas A, Ganguly N, Gummadi KP, Chakraborty A. FairRec: Two-sided fairness for personalized recommendations in two-sided platforms. Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society, 2020, p. 469-481.
13. Steck H. Calibrated recommendations. Proceedings of the 12th ACM Conference on Recommender Systems (RecSys), 2018, p. 154-162.
14. Dwork C, Hardt M, Pitassi T, Reingold O, Zemel R. Fairness through awareness. Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, 2012, p. 214-226.
15. Zliobaite I. Measuring discrimination in algorithmic decision making. Data Mining and Knowledge Discovery. 2015;31(4):1060-1089.