Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 31 No. 5s (2024)

Promoting Fairness in Recommender Systems: A Multifaceted Approach

Harshali Bhuwad[1] , Dr.Jagdish.W.Bakal[2]

[1]Research Scholar, Thadomal Shahani Engineering College, Mumbai, 400050, Maharashtra, India

[2]Principal, Pillai HOC College of Engineering Technology, Navi-Mumbai, 410207, Maharashtra, India

Article History:

Received: *16-05-2024*

Revised: *24-06-2024*

Accepted: *09-07-2024*

Abstract:

The increasing prevalence of recommender systems necessitates a critical examination of fairness concerns. Biases within data or algorithms can lead to discriminatory recommendations, hindering user experience and potentially causing societal harm. This review paper delves into the multifaceted landscape of promoting fairness in recommender systems. We explore various debiasing techniques, including data preprocessing (e.g., cleaning and filtering) to address biased data, fairness-aware data augmentation to counter representation imbalances and algorithmic debiasing approaches that promote equitable treatment across user groups. We further examine fairness aware metrics that go beyond traditional measures like click-through rates, such as parity metrics (e.g., statistical parity, equality of opportunity) and diversity metrics (e.g., coverage, novelty) to ensure fair distribution and recommendation variety for different users. Finally, we emphasize the significance of transparency and user control. By providing users with insights into recommendation rationale and empowering them to manage their data and preferences, we can build trust and foster a more equitable recommendation ecosystem. This review paper sheds light on the current state-of-the-art approaches to fair recommender systems, paving the way for future research and development in this crucial area.

Keywords: Recommender Systems, Fairness, Debiasing Techniques, Data Cleaning, Data Augmentation, Algorithmic Debiasing, Fairness-aware Metrics, Parity Metrics, Diversity Metrics, Transparency, User Control.

1. Introduction

Recommender systems have become an essential part of our daily lives, assisting us in navigating the ever-growing sea of information online. These intelligent algorithms personalize user experiences by suggesting relevant items, ranging from products and movies to music and news, based on individual preferences and past interactions [1]. By filtering the vast amount of data available, recommender systems empower users to make informed choices aligned with their interests [2]. However, while undeniably beneficial, there is a growing concern regarding fairness and bias within these powerful algorithms.

Traditionally, research focused primarily on the value recommender systems deliver to consumers. However, recent years have shed light on the potential for these systems to perpetuate biases and have unintended consequences [3]. These biases can stem from various factors, such as demographics, user behaviour, and even the inherent popularity of certain items. One well-known issue is bias towards popular items. Popular items tend to receive more exposure and engagement from users, leading recommender systems to prioritize them over potentially relevant but less popular recommendations. While this is a significant concern, it is just one facet of the broader and more complex issue of fairness and bias in recommender systems.

One under-explored challenge in achieving fairness is the dynamic nature of user preferences. As Yoo et al. (2023) point out, recommender systems that adapt to new user data can experience degradation in fairness over time, despite improvements in overall recommendation performance [4]. This highlights the need for novel approaches that can maintain fairness while adapting to evolving user preferences. This review paper delves into this critical issue, exploring how recommender systems can unintentionally perpetuate biases. We will analyse the various sources of bias within these systems and their impact on users and society. Additionally, we will examine promising approaches and ongoing research efforts focused on promoting fair and unbiased recommendations. By comprehensively reviewing existing research, this paper aims to illuminate the complexities of fairness in recommender systems. We hope to highlight the importance of addressing bias in this rapidly evolving field and pave the way for developing more equitable and responsible recommendation algorithms.

2. Background on Fairness and Bias in Machine Learning

Fairness in machine learning (ML) algorithms refers to the absence of prejudice or discriminatory outcomes based on irrelevant characteristics [5]. Ideally, ML models should make predictions or recommendations solely based on the relevant features used for training, without favouring or disfavouring specific groups or individuals.

However, bias can creep into ML models in various ways, leading to unfair outcomes. Bias, in this context, refers to a systematic preference or prejudice within the algorithm that can skew its outputs [6]. This bias can originate from different sources, affecting fairness in various ways. Here is a breakdown of some key types of fairness concerns in ML algorithms:

2.1 Types of Fairness Concerns in Machine Learning

Table 1 summarizes the different types of fairness concerns, their descriptions, focus areas, and examples relevant to recommender systems

Table 1: Types of Fairness Concerns in Machine Learning

Fairness Type	Description	Focus	Example (Recommender System)
Individual Fairness	Similar treatment for similar individuals	Model outputs	Users with similar preferences receive similar recommendations (regardless of race, gender, age).
Group Fairness	Avoiding discrimination against specific groups	Model outcomes for different groups	Different demographic groups (e.g., young vs. older users) have equal access to diverse recommendations.
Algorithmic Fairness	Mitigating bias in design and training	Design choices and training data	Examining training data for biases and scrutinizing algorithm design to avoid favoring certain items.

2.2 Methodologies to Address Fairness Concerns

Beyond understanding the different types of fairness concerns, it's crucial to explore the methodologies used to address them. These methodologies can be applied at various stages of the ML pipeline, from data collection and training to model evaluation and deployment. Specific approaches will vary depending on the type of fairness concern being addressed. Here is a breakdown of some common methodologies used to address fairness in recommender systems:

1. **Individual Fairness:** This principle focuses on ensuring similar treatment for individuals with similar characteristics. In recommender systems, for example, individual fairness would mean that

users with similar past interactions or preferences receive similar recommendations, regardless of irrelevant factors like race, gender, or age [7].

• **Methodologies:**

– **Counterfactual Fairness:** This approach compares the predicted outcome for an individual with a hypothetical scenario where their sensitive attribute (e.g., race) is changed. The model is considered fair if the predictions remain similar across these scenarios [5].

– **Fairness-aware Metrics:** Beyond accuracy, metrics like fairness aware ranking (fair rank) can be used to evaluate how similar individuals with similar features are treated by the model[7].

2. Group Fairness: This concept emphasizes avoiding discrimination against specific demographic groups. When applied to recommender systems, group fairness would ensure that different demographic groups (e.g., young vs. older users) have equal access to a diverse range of recommendations and are not limited by the algorithm's biases [8].

• **Methodologies:**

– **Disparate Impact Analysis:** This method evaluates whether the model's predictions have a disproportionately negative impact on certain demographic groups. For recommender systems, this could involve analysing whether specific groups consistently receive lower-rated recommendations [8].

– **Equality of Opportunity/Outcomes:** These principles aim to ensure that different groups have similar opportunities to achieve positive outcomes based on the model's recommendations [9].

3. Algorithmic Fairness: This type of fairness focuses on the design and training process of the ML model itself [10]. Algorithmic fairness aims to identify and mitigate biases within the algorithm's design choices or the data used to train it. For recommender systems, this might involve examining potential biases in the training data used (e.g., historical user data reflecting existing societal biases) or scrutinizing the algorithm's design to ensure it does not inherently favour certain types of items [11].

• **Methodologies:**

– **Debiasing Techniques:** These techniques aim to remove or mitigate biases within the training data itself. This could involve data augmentation (adding data to represent underrepresented groups) or data balancing (adjusting the weights of data points to ensure equal representation) [10].

– **Fairness-aware Model Architectures:** To reduce bias, algorithms should be designed with fairness requirements integrated into the learning process. Research in this field is ongoing, and several strategies are being investigated [10].

When assessing and reducing bias in recommender systems, it is essential to comprehend these various fairness concerns and the approaches taken to address them. We may work towards more responsible and equitable AI applications by guaranteeing fairness at the individual, group, and algorithmic levels [12].

3. Sources of Bias in Recommender Systems

Recommender systems undoubtedly help to personalise user experiences, but they can also reinforce prejudices in several ways. To create algorithms that are fairer and more egalitarian, it is essential to comprehend these causes of bias. We examine the main offenders that can compromise recommender systems here:

1. Data Bias: The term "data bias" describes the existence of bias or uneven representation in the training set of data that was used to create the recommender system. A feedback loop that intensifies already existing societal inequities can be caused by several variables, including bias:

• **Historical Biases:** Historical user data frequently reflects biases from the real world. In an investigation of gender bias in movie recommendations, for example, [13] discovered that recommender systems trained on historical data tended to favour films directed by men, even for users who did not express a preference for such films. This restricts users' access to a variety of content and perpetuates preexisting gender prejudices.

• **Limited Data:** Bias amplification is more common in recommender systems with sparse data. It is important to note that recommendations may not generalise adequately to users from varied backgrounds when training data predominantly represents a certain demographic group [7]. When users from underrepresented groups are excluded from relevant recommendations, it can result in an issue called exclusion bias.

• **Data Collection Methods:** Bias may be introduced by the methods used to acquire the data. Users who are less at ease expressing their preferences may be underrepresented in a recommender system, for example, if it only uses explicit user evaluations (such as thumbs up/down). In their investigation of algorithmic bias in social news recommenders, [14] pointed out that this can result in biassed recommendations that don't accurately represent the whole spectrum of user interests.

Consider the following pie chart in Figure 1 depicting a scenario with significant gender bias in user data for a movie recommender system. As seen, the data primarily represents the preferences of a single gender group (e.g., 80% Male). This may result in suggestions that are biassed in favour of directors or genres that are generally enjoyed by that group, disregarding the needs of the underrepresented group (e.g., 20% Female).

Figure 2 showcases an age bias within user data for a music recommender system. Most data points belong to younger users (e.g., 70% under 35), potentially leading to a bias towards music genres and artists popular with younger audiences.

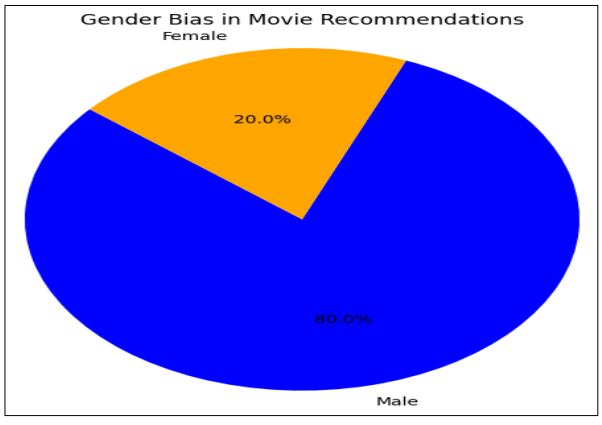

Figure 1: Gender Bias in Movie Recommender System Data

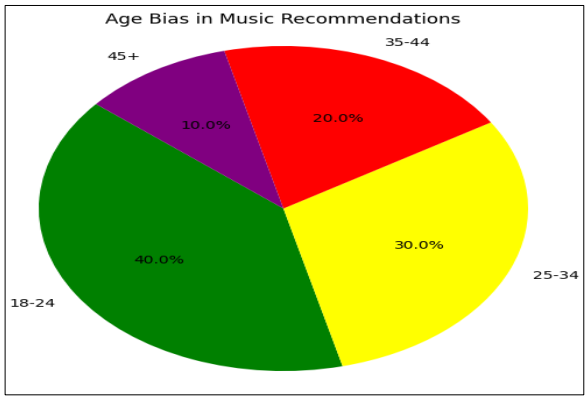

Figure 2: Age Bias in Music Recommender System Data

2. Algorithmic Bias: Algorithmic bias refers to inherent biases within the design of the recommender system itself. These biases can arise from the very choices made during the development process:

• **Algorithmic Choices:** One potential source of bias is the recommendation algorithm's choice. Algorithms that favour items with more user interactions, such as collaborative filtering (CF) algorithms, have been shown to magnify popularity bias. In their work on fairness-aware matrix factorization for recommender systems, [3] draw attention to the possibility of a long-tail problem, in which less popular but potentially important options are ignored.

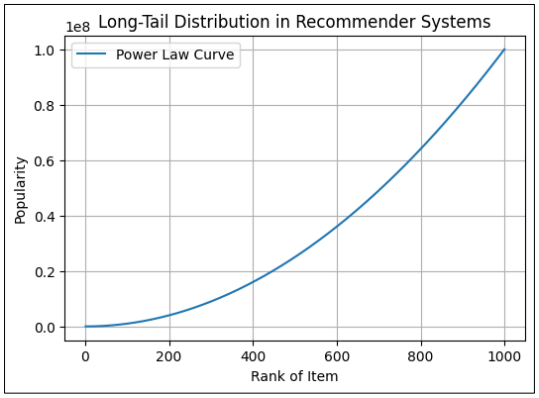

Figure 3: Long-Tail Distribution in Recommender Systems

The long-tail issue in recommender systems is highlighted in Figure 3, which uses a power law curve to illustrate the idea of algorithmic bias. The power law curve displays a steep drop in popularity along the sorted list's (X-axis) downward direction. This indicates that the recommender system tends to prioritise popular things (left side of the curve) with higher user interactions, meaning that many niche products (long tail) receive less attention.

• **Similarity Metrics:** There is potential for bias in the metrics employed to assess item similarity. A measure that favours item attributes that are regularly purchased together, for example, may penalise niche items that are not frequently purchased in conjunction with other popular items. In their investigation of fairness in recommender systems, [15] address how this may prevent recommendations from being serendipitous and restrict user exposure to a range of possibilities.

• **Exposure Bias:** Bias may be introduced into recommendations by the recommender system's presentation. The echo chamber effect, which limits a user's exposure to a variety of possibilities and reinforces their preexisting preferences, may be caused by algorithms that give priority to highly-rated goods. In his research on the drawbacks of personalisation algorithms, [16] examined these phenomena, also referred to as filter bubble bias.

3. Interaction Bias: Interaction bias refers to biases introduced through user interactions with the recommender system. These biases create a feedback loop where user behaviour unwittingly reinforces existing biases within the system:

• **Confirmation Bias:** Users tend to click on and interact with recommendations that align with their existing preferences. This reinforces the bias within the system as the algorithm prioritizes items like those the user has already interacted with. This phenomenon, explored by [17], can lead to a cycle where users are continuously exposed to content that confirms their existing beliefs and limits their exploration of new and diverse viewpoints.

• **Clickstream Bias:** To improve their suggestions, recommender systems frequently rely on user clicks and interactions. But if consumers are more likely to click on recommendations that catch their attention than on ones that truly suit them, this could result in bias. This clickstream bias might cause

the algorithm to choose sensationalised information over more insightful recommendations, as [1] pointed out in their research on recommender system biases.

• **Feedback Loop:** The system's prejudice is further reinforced by positive user feedback on biassed suggestions. This leads to a feedback cycle, as [18] highlight, wherein initially biassed recommendations gradually grow even more pronounced. This may cause consumers to be exposed to biassed content more frequently, which would reinforce the system's inherent prejudices.

By recognising these diverse origins of prejudice, scholars and programmers can endeavour to establish more just and equitable recommender systems with great thought.

4. Impact of Bias in Recommender Systems

Although recommender systems provide unquestionably valuable personalisation, users and society may suffer greatly from biassed recommendations. Here, we examine these possible effects in more detail, using knowledge from pertinent studies:

1. Reduced User Satisfaction and Trust in the System:

• **Irrelevance and Frustration:** Biased recommendations frequently fall short of expectations and offer unrelated products or information [14]. Reduced pleasure and a feeling of being neglected can result from this frustration with the system's incapacity to comprehend user choices [1].

• **Lack of Discovery and Serendipity:** Excessive exposure to a variety of potentially fascinating stuff can result from biased algorithms that favour popular things or reinforce preexisting preferences [17]. It can also make user experiences monotonous and predictable. Serendipity is the process by which wonderful things happen by accident.

• **Loss of Trust and Disengagement:** User confidence in the system may be damaged by frequently receiving discriminating or irrelevant recommendations [15]. Consumers may completely stop using the platform or go to another one they believe to be more objective[19].

2. Perpetuation of Social Inequalities and Discrimination:

• **Amplification of Existing Biases:** Social prejudices in the training data sets of recommender systems have the potential to be amplified. For example, racial or gender biases may result in recommendations that limit opportunities for specific user groups or perpetuate preconceptions [13]. This may make already-existing socioeconomic injustices worse.

• **Exclusion and Filter Bubbles:** Users may be excluded from pertinent recommendations due to biassed algorithms, or they may be kept in "filter bubbles" where they are only exposed to information that confirms their preexisting opinions [16]. This may impede social mobility and restrict exposure to a range of viewpoints, which could result in social polarisation.

• **Unfair Access to Resources and Opportunities:** Biased algorithms can maintain prejudice and restrict access to resources for specific groups in crucial domains such as loan approvals or employment recommendations [7]. This can impede equal opportunity and have a major detrimental effect on people.

3. Limited Exposure to Diverse Content and Viewpoints:

• **Echo Chambers and Confirmation Bias:** Users who are largely exposed to content that supports their preexisting ideas may find themselves in "echo chambers" caused by biased recommender systems [17]. This may restrict their exposure to contrasting ideas and impede their ability to critically analyse and make well informed decisions.

• **Reduced Knowledge and Understanding:** A user's capacity to learn about new subjects and comprehend various points of view may be hampered by limited exposure to diverse content [18]. This may impede intellectual development and result in a lack of well-rounded knowledge.
• **Reduced Openness and Tolerance:** Reduced tolerance for opposing ideas and openness to new ones can result from a continual barrage of information that supports preexisting beliefs [20]. This could stifle fruitful discussion and deepen social division.

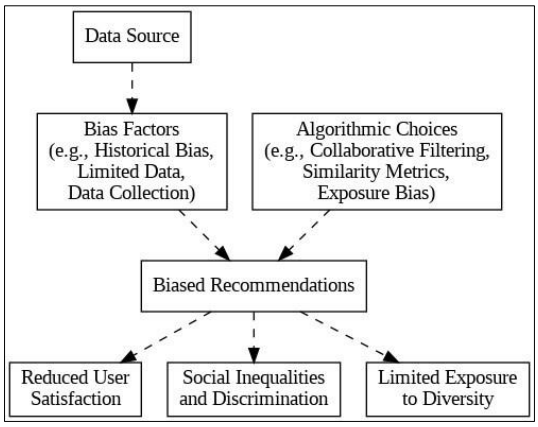

Figure 4: Impact of Bias in Recommender Systems

The figure 4 illustrates the key factors that contribute to bias in recommender systems (data source, algorithmic choices) and their potential consequences on user experience (reduced satisfaction) and social aspects (inequalities, discrimination, and limited exposure to diversity).
These potential consequences highlight the importance of mitigating bias in recommender systems. By promoting fairness and inclusivity, we can ensure these powerful tools enhance user experience, contribute to a more equitable and informed society, and avoid the pitfalls outlined by researchers like [14],[1], [17], and [15].

5. Approaches to Fair Recommendation Systems

The potential pitfalls of bias in recommender systems necessitate the development of fair and inclusive algorithms. Researchers have explored various approaches to mitigate bias and promote fairness in recommendations. Here, we delve into some key strategies, drawing insights from relevant research:

1. **Debiasing Techniques:** This approach focuses on identifying and removing biases from the data or within the algorithm itself. Here are some key techniques:
• **Data Preprocessing:** Biased entries can be found and eliminated from the training data using methods like data cleaning and filtering. If a user's gender or race is not directly related to the recommendation task, this could entail removing user attributes that are known to be associated with bias [7]. To prevent overcorrection and unexpected outcomes, however, prudence is required [21].
• **Fairness-aware Data Augmentation:** Fairness can be increased by creating data points artificially to offset imbalances or underrepresentation in the training set. This can be accomplished by using methods such as creating artificial users with demographic profiles that are balanced [1].
• **Algorithmic Debiasing:** Modifying the recommendation algorithms themselves to be less susceptible to bias. This could involve techniques like counterfactual fairness, where the algorithm considers how a recommendation for a similar user from a different demographic group might differ [22]. For

instance, in a job recommendation system, a debiased algorithm might prioritize skills and experience over gender when suggesting suitable candidates.

• **Debiasing with Counterfactual:** A recent approach [23] introduces counterfactual explanations for fairness in recommender systems. This method focuses on identifying attribute-level explanations for potential bias in recommendations. It utilizes Heterogeneous Information Networks (HINs) to represent user attributes and their relationships, enabling the consideration of discrete attributes like gender or location. The paper leverages off-policy reinforcement learning to identify the most impactful attribute changes for improving fairness. By explaining how a recommendation might differ if a user belonged to a different demographic group, this approach enhances transparency and facilitates human oversight.

2. Fairness-aware Metrics: Traditional recommender system evaluation relies heavily on accuracy metrics like click-through rate (CTR).

However, these metrics can overlook issues of fairness. Fairness-aware metrics go beyond accuracy to consider how recommendations are distributed across different user groups. As examples, consider:

• **Parity Metrics:** These metrics ensure that similar recommendations are offered to users from different demographic groups. Examples include statistical parity (equal recommendation rates) and equality of opportunity (equal chances of receiving a top recommendation) [13].

• **Diversity Metrics:** These metrics measure the variety of recommendations offered to users, preventing the system from getting stuck in "filter bubbles" and promoting exposure to diverse content. Examples include coverage (the range of items recommended) and novelty (the proportion of recommendations unseen by the user before)[15].

It's crucial to remember that accuracy and fairness frequently have to be traded off. Scholars are currently devising techniques that strike a compromise between these conflicting objectives [24] and employ a range of indicators to obtain a comprehensive understanding of the system's functioning.

3. Transparency and User Control:

• **Transparency:** Providing users with insights into how recommender systems work and the factors influencing their recommendations can help build trust. This could involve explanations for why specific items are recommended and the ability to understand the system's reasoning [25].

• **User Control:** Empowering users to control the data used for personalization offers them more agency. This could include mechanisms to allow users to adjust their recommendation profiles, remove irrelevant data points, or opt-out of certain features [26].

These methods present encouraging paths for reducing bias and advancing equity in recommender systems. We can guarantee that these recommender systems provide a satisfying user experience that is both individualised and equitable by utilising a blend of debiasing strategies, fairness-aware metrics, and transparency with user control.

6. Challenges and Future Directions in Fair Recommendation Systems

While significant progress has been made in developing fair recommender systems, there are still substantial challenges to overcome. Here, we delve into some key hurdles and promising future research directions, drawing insights from recent research:

• **Defining and Measuring Fairness:** Fairness is a multifaceted concept, and there is no single universally accepted definition. Different stakeholders (users, developers, policymakers) may prioritize different aspects of fairness. Metrics like parity and diversity offer valuable insights, but they might not capture

the full picture. Future research should explore comprehensive frameworks to define and measure fairness that consider various stakeholders' perspectives and the potential for unintended consequences [27]. For instance, a fairness metric might account for not only equal opportunity for all groups but also the quality of recommendations received [28].

• **Balancing Fairness with Other Goals:** In recommender systems, achieving justice frequently means sacrificing other desirable characteristics like accuracy and personalisation. Debiasing algorithms, for example, may initially result in suggestions that are less accurate. Subsequent investigations ought to delve into strategies that reconcile these conflicting objectives, possibly using approaches such as context-aware fairness considerations or multi-objective optimisation [29]. Depending on the recommendation task and user circumstance, this may include dynamically modifying the fairness constraints.

• **Mitigating Bias while Ensuring User Privacy:** Concerns about privacy arise because many debiasing methods rely on user data. If users worry about discrimination or improper use of their data, they may be reluctant to divulge sensitive information. Future studies could look on fairness-aware, privacy-preserving strategies that provide equitable results without jeopardising user privacy [30]. This could entail using differential privacy strategies that introduce noise to safeguard individual data points while maintaining statistical features for fairness analysis, federated learning techniques, or anonymised data.

• **Fairness across Different Recommendation Tasks and Contexts:** Even though fairness has been the subject of research in a variety of circumstances (such as loan approvals and employment recommendations), the issues and difficulties may vary. Future studies should investigate how fairness might be addressed in various domains, considering elements such as the potential for social impact and the sensitivity of the recommendation task (e.g., fairness in criminal justice recommendations is crucial) [6].

• **Evolving Nature of Bias and Societal Norms:** Societal norms and understandings of fairness are constantly evolving. What is considered fair today might not be considered fair tomorrow. Future research should develop methods that can adapt to these changes, potentially through incorporating human-in-the-loop fairness feedback mechanisms or ongoing evaluations based on evolving societal standards [21].

7. Future Research Directions:

Building on these challenges, several promising avenues exist for future research in fair recommender systems:

• **Explainable and Transparent Fairness Mechanisms:** Enhancing user confidence and facilitating human supervision can be achieved by creating interpretable algorithms that can justify suggestions and exhibit fairness considerations [31]. To demonstrate to users how suggestions might vary for users with varied traits, counterfactual explanations are one strategy that can be used in this regard.

• **Counterfactual Fairness and Causal Inference Techniques:** More robust algorithms that consider fairness can be achieved by utilising counterfactual reasoning and causal inference to comprehend how

suggestions might vary for users belonging to different demographics [22]. By doing this, algorithmic bias that may not be immediately visible from traditional measures can be found and addressed.

• **User-Centric Fairness Controls:** Enhancing user agency and recommender system fairness can be achieved by giving users more precise control over how their data is utilised for personalisation and by giving them the opportunity to voice concerns about fairness [26]. Users may be able to modify their recommendation profiles, report recommendations that are biased, or choose not to use specific personalisation elements.

• **Interdisciplinary Collaboration:** To create strong fairness frameworks and handle the intricate societal ramifications of bias in recommender systems, cooperation amongst academics in computer science, social sciences, law, and ethics is essential [32]. This can guarantee that technological solutions are based on moral principles and actual social circumstances.

By tackling these challenges and pursuing these exciting research directions, we can move closer to achieving fair and unbiased recommender systems that benefit all users and contribute to a more equitable and just society.

8. Conclusion:

Recommender systems are powerful tools that shape our online experiences, but they can perpetuate bias if left unchecked. This review paper has explored the various sources of bias that can infiltrate these systems, including data imbalances, algorithmic limitations, and even societal prejudices reflected in user data. These biases can have a significant impact on users, leading to unfair outcomes like unequal access to opportunities or exposure to limited viewpoints.

Fortunately, researchers are actively developing methods to mitigate bias and promote fairness in recommender systems. Debiasing techniques aim to remove biases from data or algorithms, while fairness-aware metrics help us evaluate systems beyond just accuracy and ensure fair distribution of recommendations across user groups. Transparency and user control are also crucial elements, allowing users to understand how recommendations are generated and giving them agency over their data and personalization experience.

The pursuit of fair recommender systems is an ongoing journey. Research is needed to develop comprehensive frameworks that define and measure fairness, considering the perspectives of various stakeholders and the potential for unintended consequences. Furthermore, balancing fairness with other goals like accuracy and personalization remains a challenge, requiring innovative approaches that can optimize for both. Privacy-preserving techniques for debiasing and interdisciplinary collaboration between computer scientists, social scientists, and ethicists are also crucial areas for future exploration.

These advancements hold significant implications for recommender system designers and users. Designers must prioritize fairness considerations throughout the development process, employing robust debiasing techniques and fairness-aware evaluation metrics. Transparency tools and user controls empower users to understand these systems and advocate for their fair treatment. Ultimately, by working together, researchers, developers, and users can create fairer and more equitable recommender systems that benefit all and contribute to a more just and inclusive online environment.

Declarations

Funding information

No funding was involved in the present work.

Conflicts of interest

Authors declare that there is no conflict of interest.

Code availability

The codes will be made available upon reasonable request to the authors.

Authors contributions

The work presented in the paper was primarily done by first author, including conceptualization, literature reading, data gathering, coding, result analysis, and manuscript writing. A review was conducted by second author, and visualization work was carried out by both authors.

Refrences

[1] e. a. Zhang, Yifan, Modeling clickstream for effective and scalable coldstart recommendation, Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining, 2014.
[2] H. Abdollahpouri, Popularity bias in recommendation: A multistakeholder perspective, 2020. arXiv:2008.08551.
[3] N. Neophytou, B. Mitra, C. Stinson, Revisiting popularity, and demographic biases in recommender evaluation and effectiveness, in: M. Hagen, S. Verberne, C. Macdonald, C. Seifert, K. Balog, K. Norvag, V. Setty (Eds.), Advances in Information Retrieval, Springer International Publishing, Cham, 2022, pp. 641–654.
[4] H. Yoo, Z. Zeng, J. Kang, R. Qiu, D. Zhou, Z. Liu, F. Wang, C. Xu, E. Chan, H. Tong, Ensuring user-side fairness in dynamic recommender systems, in: T. Chua, C. Ngo, R. Kumar, H. W. Lauw, R. K. Lee (Eds.), Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, May 13-17, 2024, ACM, 2024, pp. 3667–3678. doi:10.1145/3589334.3645536. URL https://doi.org/10.1145/3589334.3645536
[5] M. H. T. P. O. R. Dwork, Cynthia, S. Oh, Fairness through awareness, 2012.
[6] e. a. Selbst, Adrianna, Fairness and abstraction in sociotechnical systems, FAT* ’19: Proceedings of the Conference on Fairness, Accountability, and Transparency, 2019, pp. 59–68.
[7] S. R. Pedreschi, Fosca, M. Giannotti, integrating inductive data mining with complementary knowledge: a review of support for constraint-based mining, in: Knowledge and information systems, 2012, pp. 197– 219.
[8] M. H. Barocas, Solon, A. Narayanan, Fairness and machine learning, 2019.
[9] P. E. S. E. T. M. Hardt, M., Machine learning and algorithmic fairness, arXiv preprint arXiv:1606.08653, 2016.
[10] S. S. A. B. Mehra, Siddhartha, A. Rajagopal., Evaluating bias in machine learning algorithms: A survey, 2019.
[11] C. O. J. S. P. C. J. S. Amodei, Dario, I. Sutskever, Concrete problems in ai safety, arXiv preprint arXiv:1606.06565, 2016.
[12] S. S. Mitchell, Melanie, E. Brynjolfsson, Algorithmic fairness in practice, Nature 568.7753, 2019, pp. 77–81.
[13] e. a. Eynul, Talha, Fairness in recommendation systems: A survey, ACM Computing Surveys (CSUR) 54.1, 2020, pp. 1–36.
[14] e. a. Ekstrand, Michael, Algorithmic bias in social news recommender systems, Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency, 2019.
[15] e. a. Liu, Liyuan, Fairness in recommender systems: A fairness-aware ranking approach, Knowledge and Information Systems 82.2, 2020, pp. 373–393.
[16] E. Pariser, The filter bubble: What the internet hides from you and why it matters, Penguin Press, 2011.
[17] e. a. Boich´e, Adrien P., The role of confirmation bias in recommendation systems, in: Proceedings of the 2015 ACM Conference on Research in Information Systems, 2015. 20
[18] e. a. Ge, Mengqi, Revisiting fairness in recommendation systems: A personalized approach, in: Knowledge and Information Systems, 2019, pp. 503–524.
[19] Y. Tintarev, J. M. Masthoff, Long tail recommendation with user preferences-related confidence, ACM Transactions on Information Systems (TOIS) 35.2, 2017, pp. 1–20.

[20] C. R. Sunstein, Republic.com 2.0, Princeton University Press, 2009.
[21] e. a. Friedler, Sorelle, Fairness in computer systems, Communications of the ACM 62.1, 2019, pp. 51–61.
[22] e. a. Kusner, Matt, Algorithmic fairness under counterfactual interventions, arXiv preprint arXiv:1703.06061, 2017.
[23] X. Wang, Q. Li, D. Yu, Q. Li, G. Xu, Counterfactual explanation for fairness in recommendation, ACM Trans. Inf. Syst. 42 (4) (mar 2024). doi:10.1145/3643670. URL https://doi.org/10.1145/3643670
[24] e. a. Ye, Niantian, Fairness-aware recommender systems: Techniques, advances, and challenges, IEEE transactions on knowledge and data engineering 30.9, 2018, pp. 1663–1686.
[25] e. a. Mittelstadt, Brent Daniel, Explaining algorithms in healthcare: A case study of deep learning for cancer identification, Nature Machine Intelligence 1.11, 2019, pp. 131–138.
[26] e. a. Shin, Juno, User control for algorithmic fairness, Proceedings of the 2021 ACM Conference on Human Factors in Computing Systems, 2021, pp. 1–13.
[27] e. a. Borges, Inˆes, A survey on fairness in recommender systems, ACM Computing Surveys (CSUR), 2021.
[28] e. a. Verma, Aditya, Fairness considerations in personalized ranking, Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2020, pp. 138–148.
[29] e. a. Singh, Amanpreet, A survey on fairness in multi-objective recommender systems, arXiv preprint arXiv:2204.03282, 2022. 21
[30] e. a. Wang, Yuefeng, Privacy-preserving fair recommendation with federated learning, Proceedings of the 2023 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2023.
[31] e. a. Xu, Yingying, Explainable recommendation systems: A survey, ACM Computing Surveys (CSUR), 2021.
[32] e. a. Jobin, Anna, The ethics of artificial intelligence, Nature 569.7753, 2019, pp. 334–341.