Information Fusion 100 (2023) 101906

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

A survey on fairness-aware recommender systems☆

Di Jin [a], Luzhi Wang [a], He Zhang [b], Yizhen Zheng [b], Weiping Ding [c,*], Feng Xia [d], Shirui Pan [e,*]

[a] College of Intelligence and Computing, Tianjin University, China
[b] Department of Data Science and AI, Faculty of IT, Monash University, Australia
[c] School of Information Science and Technology, Nantong University, China
[d] School of Computing Technologies, RMIT University, Australia
[e] School of Information and Communication Technology, Griffith University, Australia

ARTICLE INFO

Keywords:
Recommender systems
Fairness
Trustworthiness
Survey

ABSTRACT

As information filtering services, recommender systems have extremely enriched our daily life by providing personalized suggestions and facilitating people in decision-making, which makes them vital and indispensable to human society in the information era. However, as people become more dependent on them, recent studies show that recommender systems potentially own unintentional impacts on society and individuals because of their unfairness (e.g., gender discrimination in job recommendations). To develop trustworthy services, it is crucial to devise fairness-aware recommender systems that can mitigate these bias issues. In this survey, we summarize existing methodologies and practices of fairness in recommender systems. Firstly, we present concepts of fairness in different recommendation scenarios, comprehensively categorize current advances, and introduce typical methods to promote fairness in different stages of recommender systems. Next, after introducing datasets and evaluation metrics applied to assess the fairness of recommender systems, we will delve into the significant influence that fairness-aware recommender systems exert on real-world industrial applications. Subsequently, we highlight the connection between fairness and other principles of trustworthy recommender systems, aiming to consider trustworthiness principles holistically while advocating for fairness. Finally, we summarize this review, spotlighting promising opportunities in comprehending concepts, frameworks, the balance between accuracy and fairness, and the ties with trustworthiness, with the ultimate goal of fostering the development of fairness-aware recommender systems.

1. Introduction

Recommendation systems (RSs) are information filtering systems that are expected to suggest products and services, i.e., items, that most likely interest a user [1]. The suggestions are related to various decision-making processes for a user, such as which products to purchase, which videos to watch and which songs to listen. As the world becomes more information-overloaded, recommender systems are particularly useful when users need to choose an item from an overwhelming selection offered by a service. Recommender systems are pervasive and have been utilized in various fields including e-commerce [2–4], economics [5–7], education [8–10], etc. For the e-commerce industry, Amazon, for example, distributes product information provided by merchants to users based on their history of purchases or website interaction. Remarkably, its recommendation engine accounts for three-quarters of its total revenue [11]. In the economics industry, the loans domain can be recommended to users based on their attributes information such as annual income, address, occupation, and so on [12]. Lastly, in the education industry, Massive Open Online Course (MOOC), one of the leading online learning providers, recommends courses to users based on their historical opinions [13].

However, as the development of recommender systems surges and the reliance on them grows, these systems can lead to huge negative impacts on society and individuals due to unfairness. The causes of unfairness and their undesirable outcomes are numerous and significant. For instance, a loan domain recommender system may recommend loan domains influenced by a user's sensitive personal attributes, such as age, gender, and race. As a result, minority groups such as senior citizens, women workers, and ethnic minorities may suffer worsening financial situations by being recommended loan domains with higher interest rates, which is obviously unfair. The second example is MOOC course recommendations [13], in which more than 40% of the courses are taught by American teachers, while the remaining courses from 73

☆ This project is funded by the National Natural Science Foundation of China (No: 62272340 and 61976120) and ARC Future Fellowship (No. FT210100097).
* Corresponding authors.
E-mail addresses: jindi@tju.edu.cn (D. Jin), wangluzhi@tju.edu.cn (L. Wang), he.zhang1@monash.edu (H. Zhang), yizhen.zheng1@monash.edu (Y. Zheng), dwp9988@163.com (W. Ding), f.xia@ieee.org (F. Xia), s.pan@griffith.edu.au (S. Pan).

https://doi.org/10.1016/j.inffus.2023.101906
Received 29 May 2023; Received in revised form 27 June 2023; Accepted 28 June 2023
Available online 4 July 2023

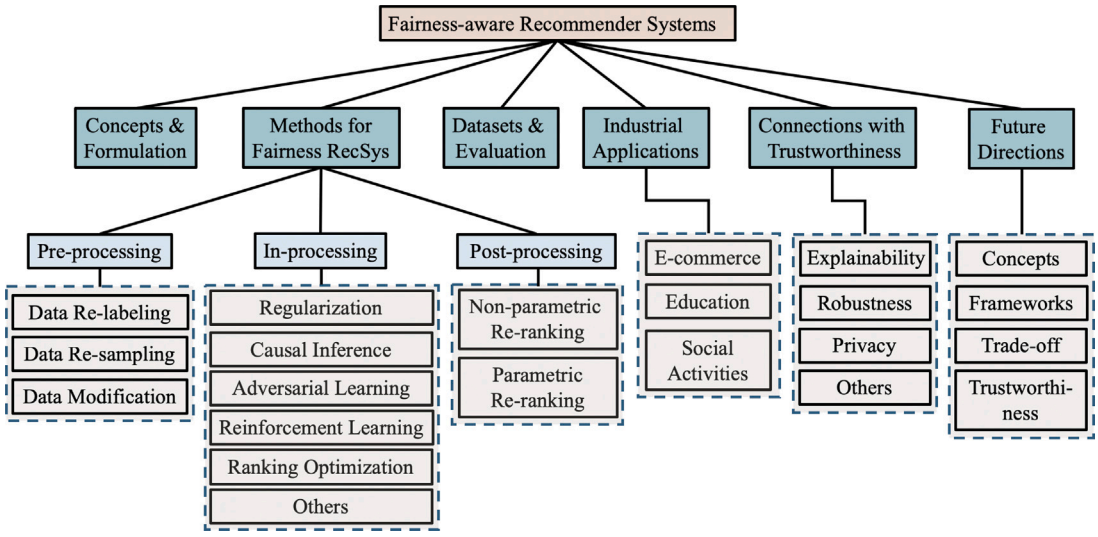


Fig. 1. The organizational layout of this survey.

countries are rarely recommended to or enrolled in by online users. Due to the vast geographic imbalance in MOOC recommendations, teachers in smaller or less-known places faced a great disadvantage when attracting students. In a similar way, Amazon tended to recommend items from larger merchants over smaller merchants. Because of this, smaller merchants would find it difficult to compete with large merchants, even if they offered better prices or quality. Although recommender systems are supposed to mitigate these unfairnesses, they can actually exacerbate them in the recommendation pipeline.

Driven by these unintentional issues, people increasingly yearn for fairness since it is critical and essential in developing recommender systems that can be trusted. First, fairness benefits users, items, and even recommender systems themselves [14]. For example, in a fair recommender system, users can obtain more relevant information, including niche information, which can aid in breaking out of the cocoon of information. As minor items are allocated more exposure, the Matthew effect [15] is lessened, encouraging providers to enhance their creativity and diversity. By providing equal quality of service for objects from diverse backgrounds, fair recommender systems can also gain long-term interest due to positive feedback from users and item providers. Second, a global consensus has recently been built on enhancing the trustworthiness of AI systems [16–18], including fairness, robustness, explainability, privacy, etc. Devising fairness-aware recommender systems directly contribute to the trustworthiness of RSs [18]. Moreover, studying the connections between fairness and other aspects (e.g., robustness [19]) benefits the comprehensive building of trustworthy systems [16,17]. Finally, compliance with laws and regulations [14] requires recommender systems to be fair when interacting with people because fairness is one of the cornerstones of keeping social order. For example, discrimination against vulnerable groups of people based on sensitive information (e.g., gender, age, race [20,21]) is forbidden by current anti-discrimination laws [22], which also require that similar people should be treated similarly to ensure equality.

Although the fairness of general machine learning tasks (e.g., image classification) has been extensively explored [23], building fairness-aware recommender systems is no-trivial since the following challenges. **(1)** *Diverse and unique fairness concepts*. Various real-world scenarios of RSs encompass diverse fairness concepts, such as group or individual fairness [24,25], static or dynamic fairness [26,27]. Moreover, the unique characteristics (e.g., severing sellers and buyers simultaneously, and changing recommendation trends driven by time) of recommender systems call for elaborated fairness concepts. For example, the fairness concerning multi-party benefits (e.g., multi-sided fairness) or dynamic evolution (e.g., dynamic fairness [27]) should be taken into account since recommender systems interact with both users and item providers in a dynamic process. Therefore, achieving fairness for different recommender systems remains a complex task. **(2)** *Full life cycle of fairness demands.* Unfairness exists in the whole life-cycle of recommender systems because of their dynamic interactions with users and items (as shown in Fig. 2). This fact requires researchers and practitioners to enhance fairness at each stage of developing recommender systems to avoid learning historical unfairness from the last loop and even amplifying it into the subsequent feedback loop. **(3)** *Adverse effects of improving fairness.* Existing studies have demonstrated the existence of the trade-off between fairness and accuracy in recommender systems [28], which makes designing recommender systems that simultaneously prioritize fairness while maintaining satisfactory performance not easy. Moreover, improving fairness of recommender systems can potentially influence other aspects (e.g., robustness, privacy) of trustworthiness. For example, a recent work shows that individual fairness is at the cost of privacy [29]. Thus, evaluating the influence of fairness should be taken into account when comprehensively building fairness-aware recommender systems. Therefore, it is imperative to summarize current efforts and advancements in building fairness-aware recommender systems, which contribute to developing trustworthy, responsible, and socially beneficial AI services.

In this survey, **fair recommender systems** refer to recommender systems that can adapt to different users/items and provide indiscriminate recommendation services to them. Related reviews include surveys on recommender systems, fair recommender systems, and trustworthy recommender systems. Our survey differs from those surveys in that it elaborates on the existing advancements in the fair recommender systems as well as how fairness interacts with other aspects of trustworthy recommender systems. **(1)** Surveys on recommender systems review the advancements of performance-oriented methods on recommender systems. Wu et al. [30] explore the impact of Graph Neural Networks on different categories of recommender systems. Deldjoo et al. [31] investigate the impact of adversarial learning on the security and accuracy of recommender systems. Wu et al. [32] introduce the modeling of collaborative filtering techniques in different types of recommender systems (e.g., content-rich recommender systems, sequential recommender systems). **(2)** Recently, surveys on fair recommendation systems have also emerged. For example, Wang et al. [14] explore various perspectives on unfairness issues and provide an overview of existing approaches, which include data-oriented, ranking, and re-ranking methods. Zehlike et al. [33] delve into how ranking methods impact the fairness of recommender systems. They introduce the fair ranking framework and

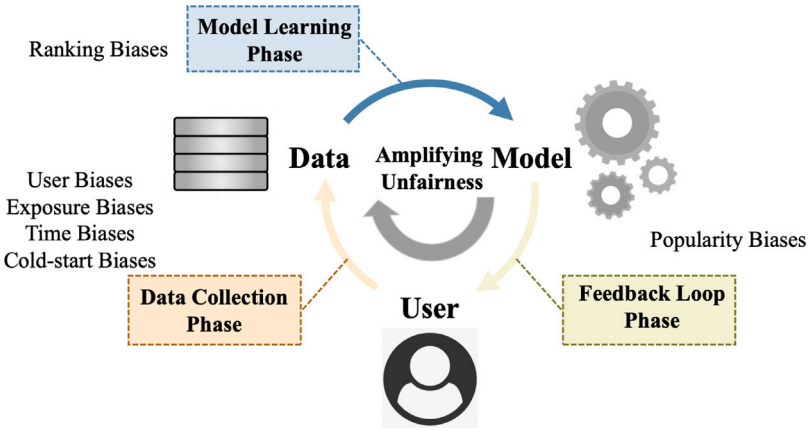


Fig. 2. The lifecycle of recommender systems. In the lifecycle, the data collection phase collects data from users; the model learning phase feeds data into the model for training; the loop phase provides recommendation results to users. In the lifecycle, biases in each phase will have an unfair impact on the recommender system. With the operation and feedback loop of recommender systems, existing biases can be potentially amplified in the following phases of recommender systems.

summarize the evaluation of fair ranking methods. Chen et al. [34] describe the existence of biases in recommender systems and describe ways to address them. Unlike these surveys, our survey categorizes current methods into pre-processing, in-processing, and post-processing methods; we also present comprehensive and fine-grained taxonomy on each of them (e.g., data re-labeling, data re-sampling, and data modification in pre-processing methods). **(3)** Current surveys on trustworthy recommender systems allocate their attention to several different aspects of trustworthiness (e.g., explainability, robustness) [18, 35] or present a conceptual framework of trustworthy recommender systems [36]. In contrast, our survey focuses on comprehensively summarizing current advancements in fairness and discussing the influence on other aspects of trustworthy recommender systems from the view of enhancing fairness.

We introduce the theory and practice of fairness-aware recommender systems in this survey. As a first step, we describe the concepts and formulations of recommender systems and fairness in Section 2. Then, we categorize and illustrate methods promoting fairness in different stages of recommender systems in Section 3 after carefully analyzing related literature regarding fairness in recommender systems. Specifically, these processing stages include pre-processing, in-processing, and post-processing. In Section 4, we collect and collate datasets and evaluation metrics used in literature exploring fairness in recommender systems. Following that, we discuss the industrial applications of recommendation systems that consider fairness in e-commerce, education, and social activities in Section 5. Moreover, Section 6 explores the connection between fairness and other ethical principles of a trustworthy recommender system such as explainability, robustness, privacy, and so on. Section 7 provides a big picture of future directions of fairness-aware recommender systems from different dimensions, including concepts, frameworks, trade-off, and trustworthiness. Last but not least, Section 8 summarizes this survey's importance and influence within the context of trustworthy recommender systems. Fig. 1 illustrates the organizational layout of this survey, as well as the logical connections between each section. The contributions of this survey can be enumerated as follows:

- **A Holistic Taxonomy of Fairness-aware RS Methods.** Compared to previous surveys, this study presents a more comprehensive taxonomy for methods that improve fairness in recommender systems. These methods are grouped in accordance with their roles in the three phases of implementing recommender systems: pre-processing, in-processing, and post-processing. A useful and intuitive framework is provided for the adoption or understanding of these methods by interested researchers.
- **Building Connections between Fairness and Other Ethical Principles.** This study demonstrates how fairness relates to other ethical principles in trustworthy recommendation systems. It encourages people to consider holistically while advocating fairness. A question such as, does promoting fairness affect other ethical dimensions of trustworthy recommendation systems, e.g., explainability and robustness, should be asked.
- **Evaluation of Existing Challenges and Future Direction.** We highlight the existing limitations and challenges of existing fairness-promoting methods for recommender systems. In future works, these problems should be further considered and addressed. In particular, the definition of fairness is not consistent between studies, which can easily cause confusion. Additionally, despite improvements in fairness, existing methods neglect the relationship between fairness and other ethical dimensions of trustworthy systems, which can deteriorate other ethical metrics.

2. Concepts and formulations

2.1. Recommender systems

A recommender system serves as an information filtering system that learns and predicts the user's interest in an item [37]. According to relevant information such as items, users, and the interaction between them, the recommender system provides users with personalized services and recommends suitable items to them [38]. Given a user $u \in U$, an item $v \in V$, where U is a user set and V is an item set, the goal of the recommender system is to learn an information filter $f(\cdot)$ to capture user preference. The predicted score $y_{u,v}$ for the user's preferences in items is [30]:

$$y_{u,v} = f(u, v). \tag{1}$$

The lifecycle of the recommender system can be abstracted into a feedback loop composed of users, data, and models. The loop consists of three phases as shown in Fig. 2, including, a data collection phase, a model learning phase, and a feedback phase. The data collection phase is established between users and data, and it collects users' and items' attribute information, as well as user–item interaction data. The learning phase is built with the data and the model. In specific, this refers to the development of a recommendation model based on the collected data. A recommender system utilizes historical data to forecast the probability of an item being recommended to a user. This step delivers the suggested results to users to satisfy their information needs. This phase will impact users' future behavior and decisions.

The recommender system can be divided into several types according to the different scenarios:

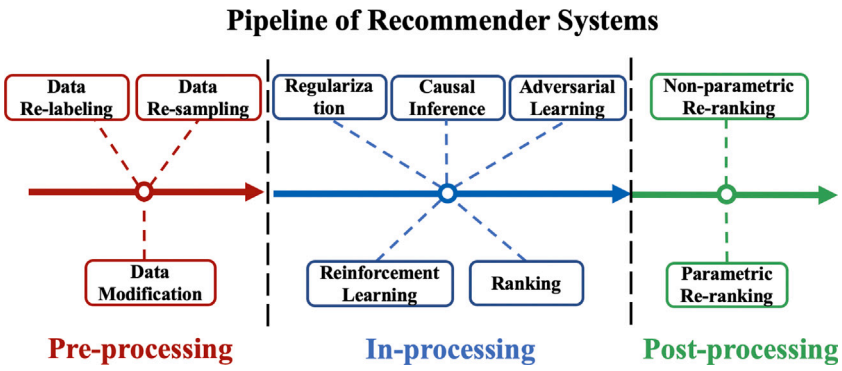


Fig. 3. Fairness-aware methods in the three stages. The pre-processing stage encompasses several methods like data re-labeling, data re-sampling, and data modification. These pre-processing methods are independent of recommendation models. The in-processing stage is mainly to improve the fairness of the model from a method perspective. The in-processing stage involves various techniques, including regularization-based methods, causal-inference-based methods, adversarial-learning-based methods, reinforcement-learning-based methods, and ranking-optimization-based methods. The post-processing stage also employs model-independent methods, treating the model as a black box and solely processing the model's output results. The post-processing methods include non-parametric re-ranking methods and parametric re-ranking methods.

- **Session-based RSs** take each session as the input unit, capturing the user's short-term preferences and dynamic interests as reflected in session transitions [39].
- **Conversational-based RSs** are recommender systems that can interact with users in multiple rounds in real-time, eliciting users' dynamic preferences and taking actions based on their current needs [40].
- **Content-enriched RSs** incorporate some auxiliary data related to users and items to enhance representation learning and semantic relevance. These auxiliary data may include textual content, knowledge graphs, etc [32].
- **Social RSs** are defined as any recommender system for the social media domain. This type of recommender systems improves recommendation performance by incorporating social relations into the recommender system [13].

There are unique unfairnesses in different types of recommender systems, and we introduce these unfairness issues and the biases that generate unfairness in the follow-up.

2.2. Fairness

Fairness is a concept that originated in sociology, economics, and law [41–43]. Its definition in the Oxford English Dictionary is "imperfect and just treatment or behavior without favoritism or discrimination". In the context of recommender systems, fairness requires that recommender systems treat all users and items equally. For example, in loan approval based on recommender systems [12], the systems are fair if the approval of an applicant seeking a loan is not influenced by the user's attributes (e.g., gender, race). We can develop fair recommender systems when bias or unfairness is eliminated from systems [14]. In this section, we first categorize causes of unfairness and then present common fairness expressions that can measure the existence of bias.

2.2.1. Causes of unfairness

As depicted in Fig. 2, the unfairness of a recommender system appears in its whole lifecycle, including the data collection phase, the model learning phase, and the feedback loop phase. According to the position of bias in the three phase, we divide bias in the recommender system into three categories, namely the data bias, model bias, and feedback bias. Subsequently, we will elucidate these biases by showing how they result in unfairness in recommender systems.

Data Bias. Currently, recommender systems are always trained on large datasets, which usually contain user information like user behaviors (e.g. incorrect clicks), sensitive attributes (e.g. gender), and the interaction between users and items. However, training datasets potentially include various biases. These biases can lead to unfair recommendations, which may cause undesirable or even disastrous consequences for human life and society. Next, we will introduce some common biases in datasets.

User Bias. User attribute bias is a common user bias. In recommender systems, some sensitive attributes like age [44], geographical location [45], gender [46], and profession [47] are important sources of information for recommender systems to understand user preferences. However, sometimes users' attributes (e.g., age, gender) can cause biased recommendation results. For example, age may be an effective feature in a music recommender system as they can recommend music to users according to their ages [48]. In general, youngsters have a higher preference for hip-hop music than senior people. However, sometimes users of a certain age group may wish to jump out of the cocoon of age information and explore different kinds of music. For the gender attribute, some researchers [49–51] observe the promotion of high-paying jobs varies by gender. In MOOC, a recommender system attempt to guide course resources to students [52], the geographic location and nationality of teachers also influence which courses are recommended to students [13].

In addition to user attribute bias, user selection bias is also an important part of user bias. User selection bias refers to the behavioral bias when the user selects items, which usually exists in user explicit feedback. The existence of biases in user feedback data can lead to inconsistencies between user preferences and behavior records [53]. For instance, music streaming media recommender systems provide playlists based on the music preference of users [54], which can be affected by the user feedback bias. Specifically, during the recommendation process, the model is updated online based on user feedback. In a music recommender system, explicit feedback (e.g., ratings of items [55]) can be the track marked "favorite" by the user. However, users may have wrongly clicked like tags, which will lead to explicit feedback bias [56].

Exposure Bias. Exposure Bias refers to the situation where users only have access to a portion of the available item set [57]. The exposure bias is often present in implicit feedback data, which usually refers to whether the user interacts with the item, including some purchases, clicks, and other behaviors [58]. For example, e-commerce sites like Amazon allow users to provide feedback on recommended items beyond the user interface by searching and browsing various product pages. The user's browsing behaviors may contain incorrect clicks, and these mistakes can make the recommender system misjudge the user's preference, which results in an unfair recommendation [11]. In a music recommender system, a handful of popular artists may garner the majority of traffic, thereby underexposing less mainstream artists. If this bias remains unaddressed, recommender systems could adversely affect the experience of diverse users and items on the platform due to continuous interaction with biased recommendations,

and thus the training of models using biased interactions in subsequent timeframes [26]. In music streaming media recommender systems, the explicit feedback can be the track marked "favorite" by the user, and implicit feedback can be the number of times a track has been played. However, users may pay less attention to the music app while doing other activities, such as exercising, reading late at night, or commuting, and the music is looped multiple times, which can lead to an invisible feedback bias [56].

Time Bias. Some time-related recommendations, such as news recommendations, session-based recommendations, and job recommendations, heavily rely on direct user–item interactions to understand user preferences and provide specialized recommendations. However, freshly released data could unfairly overrepresent users' long-term interests, which is unfair to learn users' preferences. If we continue to disclose items at time $t+n$ according to the fairness limitations at time t, even though an item that is popular at time t may no longer be popular at time $t+n$, we will neglect the long-term fairness dynamics [27]. This will ignore the long-term dynamic process of fairness, which leads to recommended bias. For example, the goal of job recommendations is to recommend job advertisements to job seekers. Job seekers prefer to click after seeing a new job advertisement. New job advertisements obtain higher click rates as a result of this behavior than older job advertisements. Consequently, jobs of longer-lasting professionals are less likely to be recommended. However, job seekers are driven by their long-term career aspirations when contemplating jobs. Advertising recommendations that are consistent with these preferences are more advantageous to job seekers [59]. Therefore, it is unfair to recommend recent advertisements to users without considering users' long-term interests.

Cold-start Bias. The above biases only consider unfairness in the case of warm-start recommendation. The primary source of unfairness in this instance is data biases (e.g., clicks or pageviews). However, the recommender system will meet a cold start problem when lacks data. When a new user or item enters the system, the cold-start problem occurs at which point the recommender system fails to provide personalized recommendations because it lacks sufficient data on user behaviors or item attributes [60]. Solving a cold-start problem is usually using prior knowledge from warm-start recommendations to train cold-start recommender systems. The cold-start bias refers to the bias of this prior knowledge in warm-start recommender systems. These biases will be brought into the cold-start recommender system when training, and cause the unfairness phenomena [61]. The unfairness phenomena can be particularly problematic because the unfairness caused by cold-start recommendations can persist and accumulate throughout the lifespan of the item, making it increasingly difficult to mitigate unfairness [62].

Model Bias. The model learning phase uses the collected data in the data collection phase to train the recommendation model, the core of which is to deduce user preferences from past interaction data to predict the possibility of users choosing unvisited targets. Models with design flaws (e.g., ranking bias) may further magnify biases in the input data.

Ranking Bias. Some loss functions can further exacerbate unfairness during the training of recommender systems. These loss functions affect the predicted score of the item, which in turn affects the recommendation ranking list of recommender systems. For example, Wan et al. [63] demonstrate that point losses (e.g., MSE loss) and pairwise losses (e.g., BPR loss) are sensitive to popular items. These loss functions give popular items higher scores than unpopular items, which amplifies exposure bias during the training of recommender systems. Zhu et al. [64] also prove that the BPR loss lacks the fairness constraint of equal opportunity ranking for reducing bias.

Feedback Bias. A feedback loop exists in every recommender system, which provides recommendation results of a model to users for selection. In the feedback loop, users and a recommender system interact and co-evolve. Users' preferences and behaviors are updated through the recommender system, and the recommender system uses the updated data for self-reinforcing. This feedback loop mechanism not only generates bias, but also exacerbates the bias over time, resulting in a gradual deterioration of the fairness ecosystem. In the following paragraph, we introduce the popularity bias, which is a typical feedback bias.

Popularity Bias. A few popular items are frequently recommended, while most others are disregarded. Users consume these recommendations, and their responses are recorded and added to the system. Over time, the recommender system recommends popular items to users, continuously collects user feedback on popular items, and adds them to the training set, making the data distribution more unbalanced, which will result in more and more recommendation results focusing on popular items [65]. The existence of popularity bias will also constantly change the user's preference representation, making it challenging for the recommender system to capture the user's true preference. For example, Naghiaei et al. [66] investigate the impact of popularity issues on book recommender systems. Their work shows that recommender systems tend to recommend popular items frequently, and there is a strong correlation existing between the popularity of books and the frequency with which they are recommended. Most books are not exposed to users by the recommender system, while popular books are highlighted more frequently.

2.2.2. Fairness expressions in recommender systems

A recommender system can be employed in multiple scenarios, and the unfairness concerns in each scenario are different. Next, we present the expressions of fairness in recommender systems from different perspectives.

Individual Fairness vs. Group Fairness. Given two similar users $u_i, u_j \in U$, individually fairness-aware recommender systems hope to give similar prediction scores for samples u_i and u_i [24]. For example, in a healthcare recommendation system, two patients exhibiting similar pathologies should receive recommendations of equivalent quality [67]. Methods for individual fairness solve the problem of statistical equality through pairwise comparisons between similar users. The formal definition of individual fairness can be expressed as $f(u_i, v) \approx f(u_j, v)$. where $f(\cdot)$ is a recommender system, v is an item.

Group fairness necessitates that protected groups receive treatment akin to that of advantaged groups [25]. As a typical group fairness, demographic parity [15] requires that the probability that the protected group (e.g., female group) is predicted to be a positive sample (e.g., job offers) is equal to the probability that the advantaged group (e.g., male group) is predicted to be a positive sample. Given a protected group U_i and an advantaged group U_j, the demographic parity of the recommender system can be formalized as $Pr(y_{U_i,v} = 1|U_i) = Pr(y_{U_j,v} = 1|U_j)$.

Static Fairness vs. Dynamic Fairness. Static fairness is defined as providing a short-term fair recommender system regardless of changes in the recommendation environment. Zhang et al. [68] propose that the concept of recommendation fairness proposed by most methods is static fairness, i.e., the protected group is fixed during the recommendation process. As an example, traditional matrix-based recommendation strategies are aimed at maximization of users' immediate gratification, assuming that their preferences remain static. However, they potentially ignore users' long-term interests.

Dynamic fairness [27] is defined as considering dynamic factors in the environment to maintain fairness. Dynamic fairness is related to time bias. Some works argue that recommending an item that was recommended a long time ago should have the same probability as recommending an item that was recommended recently. Assuming v_t represents an item appears in time t and v_{t+1} represents an item appears in time $t+1$, the dynamic fairness can be defined as: $Pr(y_{u,v_t} = 1|v_t) = Pr_t(y_{u,v_{t+1}} = 1|v_{t+1})$. In real-world recommender systems, it can also be expressed as $\frac{Pr(y_{u,v_t}=1|v_t)}{Pr(y_{u,v_{t+1}}=1|v_{t+1})} < \xi$, where ξ is a slack factor used to adjust the dynamic fairness granularity for recommender systems.

Table 1
Summary of methods. We classify existing methods based on the phase of implementation, the technology utilized, the bias issues encountered, the suggested application scenarios, and the official method names.

Stage	Technologies	Bias	Background	Methods
Pre-processing	Data re-labeling	Popularity Bias	General RSs	IFNA[a] [69]
		User Bias	General RSs	DPTC[a] [70]
	Data re-sampling	User Bias	General RSs	MPML[a] [71], HPO [72], RNS [73], HFD[a] [74]
	Data modification	User Bias	General RSs	HURR[a] [75], CFAI[a] [76], RSNMF [77], EPTB[a] [78]
In-processing	Regularization	Cold-start Bias	Cold-start RSs	CLOVER [79]
		Exposure Bias	Social RSs	MinDiff [80], FRRPC[a] [81], SERec [82], IDLR[a] [83]
		Popularity Bias	General RSs	FARL[a] [84], CPBLR[a] [85]
		User Bias	General RSs	CFIFD[a] [86], FFFU[a] [87], SLIM [88], FaiRecSys [89], MMIUR[a] [64], F2VAE [90], FairRec [91], IURPF[a] [92], FRRP[a] [81], FATBR[a] [93]
	Causal inference	Cold-start Bias	Cold-start RSs	CLOVER [79]
		Exposure Bias	Social RSs	DHRS[a] [94], ABPUL[a] [95]
		Popularity Bias	General RSs	IANP[a] [96], PDA [97], MACR[a] [98], DANCER [99]
		User Bias	Content RSs	FairTED [100]
			General RSs	CSBRS[a] [101], SHT [102], RTDLE[a] [103], TPFB[a] [104], AdaRequest [105], InvPref [106], DeSCoVeR [107]
			Session-based RSs	AIPB[a] [20]
	Adversarial learning	User Bias	General RSs	FairRec [91], FRFC [108], CGWG[a] [109]
	Reinforcement learning	Exposure Bias	Social RSs	FCPO [27], MORL [110]
		Popularity Bias	Conversational RSs	Popcorn [111]
	Ranking	Exposure Bias	Social RSs	RDC [112]
		Popularity Bias	General RSs	CPR [63]
		User Bias	General RSs	MMIUR[a] [64]
	Others	Exposure Bias	Session-based RSs	CLRec [113], SAR-Net [114]
		User Bias	Conversational RSs	CUAB[a] [56]
		User Bias	Social RSs	FairSR [115]
Post-processing	Non-parametric re-ranking	Exposure Bias	Cold-start RSs	GEN [62]
			General RSs	CPFair [116], FairMatch [117], TFROM [118]
	Parametric re-ranking	Exposure Bias	General RSs	HyperFair [119]

[a]Indicates the method has no specific name. We named it with the abbreviation of its article name.

Single-sided Fairness vs. Multi-sided Fairness. There are multiple stakeholders in a recommender system. Single-sided fairness refers to maintaining the interests (i.e., fairness) of a single role. For example, when recommending items to users, it only pays attention to whether the user has received a fair recommendation. Multi-sided fairness refers to maintaining the interests of multiple roles [116]. For example, consumers expect the recommender system to fairly recommend products they are interested in, while suppliers wish their products to be fairly exposed to consumers. Multi-sided fairness aims to maintain the fairness of both consumers and suppliers. A typical multi-sided fairness is to increase diversity [120] in recommendation results, ensuring that the average exposure of items across multiple aspects is balanced [121]. In addition, it is plausible to enforce proportional exposure of both groups relative to their average utility, i.e., $\frac{\sum_{i=1}^{|V_i|} Pr(y_{u,V_i}=1|V_i)}{|V_i|} = \frac{\sum_{j=1}^{|V_j|} Pr(y_{u,V_j}=1|V_j)}{|V_j|}$, where V_i and V_j are two different item groups.

Others. Other common perspectives of fairness in recommender systems are:

- *Ranking Fairness.* The fairness of sorting is usually reflected in statistical equality or equality of opportunity. A simple interpretation of statistical parity in ranking is the assurance that the proportion of protected individuals appearing within a ranking prefix exceeds a predetermined threshold [122]. In general, ranking fairness asks that similar items or groups of items receive similar visibility, they appear at similar positions in the ranking [123].
- *Causal Fairness.* Most recommender systems are based on statistics; however, this ignores causality in the original data. Unlike individual fairness and equal opportunity, causal fairness not only considers observational data but also incorporates additional causal relationships within them [97]. Counterfactual fairness is a typical causal fairness. It demands that the predictions of recommender systems remain consistent in both the factual and the counterfactual world [23,124].
- *Long-term Fairness.* Long-term fairness is affected on time [110] and considers the long-term impact of fairness interventions [125]. It is similar to dynamic fairness in that it considers the long-term impact on user preferences [126].
- *Envy-freeness Fairness.* Do et al. [127] propose a criterion of envy-freeness fairness, stating that each user should prefer their own recommendations over those of other users.

Other fairness concepts, like Rawlsian Maximin Fairness or Maximin-Shared Fairness, can be found in recent literature on fairness [14].

3. Methods for fairness-aware recommender systems

In this section, we summarize methods for improving the fairness of recommender systems and present them in three stages of implementation, including pre-processing, in-processing, and post-processing (as shown in Fig. 3). In addition, we summarize the fairness-aware recommendation methods in Table 1, specifically including the biases the methods can address, the types of fairness-aware recommendation methods, and other important information.

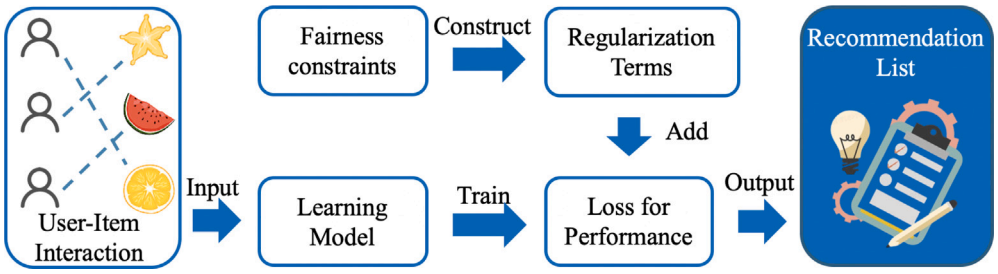


Fig. 4. Regularization-based fairness-aware methods.

3.1. Pre-processing methods for fairness recommender systems

In the pre-processing stage, fairness-enhancing methods are used to minimize biases in the training datasets. It is possible for these biases to be amplified throughout the lifecycle of a recommender system, resulting in unfair recommendations for users. We divide these fairness-aware pre-processing methods into three categories according to data debiasing methods.

3.1.1. Data re-labeling

Re-labeling changes the labels of the training dataset to remove biases in the input data. Some recommender systems predict user preferences using binary implicit feedback labels, like "click" or "not click". However, there may be noise in the feedback labels, i.e., clicks do not necessarily represent positive feedback, and missing clicks do not necessarily represent negative feedback. These "noisy" labels can lead to a drop in model performance. Wang et al. [69] design a self-supervised re-labeling framework for noise in implicit feedback. The framework dynamically generates pseudo labels for user preferences to mitigate noise in both observed and unobserved feedback data.

3.1.2. Data re-sampling

Unbalanced data distribution may lead to a decrease in the training effect of recommender systems. Montanari et al. [72] design a data sampling algorithm to ensure that the sampling is uniform and not affected by the distribution. They randomly select a certain percentage of users and delete their interaction information. This type of re-sampling method reduces the size of the dataset for approximately the same proportion of users and maintains the dynamic nature of user profiles. Celis et al. [74] propose a subsampling method that proportionally subsamples the dataset based on different sensitive attributes. Ding et al. [73] design a negative sampler that creates data resembling generates data similar to the exposure data through feature-matching techniques instead of selecting directly from exposure data. The sampler forces the distribution of positive and negative data to be balanced by adding negative samples.

3.1.3. Data modification

The main idea of data modification is to augment or modify the biased data to reduce the bias. For example, some recommender systems may utilize textual data (e.g. job recommender systems, news recommender systems), which may be missing or have errors. This incomplete or noisy data can lead to data bias. To solve the noisy data problems, Wang et al. [76] implemented natural language processing (NLP) pre-processing techniques to modify the training data. Specifically, in a crowdtesting recommendation (e.g., recommending software testing tasks to professionals), they first perform standard word segmentation for each document. Then, they remove stop words and apply synonym substitution to reduce noise. In addition, they construct a descriptive term list and perform term filtering for each document. Similarly, Sachdeva et al. [75] use a similar NLP approach as Wang et al. In addition to handling noisy data, some works focus on missing data. Collaborative filtering-based recommenders typically model user preferences as a user–item rating matrix. Since it is not possible for the user to interact with all the items. The rating matrix is thus high-dimensional and sparse (HiDS), with many missing data representing the user's unobserved preferences. Some works [77,78] apply a latent factor-based model, which provides a good job of handling the high-dimensional and sparse (HiDS) matrices, to process missing data.

3.2. In-processing methods for fair recommender systems

Currently, most recommender systems (e.g., collaborative filtering methods [96]) are devised to extract user preferences by learning the correlation in the training data [102]. However, people potentially suffer from unfairness services when using these correlation-oriented recommender systems, such as Simpson's paradox [128], popularity bias [97], user-oriented bias [105], cold-start bias [79], to name only a few. In addition to removing bias via pre-processing methods, many works design fairness-aware methods to alleviate or even eliminate unfairness during model training of recommender systems. In-processing fairness-aware methods aim to learn bias-free models. In this section, we classify these works into five categories including regularization-based methods, casual-inference-based methods, adversarial-learning based, reinforcement-learning-based methods, ranking methods, and others.

3.2.1. Regularization and consternation for fairness

To alleviate unfairness in recommender systems, regularization penalizes the predicted recommendation score in accordance with the fairness evaluation [34]. Regularization terms are widely used in various recommendation scenarios to reduce bias, such as user-oriented bias [79], group bias [81], exposure bias [82], popularity bias [84], etc.

In general, a regularization term is usually regarded as an additional loss focusing on promoting fairness. The term is added to the main loss, which is mainly responsible for improving the recommendation performance. Specifically, the total loss used for recommendation training is the sum of the performance loss and the regularization term:

$$\mathcal{L} = \mathcal{L}_{per}(\mathbf{u}, \mathbf{v}) + \mathcal{L}_{reg}(\boldsymbol{\theta}), \tag{2}$$

where L is the total loss for training, $L_{per}(\cdot)$ is the loss for optimizing the recommendation performance, $L_{reg}(\cdot)$ is a customized loss function, and θ refers to all possible parameters related to fairness evaluation. Fig. 4 shows a general framework for regularization-based fairness-aware methods. According to the composition of the regularization term, we can divide the regularization into three categories, including norm-based regularization terms, matrix-based regularization terms, and pair-wise regularization terms.

Norm-based Regularization. The norm calculates the distance between raw features and generated embeddings. It is used to evaluate the deviation of the model of learnable recommender systems. Burke et al. [88] propose the possibility of using l_1 norm, and l_2 norm as

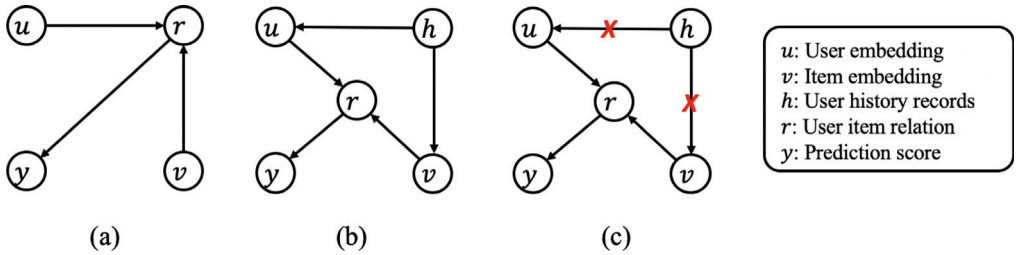


Fig. 5. A causal graph is a directed acyclic graph, wherein each node stands for a random variable, and the directed edges indicate causal relationships. In a recommender system, the recommendation process can be briefly described as calculating the matching score r between user u and item v, to predict the recommendation score Y. As shown in (a), where u is the user's representation, v is the item's representation, h represents the user's history, and y is the prediction score. $u \rightarrow r$ denotes that u is an inducement of r, and R has been effected by a direct causal from u. Similarly, there is no arrow between u and y, indicating that u has no direct causal relationship with y. The popularity bias, a prevalent form of unfairness in existing recommender systems, results from users repeatedly clicking recommended items and then recommender systems always advocating for these items. As shown in (b), the history h influences u and v, which determine r. (c) shows a method to decrease the effect of h on u and v.

regularization terms. Among multi-sided unfairness issues, user neighborhoods can constrain deviance in their opinion. The regularization term can be described as:

$$\mathcal{L}_{reg} = \lambda_1 \|W\|^1 + \frac{\lambda_2}{2}\|W\|^2 + \frac{\lambda_3}{2}\sum_i^n (b_i)^2, \tag{3}$$

where W is a user–user weight matrix, $\|\cdot\|^1$ is l_1 norm, $\|\cdot\|^2$ is l_2 norm, and b_i is a neighborhood balance regularization for reducing the probability of user neighborhoods forming. b_i is the squared difference between the weights of the protected users versus the unprotected users. Protected users are usually including sensitive attributes, and unprotected users do not have sensitive attributes. The works [87,93] employ similar techniques to alleviate group unfairness. Hu et al. [86] use the l_2 norm of user embeddings and item embeddings as a regularization term.

Matrix-based Regularization. Some regularization terms are in the form of matrices. Abdollahpouri et al. [85] introduce a matrix-based regularization LapDQ to reduce the popularity bias. The LapDQ regularizer is defined as :

$$\mathcal{L}_{reg} = tr(V^T L_D V), \tag{4}$$

where V is the item embedding matrix, $tr(\cdot)$ is the trace function, and L_D is the Laplacian of the dissimilarity matrix D. Wasilewski et al. [83] and Edizel et al. [89] also use this type of regularizer for ranking unfairness and user bias.

Correlation-based Regularization. This type of method mainly exploits correlation to reduce bias. Beutel et al. [81] propose a correlation-based pair-wise regularization term to balance the clicked and unclicked item, which is defined as:

$$\mathcal{L}_{reg} = |Corr(A, B)|, \tag{5}$$

where $Corr(\cdot)$ calculates the absolute correlation of two random variables, A and B are two random variables. The specific meaning of A in this work is the residuals between clicked and unclicked items. Variable B means the correlation between group users of the clicked items. The model is penalized if it predicts that one group clicked on an item more than the other group did. Prost et al. [80] propose a MinDiff formulation based on the above method. MinDiff minimizes the correlation of predicted probability distribution and the distribution between the clicked items and unclicked items.

Others. To further eliminate data bias in the model training, Wu et al. [91] propose an orthogonality regularization to orthogonalize the unbiased user embeddings to the biased user embeddings. It thus distinguishes between embeddings that are unbiased and those that are biased. For each user u, the orthogonality regularization can be defined as:

$$\mathcal{L}_{reg}(u^b, u^d) = |\frac{\mathbf{u}^b \cdot \mathbf{u}^d}{\|\mathbf{u}^b\| \cdot \|\mathbf{u}^d\|}|, \tag{6}$$

where $\mathbf{u}^b$ and $\mathbf{u}^d$ are the bias-aware and bias-free embeddings, respectively. For group fairness, Boratto et al. [92] customized a special regularization term to reduce the sensitive attribute in group fairness.

$$\mathcal{L}_{reg} = (\frac{\sum_i^n f(u,v) \cdot S(v)}{\sum_i^n f(u,v)} - C)^2, \tag{7}$$

where $S(v)$ represents the percentage of users who have interacted with item v, C represents the proportion of interactions between groups with sensitive attributes and a certain type of item in all interactions. This regularized optimization implies that the model is penalized if the difference between correlation and contribution of the population is significant. Zhu et al. [64] use Kullback–Leibler Divergence to normalize user prediction scores to a normal distribution, reducing the ranking unfairness in model training.

3.2.2. *Causal inference for fairness*

The causal inference in artificial intelligence explores the causal relationships between variables, i.e., how one variable determines another variable. In this survey, *causal inference for fairness* represent methods (e.g. inverse propensity score [95,129]) that trace to the source of bias and then mitigate unfairness through causal inference [98,104,106]. As shown in Fig. 5, causal graphs, which visually represent causal relationships between variables in a recommender system, are used to analyze the causes of unfairness. In this section, based on the usage of causal graphs, we categorize current causal inference methods for fairness into inverse propensity scoring, backdoor adjustment, and counterfactual inference.

Inverse Propensity Score. By analyzing the causes of bias in the causal graph, the inverse propensity score (IPS) reweights samples in order to reduce the influences of biased samples, without changing the causal relationship between variables [94]. In Fig. 5(b), for instance, the confounding variable h affects both u and i, the probability distribution $\hat{Pr}(h,u,v,r)$ in this figure can be expressed as

$$\hat{Pr}(h,u,v,r) = Pr(r|u,v,h)Pr(u|h)Pr(v|h), \tag{8}$$

while in the absence of the confounding variable (i.e., Fig. 5(c)), the distribution is:

$$Pr(h,u,v,r) = Pr(r|u,v,h)Pr(u)Pr(i)Pr(h). \tag{9}$$

To remove the bias brought by the confounding variable h, IPS methods expect $\hat{Pr}(h,u,v,r) = Pr(h,u,v,r)$, which can be achieved through multiplying $\hat{Pr}(h,u,v,r)$ by a propensity weight, i.e.,

$$Pr(h,u,v,r) = \frac{Pr(h)}{Pr(u|h)Pr(v|h)}\hat{Pr}(h,u,v,r). \tag{10}$$

Recently, some methods using IPS have been studied to conduct debiasing. A propensity framework proposed by [95] makes propensity estimations for improving exposure fairness in implicit feedback scenarios. Xiao et al. [129] fuse a deep variational information bottleneck approach with a propensity score to develop an unbiased learning

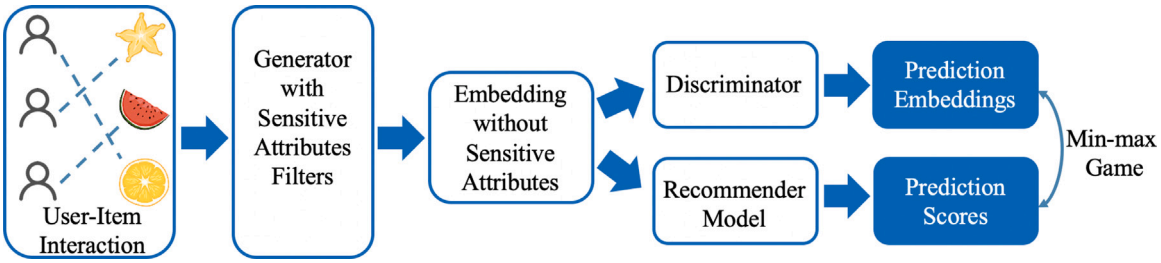


Fig. 6. Adversarial-learning-based fairness-aware methods. The adversarial-learning-based methods first use a generator and a sensitive attribute filter to remove sensitive information, resulting in embeddings without sensitive information, which are used for recommendation predictions. The corresponding sensitive attributes are predicted by the discriminator from the filtered embeddings. The generator and the discriminator engage in a max–min game.

algorithm. Since user preferences might drastically change over time, Huang et al. [99] exploit a dynamic inverse propensity score for debiasing dynamic popularity biases. To address user behavior bias, Wang et al. [101] use unbiased data to introduce propensity scores into biased training of recommender systems. Another recent work [130] analyzes the data bias mechanism in the sequential recommendation and reweights the training parameters to reduce bias using inverse propensity scores.

Backdoor Adjustment. Unlike the reweighting operation in IPS methods, *backdoor adjustment* achieves fairness by blocking off relationships that lead to biases. When removing confounding factors, backdoor adjustment methods require the causal relationship between variables satisfies the backdoor criterion. Here, the variable set Z satisfies the backdoor criterion on a causal relationship $u \rightarrow v$ in the causal graph, indicating Z satisfies (1) there is no descendant node of u in Z, and (2) Every path between u and v that leads to u is blocked by Z. For example, as shown in Fig. 5(c), the history H affects the representation of u and v. To block the influence of h on u and v, backdoor adjustment methods employ a do operation in cutting off the edge of $h \rightarrow u$ and the edge of $h \rightarrow v$, which can be formulated as $Pr(y|do(u), do(v)) = Pr(y|u, v)$.

Although the backdoor adjustment is effective in removing unfairness, it faces efficiency challenges resulting from an unlimited sample space of confounding factors. To this end, Wang et al. [131] introduce DecRS, a model that uses data approximation and KL divergence to adjust the backdoor criterion. To lessen document-level label bias in text-contained recommender systems, DeSCoVeR [107] uses causal backdoor adjustment and sentence-level keyword bias elimination techniques in a semantic context. An inference model involving popularity-bias deconfounding and adjusting (PDA) is proposed by Zhang et al. [97] as a new inference approach. It employs backdoor adjustment during model training to eliminate confusion caused by popularity bias.

Counterfactual Inference. Counterfactual inference methods for fairness construct a counterfactual causal graph [98] based on the real causal graph through some fairness-concerning actions (e.g., changing the values of some sensitive attributes like gender, age, race [20,21]). Recommender systems are fair and unbiased (i.e., counterfactual fairness in Section 2) if the recommendation results in the real world and the counterfactual world are the same. The intuition of counterfactual inference methods can be understood as "if a fairness-concerning action (e.g., modifying the value of gender) cannot change recommendation results, then results from the recommender systems are not affected by the action-object (e.g., gender)". A representative counterfactual inference method for fairness is the framework called MACR [98], which is proposed to mitigate popularity bias. MACR assumes that the recommendation interaction matrix I_{uv} is affected by user u, item v, and the user–item matching ranking score $\hat{y}_r$, i.e., $y_{uv} = \hat{y}_r * \sigma(\hat{y}_v) * \sigma(\hat{y}_u)$, $\hat{y}_v$ indicates the influence from item popularity, and $\hat{y}_u$ represents the extent of the user u interact with items. The higher value of the $\hat{y}_u$, the more likely the user is affected by the popularity of the item. Through counterfactual inference, the causal graph of the real world is transformed into the causal graph of the counterfactual world. The causal relation $I \rightarrow Y$ is removed by the following formula to alleviate the item popularity bias problem:

$$\hat{y}_r * \sigma(\hat{y}_i) * \sigma(\hat{y}_u) - c * \sigma(\hat{y}_i) * \sigma(\hat{y}_u), \tag{11}$$

here the hyperparameter c controls the influence of user and item properties on the prediction result. The inference can be interpreted logically as a ranking adjustment based on $\hat{y}_{ui}$.

FairTED [100] creates counterfactual samples of sensitive attributes to make sure that the speaker's sensitive attribute (i.e., gender) cannot influence the TED talk quality prediction. Specifically, when generating counterfactual samples, the score of presentations with female speakers are assigned as the score of presentations with the same contents and male speakers. These counterfactual samples are added into the training dataset to develop recommender systems with counterfactual fairness on gender. To mitigate the popularity bias and improve explainable fairness, Ge et al. [132] propose a framework called CEF, which uses counterfactual inference (i.e., introducing small changes in the features) to find the root cause of the model's bias. In CEF, the scores of each feature, which are calculated from counterfactual recommendation results, are regarded as fairness explanations. Moreover, Li et al. [104] counterfactually infer that user-sensitive features should be orthogonal to user embeddings, and make fair personalized recommendations by removing user-sensitive features. In the music streaming media recommender system, the user may have a situation where the music is playing incorrectly. Zhang et al. [56] suggest a counterfactual learning strategy to correct user feedback that has been incorrectly categorized.

3.2.3. Adversarial learning for fairness

Adversarial learning is a method commonly used in recommender systems to remove sensitive attributes. An adversarial-learning-based fairness-aware framework generally consists of a generator that produces node embeddings and a discriminator that predicts sensitive features from these generator outputs. By playing a min–max game with the discriminator and generator, adversarial-learning methods are able to learn fair representations. Passing a negative gradient by predicting the sensitive attribute enables the model to fool the discriminator, so that the information content of the sensitive attribute is continuously reduced. When the discriminator cannot predict the sensitive feature value, the output of the generator is considered to be decoupled from the sensitive feature. The general form of loss for adversarial-learning-based fairness-aware methods can be formulated as:

$$\mathcal{L} = \mathcal{L}_{per}(z, y) + \mathcal{L}_{adv}(y, s), \tag{12}$$

where z is a set of the generated representations of the generator, y is a set of predictions, and s is a set of the predictions of the discriminator. Fig. 6 shows a general framework for adversarial-learning-based fairness-aware methods. Recent work [79] is a study addressing cold-start bias. To swiftly adjust to cold-start new users, it suggests a comprehensive fair meta-learning framework (CLOVER) to gather a

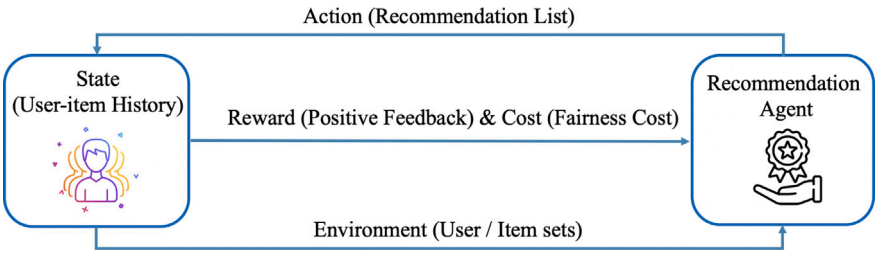


Fig. 7. Reinforcement-learning-based fairness-aware methods. A recommender system retains the learned experience in the process of interacting with the environment. When interacting in the next round, the behavior with the largest positive feedback and the smallest fairness cost will be selected.

general understanding of user preferences. CLOVER establishes that recommenders with ratings based on these representations will also fulfill counterfactual fairness if individual fair adversarial games converge to optimal solutions. Through adversarial learning, CLOVER improves the fairness of the cold-start problem in recommender systems. Based on sensitive attributes, Wu et al. [91] craft a bias-aware user embedding, which is specially used to capture the bias. Furthermore, it also learns a bias-free user embedding, which is only used to encode attribute-independent information that the user interests. The fairness of the model is guaranteed by orthogonalizing the two types of embeddings. Liu et al. [108] propose an adversarial graph neural network (GNN) to prevent users from being affected by sensitive features of neighboring users. In particular, they propose two fairness constraints to address the failure and inefficiency of adversarial classifiers in the training data. Rus et al. [109] use adversarial learning to mitigate gender bias in word embeddings obtained from recruitment-related text, which aims to provide unbiased job recommendations to job applicants. To alleviate multi-sided fairness. Liu et al. [133] design explicit and implicit adversarial fairness discriminators. Explicit discriminators aim to address biases from a local perspective, while implicit discriminators focus on addressing biases from a global perspective. The fairness generator and discriminator are trained adversarially together to mitigate bias. In their paper, Wu et al. [48] propose a fairness awareness framework relying on prompts-based bias eliminators in combination with adversarial training. Li et al. [134] devise a generative adversarial network (GAN), named FairGAN, designed to generate negative signals of users to ensure data fairness. These signals enable FairGAN to complete the best item exposure ranking.

3.2.4. Reinforcement learning for fairness

Recently, some studies take the dynamic interactions among users, items, and recommender systems into consideration, and model this feedback loop as a Markov Decision Process (MDP) [135]. With this view, reinforcement learning (RL) methods are used in training recommendation strategies from users' historical information to learn their preferences. In this survey, *reinforcement learning for fairness* indicates methods that introduce fairness-concerning feedback (e.g., dynamic fairness) during the training of RL-based recommender systems. Fig. 7 shows a general framework for reinforcement-learning-based fairness-aware methods.

Long-term fairness and dynamic fairness are two typical concepts concerning mitigating time bias (see Section 2), which can be involved in the training of RL-based recommender systems. Here we introduce some representative reinforcement learning for fairness methods. Ge et al. [27] majorly focus on the unfairness of item exposure across groups caused by time bias. According to their opinion, fairness constraints ought to change over time. For instance, an item may no longer be popular at time $t + n$, but if it is still exposed in accordance with the fairness requirements from time t earlier, the long-term dynamic changes in fairness are disregarded. To enable the model to dynamically adjust the recommendation strategy and guarantee that the fairness requirements are always satisfied with environmental changes, they propose a fairness-constrained reinforcement learning recommender system, modeling the recommendation process as a constrained Markov Decision Process (CMDP). CMDP proposes two dynamic fairness constraints for reinforcement learning. The first is the population equality constraint, which requires equal average exposure for each group of items. This constraint is enforced at all reinforcement learning iterations. The second is the exact-k fairness constraint, which requires that the length of protected candidates in each recommendation list is statistically below a given threshold. In interactive recommender systems (IRS), Liu et al. [136] dynamically maintain the long-term trade-off between accuracy and fairness by offering a method called FairRec. Using reinforcement learning, recommendations are generated by combining user preferences with system fairness in FairRec. In addition, FairRec introduces a concept called weighted proportional fairness to ensure the fairness of item exposure.

Moreover, some approaches have proposed employing multi-objective reinforcement learning to improve fairness. For instance, Ge et al. [110] investigate the Pareto optimal/effective fairness-utility trade-off problem in the recommendation process. Using multi-objective reinforcement learning, they suggest creating a fairness-aware recommendation framework (MoFIR), which introduces conditional networks and modifies the network according to user preferences. Fu et al. [111] propose a multi-objective MDP-based framework, namely Popcorn, for eliminating popularity bias in conversational recommender systems. Popcorn effectively balances recommendation performance and item popularity through a real-time semantic understanding of user history to avoid long-tail effects.

3.2.5. Ranking optimization for fairness

Ranking is an important part of the recommendation algorithm. The recommender system recommends items for a user according to the ranking of items. Specifically, the ranking algorithm sorts candidate items according to users' preference, and generates a ranking list. Top-scoring candidates receive the most exposure and are ranked first. The top-k candidates are usually returned. A learnable recommender system usually uses a loss function for ranking. Depending on the flaw in the design, some biases, such as popular bias and exposure bias, can be amplified in losses with these flaws, which may lead to unfair outcomes. In this survey, *ranking optimization for fairness* indicates methods that reduce unfairness in recommender systems by employing unbiased loss functions. Fig. 8 shows a general framework for ranking-based fairness-aware methods.

There are two types of ranking losses commonly used in a learnable recommender system: point-wise loss and pair-wise loss. A point-wise loss, such as binary cross-entropy (BCE) [137] and mean squared error (MSE) [101], minimizes the difference between the calculated recommendation score and the ground truth to capture user preferences for individual items. Pair-wise loss, such as Bayesian Personalized Ranking (BPR) [138], maximizes the user preference gap between the observed items and the unobserved items. The observed items are interacted with users and the unobserved items have no interaction with users. The BPR loss is a pair-wise ranking algorithm, that performs pairwise comparisons of items, and learns the order of relevant pairs from the comparisons to rank. The BPR loss encourages the observed item's

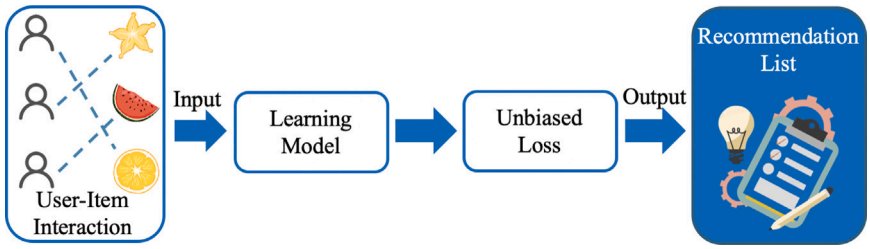


Fig. 8. Ranking-based fairness-aware methods. Such methods use a carefully designed unbiased loss for ranking.

prediction to be higher than the unobserved item's prediction. The BPR loss can be summarized as:

$$\mathcal{L}_{BPR} = -\sum_{(u,v_1,v_2)\in D_S} \ln \sigma(y_{u,v_1} - y_{u,v_2}), \tag{13}$$

where $y_{u,v}$ is the predicted score, $D_S = \{(u, v_1, v_2) \mid I_{u,v_1} = 1,\ I_{u,v_2} = 0\}$, and $I_{u,v}$ is the interaction between the user u and the item v. $I_{u,v} = 1$ indicates v is an observed item, and vice versa.

It can be seen from the above equation that the BPR loss is affected by the observed item. Popular items tend to have higher observation than unpopular items, thus, they have a high exposure probability. The more items are observed by the user, the higher the exposure for the user, which leads to popularity bias and exposure bias. Flaws in the BPR loss cause the popularity bias to be amplified during training. Therefore, Wan et al. [63] propose a cross pairwise ranking loss L_{CPR} for unbiased training:

$$L_{CPR} = -\sum_{(u_1,u_2,v_1,v_2)\in D_c} ln\sigma[\frac{1}{2}(y_{u_1,v_1} + y_{u_2,v_2} - y_{u_1,v_2} - y_{u_2,v_1})], \tag{14}$$

where $D_c = \{(u_1, u_2, v_1, v_2) \mid I_{u_1,v_1} = 1, I_{u_2,v_2} = 1, I_{u_1,v_2} = 0, I_{u_2,v_1} = 0\}$ denotes the training data, and I is the user–item interaction set. The CPR loss uses cross pair-wise interactions as training samples. Given two users and their interacted items, the unobserved data is obtained by exchanging items for the user. CPR decomposes exposure probabilities of items into a user-specific, item-specific propensity and user–item relevance, which are not independent of each other and thus contain bias. These predicted scores expressed by exposure biases can cancel the exposure biases each other according to Eq. (14), therefore the CPR loss is unbiased. There are also ranking methods that analyze the flaws of BPR loss [64,81]. However, these methods achieve unbias ranking by adding regularization terms, so we will not go into details here.

3.2.6. Others

In addition to the above-mentioned typical methods, there are also some niche methods to enhance the fairness of recommender systems. Zhou et al. [113] theoretically demonstrate that contrastive loss can replace the inverse propensity score to reduce exposure bias. Shen et al. [114] learn cross-scenario user interests through an attention network, and propose a fairness factor to gauge how important each scenario is. Zheng et al. [53] design a context-bias-aware recommendation model. They use attention networks to infer negative user preferences and eliminate contextual bias caused by the combined interaction between multiple items. Li et al. [115] design an end-to-end model for time-influenced sequential recommendation by weighting the embeddings to mitigate the unfair distribution of user attributes over items.

3.3. Post-processing methods for fair recommender systems

From the view of fairness, the recommendation results from target systems are usually not optimal, since these results potentially do not take factors concerning fairness (e.g., interactions between items, manual intentions, and differences in user preferences) into consideration. For enhancing fairness, *post-processing* methods aim to rearrange the recommendation results provided by target models, which are treated as black-box during the rearrangement (i.e., re-ranking), after the training of recommender systems. Re-ranking methods include manual-based re-ranking and algorithmic-based re-ranking. Here, we mainly introduce the re-ranking method that helps to boost the recommender system fairness, without repeating the manual re-ranking. Depending on whether or not the manual intervention is required, current re-ranking methods can be categorized into manual re-ranking and algorithmic-based re-ranking methods. In this survey, we focus on the algorithmic-based re-ranking methods and divide them into *non-parametric re-ranking* and *parametric re-ranking*, according to if a parameter learning process is required to conduct re-ranking. Fig. 9 shows a general framework of re-ranking methods for fair recommender systems.

3.3.1. Non-parametric re-ranking

In this survey, *non-parametric re-ranking* methods represent learning-free algorithms to conduct re-ranking concerning fairness. Heuristic methods used to achieve fairness are generally non-parametric re-ranking strategies, which treat the re-ranking process as an integer programming problem. Under specified fairness constraints, the optimal re-ranking outcomes are then found through heuristic search methods. Specifically, given original top-k recommendation results for each user from recommender systems, heuristic methods are designed to maximize the total preference score with respect to fairness by rearranging original recommendations, which can be formulated as

$$\begin{aligned} &\arg\max_{I_{uv}} \sum_{u=1}^{n}\sum_{v=1}^{N} I_{uv}S_{uv}, \\ &s.t. C(g_1, g_2) < \xi, \\ &\quad I_{uv} \in [0, 1], \end{aligned} \tag{15}$$

where I_{uv} is an interaction matrix that reflects whether item v should be recommended to user u, $C(\cdot)$ represents a fairness constraint function, g_1 and g_2 are two different groups. The objective function is to maximize the recommendation score under fairness constraints, which is composed of user–item matching and user preference. By treating the optimization as a 0–1 integer programming problem, a series of works employ heuristic algorithms to conduct re-ranking. Here we present some representative methods.

To ensure fairness across various groups, Fu et al. [139] utilized a heuristic re-ranking algorithm that constructs bias-constrained explainable recommendations on knowledge graphs. Similarly to this, Li et al. [15] proposed a similar approach, employing a heuristic-based method for re-ranking users, aiming to enhance group fairness considering their attributes. According to Wu et al. [118], there is a model called TFROM, which is designed to enhance exposure fairness for both users and providers. In this model, the recommendation list's length is equated to the capacity of a knapsack, and the items symbolize the objects within this knapsack. TFROM uses a heuristic search algorithm to solve the knapsack problem. Zhu et al. [62] present a novel re-ranking framework including a score scaling model, which is used to re-rank the biased recommendation results. Naghiaei et al. [116] convert the recommendation optimization problem into a 0–1 integer programming

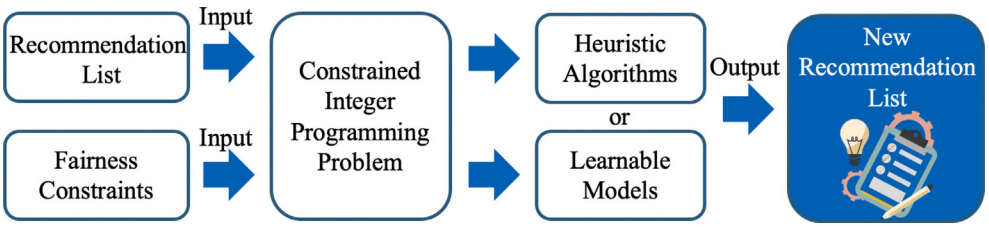


Fig. 9. The pipeline of re-ranking methods for fair recommender systems. These methods use recommendation lists and fairness constraints to construct integer programming problem, and use heuristic algorithms or learnable models to re-rank the recommendation lists.

and knapsack problem, and then propose employing greedy algorithms or solvers (e.g., Gurobi) to perform re-ranking. Moreover, Mansoury et al. [117] transform the item exposure inequity problem in popularity into a graph maximum flow problem to conduct re-ranking. Recent work [140] design a two-stage approach to transform the problem into a max–min problem, using greedy strategies to ensure maximum exposure for producers and sorting to ensure fairness for each user.

3.3.2. Parametric re-ranking

In this survey, all re-ranking methods that require a learning process to improve the fairness of recommender systems during the post-processing stage are called *parametric re-ranking* methods. Existing study [62] shows that, compared with non-parametric re-ranking methods, parametric re-ranking methods own better debiasing ability and are more competent when enhancing equality opportunity-based fairness. Here we introduce several typical parametric re-ranking methods. Given the recommendation results from the target systems, Zhu et al. [62] employ a learnable autoencoder module to enhance fairness and ensure the rearranged results have the same distribution as the original recommendations.

4. Datasets and evaluation for fair RSs

In this subsection, we provide a brief overview of the most commonly utilized datasets and assessment metrics in the context of fairness-aware recommender systems.

4.1. Datasets

In this section, we will present several datasets frequently employed in experiments related to the credibility of recommendation systems. Due to the complexity of fairness study in recommender systems, one specific dataset can be applied in addressing different bias issues. For example, the Amazon Review dataset is used in mitigating user attribute bias, model weights bias, and popularity bias. It is beneficial to use a variety of datasets to detect unfairness in recommender systems and evaluate fairness-enhancing methods. Here we summarize commonly used datasets in Table 2, demonstrating their basic information, and possible scenarios (e.g., exposure bias, and what kind of recommender systems can be used).

In addition to the datasets mentioned in Table 2, some other datasets can also be used to mitigate unfairness. For example, Gowalla [154], Ctrip [118] can be used to solve the Pairwise bias. CIKM and AliEc can be used to solve the user attribute biases [48]. MinD [155] and COCO [13] can be used to solve user behavior biases. CiteULike and XING [62] can be used to solve cold-start bias. Clothing, CellPhones [139] are used to solve Group fairness. Ciao [98] can be used to solve Feedback bias.

4.2. Evaluation

Evaluating a fairness-aware recommender system can be measured in terms of its accuracy and fairness. To avoid bias in recommendation results, this section presents accuracy metrics that will not harm fairness first, then metrics for assessing the fairness of recommender systems.

Accuracy Metrics for Fairness. There are some evaluation metrics used for evaluating the quality of a model. The common evaluation metrics are as follows:

- *Area Under Curve (AUC)* [156] is the region contained within the Receiver Operating Characteristic (ROC) curve and the coordinate axis. The maximum possible value of AUC is one. The closer the AUC is to 1, the higher the authenticity of the detection method; and vice versa.
- *Precision (P)* [157] is the proportion of correctly classified positive examples to all positive examples that are classified, and Precision@k (P@k) represents the Precision of the top-k items in the list.
- *Recall (R)* [158] refers to how many of the positive samples are successfully found by the model. Recall@k indicates the recall of the first k items in the result list.
- *Average Precision (AP)* [159] calculates the average precision. The higher the value of AP@K, the more relevant items are present in the top$-k$ recommendations and the higher these relevant items are ranked.
- *Normalize Discounted Cumulative Gain (NDCG@k)* [160] provides different degrees of relevance, ranking the results according to relevance and weighting them uniformly.

Metrics to Evaluate Fairness. There are some commonly used metrics for measuring the fairness of recommender systems:

- *Gini coefficient* is an economic indicator that is used to measure the income inequality of a country or region [28]. It is also often used to measure the fairness of recommender systems, which measures the value inequality of a frequency distribution [27]. The value of Gini coefficient is between 0 and 1, and smaller values represent greater fairness. Its definition is:

$$Gini\ coefficient = \frac{1}{2n^2\overline{ev}} \sum_{i=1}^{n} \sum_{j=1}^{n} |ev_i - ev_j|, \tag{16}$$

  where n indicates items' number, ev_i represents the exposure frequency of item v_i, and $\bar{ev}$ donates the average of items' exposure frequency.
- *KL-divergence* computes the divergence between two distributions [161], which can be used to measure various biases. Taking popularity bias as an example, KL-divergence can measure the distribution difference between user history records and popularity in recommendation results [20]. The recommendations are fairer when the value of KL-divergence is lower. The expression of KL-divergence is

$$D_{KL}(D_1(v), D_2(v)) = \sum_{v \in V} D_1(v) log \frac{D_1(v)}{D_2(v)}, \tag{17}$$

Table 2
Summary of the selected datasets. We present the dataset, detailing its cardinality—specifically the number of users, items, and edges, its applicability to fairness issues, and associated reference relations.

Datasets	#Users	#Items	#Edges	Fairness Issues	References
Amazon	3,915	2,549	77,328	Exposure Bias	[118]
				Training Bias	[141]
				User Bias	[64]
Beauty	22,363	12,101	198,502	Exposure Bias	[113]
				User Bias	[139], [15], [134]
Epinion	49,290	139,738	664,828	Exposure Bias	[116]
				Popularity Bias	[142]
				User Bias	[143]
Lastfm	1,892	17,632	92,834	Exposure Bias	[116], [21], [140], [26]
				Training Bias	[141]
				User Bias	[127]
MovieLens	943	1,349	99,287	Exposure Bias	[116], [26]
				Popularity Bias	[111], [99]
				User Bias	[104], [144]
MovieLens 1M	6,040	3,706	1,000,209	Cold-start Bias	[62], [79]
				Exposure Bias	[145], [146], [21]
				Time Bias	[27], [115]
				Training Bias	[141]
				Popularity Bias	[143], [147], [148]
				User Bias	[64], [131], [142], [149]
Yahoo!R3	5,400	1,000	183,179	Popularity Bias	[147], [150], [148]
				Training Bias	[141]
				User Bias	[151], [112], [152]
Yelp	25,677	25,815	731,671	Exposure Bias	[132]
				Popularity Bias	[98], [111]
				Training Bias	[141]
				User Bias	[64], [142], [153]

where $D_1(\cdot)$ refers to a distribution of item v and $D_2(\cdot)$ refers to a fair distribution of item v.

- *Difference* mainly considers that if the distance between the two recommendation results is less than the fairness constraint coefficient, then the recommendation system is considered to be fair [15], i.e.,

$$|y_{v_i} - y_{v_j}| < \xi, \tag{18}$$

where y_{v_i} is a recommender system prediction of item v_i, ξ is strictness parameter of fairness requirements. If the prediction score of v_i and v_j is smaller than the fairness constraint coefficient, we consider that the recommender system is fair.

In addition to the above metrics, we can also employ other metrics to evaluate fairness-aware recommender systems. For example, Lesota et al. [20] propose measures to account for prevalence bias from the median, various statistical moments, and measures of similarity that consider the entire prevalence distribution, including Mean, Median, Variance, Skew, Kurtosis, Kendall's τ rank-order correlation. If a recommender system offers suggestions that suit the preferences of a certain group of users, but fails to accurately reflect the preferences of another user group, it can be deemed unfair. An analytical metric called miscalibration is used by Abdollahpouri et al. [148] to measure the degree to which a recommender system responds to the true preferences of the user. If the Hellinger distance [148] between two distributions exceeds a threshold, then it is called miscalibration. According to Fu et al. [111], two measures are used to measure popularity bias in conversational recommender systems (CRSs): personalized average recommendation popularity (PARP) and popularity-rank correlation for users (PRU). Mena-Maldonado et al. [147] present two novel metrics for describing the relationship between prevalence and relevance distributions. The article experimentally identifies scenarios for the use of true-positive metrics or false-positive metrics. Ge et al. [27] propose a fairness indicator known as the popularity rate, which represents the proportion of popular items in a recommended list in relation to the total number of recommended items.

5. Industrial applications of fair RSs

This session focuses primarily on the application of fairness-aware recommender systems to real-world situations. We introduce the impact of fairness-aware recommender systems on life from application scenarios such as e-commerce and finance. Table 3 highlights real-world applications of fair recommender systems.

5.1. Fairness in E-commerce recommendations

There are many applications of the fair correction recommendation system in the area of e-commerce. To address the uneven data distribution across scenarios and the systematic bias of disadvantaged items that emerge in the re-ranking phase, Shen et al. [114] have introduced a solution to the aforementioned challenges in the form of a Scenario-Aware Ranking Network (SAR-Net). The SAR-Net is designed to capture a user's interests across various scenarios using two tailored attention modules. By assessing the importance of individual samples and adjusting predictions accordingly, the SAR-Net aims to mitigate data bias arising from human intervention. Wu et al. [162] introduce that in e-commerce websites such as Amazon and JD.com, a feeds-based recommendation has become a mainstream recommendation mode, and users can scroll down to view more products recommended by

Table 3
Industrial applications of fairness-aware recommender systems. We delineate the applications by outlining their respective domains of existence, the prevalent bias that they grapple with, and the methodological approaches they adopt to counteract these biases.

Domains	Sub-domains	Bias	Methods	Reference
E-commerce		Exposure Bias	Ranking	[114]
		User Bias	Re-ranking	[162]
Education		Exposure Bias	Re-ranking	[163], [164]
		Popularity Bias	Ranking	[165]
Social activities	Job recommendation	Exposure Bias	Counterfactual learning	[59]
		User Bias	Adversarial learning	[109]
	News recommendation	User Bias	Regularization	[155]
		User Bias	Adversarial learning	[53]
		Popularity Bias	Adversarial learning	[166]
	Streaming media recommendation	Popularity Bias	Ranking	[167]
		Time Bias	Backdoor adjustment	[168]
		Popularity Bias	Re-ranking	[123]
		User Bias	Adversarial learning	[169]

feeds. In the rolling feeds recommendation, four grids with pictures are usually played on the same mobile phone screen, and the similarity of the four products will affect the user's judgment, thus introducing bias. The article considers this phenomenon to be a contextual bias. The authors propose an unbiased counterfactual learning method to eliminate contextual bias. The proposed method is applied to a real-world e-commerce website, JD.com.

5.2. Fairness in education recommendations

Users are hoped to receive an education without bias in a fair education recommender system. Gómez et al. [163] believe that the geographic location of teachers has a strong impact on visibility and exposure. The re-ranking approach overcomes these phenomena by ensuring that each group receives the exposure expected, thereby ensuring that different providers are treated fairly. Boratto et al. [165] explore how recommender systems in the context of popularity-biased massively open online courses. A comparison is made of existing algorithms relating to the popularity of courses, catalog coverage, and popularity of course categories. Marras et al. [164] provide a formal definition of the online education recommendation principle and propose a novel re-ranking method that is conscious of fairness, in an effort to strike a balance between personalization and recommendation opportunities.

5.3. Fairness in social activities

5.3.1. Job recommendations

Job recommendation systems match job seekers with job information, and recommend job listings that meet job seekers' wishes, or lists of talented candidates that meet the requirements of the recruiter [170]. However, user-sensitive attributes may cause discrimination for users in job recommender systems and reduce users' trust in them. Several researches are devoted to mitigating discrimination and improving the fairness of job recommendation systems. Chen et al. [59] find that click-through rates for job advertisements decreased over time. Thus, they adopt the inverse propensity weighting method and customize a new loss function to rank the deviation of ad exposure position. In a recent paper, Rus et al. [109] demonstrate that gender bias can be removed from 12 million job openings and 0.9 million resumes through the use of a generative adversarial network, providing fair job recommendations to mitigate the pay gaps between different genders.

5.3.2. News recommendations

News recommendation systems mainly recommend news to users on digital news sites. Qi et al. [155] present ProFairRec, a news recommendation framework that prioritizes provider fairness. By integrating adversarial learning, the framework ensures that representations of fair news from providers remain unbiased during the recommendation process. They suggest the use of orthogonal regularization of provider-fair and biased representations to decrease the bias associated with news providers. Zheng et al. [53] argue that the contextual bias among news items may not be fully captured due to interactions among multiple items. To address this, they propose a novel context-bias-aware recommendation model aimed at eliminating context bias and achieving fairness in recommendations. Qi et al. [166] propose user encoders with popularity awareness to eliminate popularity bias from user behavior and achieve accurate interest modeling. In news recommendation, several recommender systems utilize multiple heads to capture correlations between news items based on representations from the news that users view. Yi et al. [171] argue that news click behavior may also be biased by the way news is presented on online platforms. So this paper proposes a bias-aware personalized news recommendation approach called DebiasRec. Debiasrec trains a biased news recommendation model from biased click behavior and inferring the biased interests of users from the clicked news articles.

5.3.3. Streaming media recommendations

Streaming media recommendation systems include music recommendation, video recommendation, etc. As a means of increasing transparency and fairness to artists in music recommendation systems, Kirdemir et al. [167] find the presence of video recommendations in YouTube's structural and systematic biases in YouTube. By employing a graphical probabilistic approach, this study evaluates the structural properties of video recommendations. To eliminate undesired temporal bias, Zhan et al. [168] propose a duration-fault quantization (D2Q)-based watch-time prediction framework that allows for industrial production systems for scaling. The framework has been implemented within the Kuaishou App, a commercial video streaming platform. This has resulted in a substantial enhancement in predicting real-time video viewing time, thereby significantly improving real-time video consumption. A study by Shakespeare et al. [172] examines whether state-of-the-art collaborative filtering algorithms exacerbate or ameliorate artist gender biases. This work designs two methods for determining why differences are attributed to changes in the distribution of inputs based on gender and user preferences. Melchiorre et al. [173] construct a dataset comprising information about the music consumption habits and personality traits of Twitter users. This

Table 4
Connections between Fair-aware RSs to Trustworthy RSs. We present various categories of fairness recommendation methodologies associated with trustworthiness properties, along with a succinct summary of their content.

Trustworthiness	Methods for Fair RSs	Ref.	Sketches
Explainability	Causal inference	[174]	Causal analysis on the relationship between users' past and future behaviors.
		[132]	An explainable weighting method is employed to rank counterfactual recommendation outcomes effectively.
		[175]	A counterfactual analysis and explanation are provided to bolster the effectiveness of explanations and promote fairness in the process.
		[139]	Presenting a fairness-constrained method that utilizes heuristic re-ranking to address the issue of unfairness recommendations based on knowledge graphs.
Robustness	Data modification	[19]	Constructing antidote datasets to improve the fairness and robustness of recommender systems.
	Causal inference	[176]	Utilizing an inverse propensity score helps eliminate polarity bias in group recommendations, ensuring a more robustness and fairness outcome.
	Regularization	[177]	Improving robustness and control fairness through L2 regularization loss.
Privacy	Causal inference	[79]	Training models on risk-free, user-approved privacy data.
	Others	[21]	Preventing the exposure of sensitive information within the learned embeddings.
Discrimination	Re-ranking	[13]	Utilizing re-ranking methods helps minimize discrimination, promoting fairness and equal representation in the results.
Diversity	Re-ranking	[26]	Transforming the issue into a maximum flow problem to improve the diversity.
	Regularization	[178]	Incorporating regularization terms can enhance fairness and diversity, ensuring a more balanced and inclusive outcome.

work analyzes the recommendation algorithms SLIM and EASE Mult-VAE. Their research results reveal notable differences in performance between user groups scoring high and low on certain personality traits. The recent work [123] proposes a fairness-conscious re-ranking framework for quantifying and mitigating algorithmic bias due to data bias. In an online A/B test of representative rankings of LinkedIn Talent Search or recommendations, the authors propose a strategy aimed at distributing ranking outcomes according to one or more safeguarded attributes, with the goal of achieving fairness principles like equal opportunity and population parity. This large-scale deployment of a framework that deploys LinkedIn Recruiter to ensure fairness in the recruitment space without impacting business metrics has the potential to positively impact over 630 million LinkedIn members. Wu et al. [169] use adversarial learning to reduce bias arising from user-sensitive attributes. Furthermore, they utilize KL divergence to capture less candidate-aware bias.

6. Connections with other trustworthy dimensions

Recommendation systems are one of the most crucial parts of the lives of people today. However, some recommendations lack moral basis and restraints, which undermines user confidence and possibly transgresses the law. A crisis of trust in the recommendation system can be sparked by a significant volume of biased training data or biased recommendation algorithms. Thus, it is uttermost important to have trustworthy recommender systems. Ideally, a trustworthy recommender system should be open and transparent, and its methods for obtaining results should be explainable. The trustworthiness of a recommender system is mainly evaluated based on four ethical principles, including explainability, robustness, privacy, and fairness. Here we introduce other ethical principles beyond fairness.

- **Explainability.** Explainability requires that the decisions of the recommender system should be understandable by people. Specifically, it requires that the decision-making process, and input and output relationships of recommender systems should be logically explained. However, most of the current recommendation models operate in the form of "black boxes", sometimes it is not always possible to explain why a recommender system produces a particular output or decision, which can lead to users' distrust of recommendation models. Therefore, explainability is crucial for building user trust in recommender systems.
- **Robustness.** The robustness of recommender systems refers to their ability to continue to operate normally when threatened or attacked. It requires the recommender system to be safe, reliable, and robust enough to handle errors or inconsistencies in all life cycle stages of the recommender system. Recommendation systems generally rely on user history records to build algorithm models. User history records contain a lot of junk data generated by systems and humans, which may lead to a decrease in the fairness and accuracy of the model. Thus, a trustworthy recommendation algorithm must be robust.
- **Privacy.** Personal data collected by a recommender system should be safe and able to protect personal privacy. Private data and privacy must be protected throughout the life cycle of the recommender system. Private data encompasses both the information shared by the user and the information derived from the user's interactions with the system. A trustworthy recommendation system should take responsibility for preventing unlawful and unfair discrimination against users as a result of the collected data.

A recommender system may contain a variety of untrustworthy issues, and biases that cause unfairness may also cause other untrustworthiness. For example, privacy and robustness issues caused by data bias (age and gender). The issue of fairness may arise simultaneously with other trustworthy issues in a recommender system. We are concerned with the fairness of the recommender system and introduce other trustworthy ethical principles based on the fairness content. We describe the connection between fairness and each trustworthy property below. Table 4 outlines the connections between fairness and other trustworthiness properties.

6.1. Connections with explainability

Some methods for mitigating fairness issues can also be added to the explainability of models. For example, methods based on causal inference analyze the causes of bias and provide explanations for the decision-making process of recommendation models. Xu et al. [174] argue that the explainability of recommender systems involves causal

Table 5
Future directions for fairness-aware recommender systems. We categorize future directions by segmenting them into distinct trajectories, elucidating the current deficiencies inherent in the fairness-aware RSs, the potential challenges to be encountered, and elucidating the advantages of resolving these issues.

Directions	Current shortcomings	Future challenges	Challenges-solving benefits
Concepts	Various fairness concepts exhibit both distinctions and interconnected characteristics.	Create customized fairness concepts for diverse recommendation scenarios.	Standardize industry concepts while promoting fairness in diverse scenarios.
Frameworks	There is not a one-size-fits-all framework for addressing fairness concerns.	Apply suitable fairness methods to specific fairness issues.	Ensuring the most effective solutions are applied.
Trade-off	Balancing fairness can sometimes affect accuracy, causing imperfect outcomes.	Finding a balance between fairness and recommendation performance.	Ensuring optimal results while promoting equitable treatment for all users.
Trustworthiness	Fairness-trustworthiness interactions are underexplored in current researches.	Explore the interaction rules between fairness and trustworthiness.	Foster trustworthy properties in fairness RSs to increase trustworthiness.

analysis between the previous and future behaviors of users, which is bound to answer counterfactual questions. An example of the question can be "What would happen if a different set of items were purchased". Counterfactual inference provides a fair framework for recommender systems, where the constructed counterfactual world explains why the model makes the output decisions. Therefore, counterfactual reasoning can simultaneously promote the explainability and fairness of recommender systems. Ge et al. [132] propose a counterfactual explainable fairness framework for group fairness. Specifically, they propose an explainable weighting method to rank the counterfactual recommendation results, which can be seen as an explanation for the final recommendation. Cornacchia et al. [175] propose a model that fuses natural language processing and counterfactual inferencing to provide recommendations for the loans domain. This model provides users with fairness and transparent advice. The path-based method is a common method to make improvements to recommender systems' explainability. Fu et al. [139] make improvements on user–item path distribution and fairness of the recommender system by designing a fairness-aware ranking algorithm.

6.2. Connections with robustness

Robustness can be reflected in the ability to defense attacks on data and models. Data bias cause unfair, it is also vulnerable to attack, which makes the model less robust. Fang et al. [19] propose to construct antidote data that mimics the rating behavior of users to mitigate data bias. These data are not considered anomalous data for attacking, thus it can make an improvement on the model's robustness. The same method also be adopted by Rastegarpanah et al. [87]. Dokoupil et al. [176] argue that recommender systems should utilize as much unbiased data as possible, whereas real-world training data is biased. In this case, to make the recommender system robust, they use an inverse propensity score to remove polarity bias in group recommendations. As a result, the fairness of group recommendations has been improved. Zhu et al. [177] design a local model and a global model to improve robustness and control fairness through l2 regularization loss. They improve model robustness and fairness through continuous gradient optimization.

6.3. Connections with privacy

Training data in recommender systems may contain some user-sensitive attributes that may be considered private by users. Even if the user's sensitive data is well protected, privacy leakage may occur during the interaction with the recommender system. The leaked private information can be maliciously obtained by other users, which brings data bias to the recommendation system and causes unfairness. Some works [51,79] establish a connection between fairness and privacy. Recommender systems may be unfair if users' private information is used extensively for personalization and when protected private data attributes like gender and ethnicity are misused. A bipartite graph is typically created organically by users and items in recommender systems. Directly abusing some user–item representations will nevertheless cause the leakage of user-sensitive information, even if user–item interactions do not contain any user-sensitive data. This is because user behavior and attributes are found to be correlated in social theory. For example, a person's privacy (such as his gender) can be inferred from his actions. Each user's embedding has a hidden connection to the behavior of similar users and users who have the same items, in addition to being tied to the user's behavior.

Similarly, Wu et al. [21] make an effort to keep user privacy and sensitive information hidden from the recommender system. They transform the fairness-aware recommendation problem into learning fair user and item representations, and provide a GNN method (FairGo) to avoid any sensitive information from being revealed from the learned embeddings. To make a fair recommendation and prevent the spread of high-level sensitive information, FairGo designs an ego network, which is user-centric and links the purchased products of the user and the item. Fairgo designs an aggregation algorithm that prevents high-order information propagation in the ego network and achieves representation fairness. The ego network embedding and user–item embedding are mapped into the same space after learning the embedding. In this space, filters of sensitive information are used for filtering, and finally, fairness training is performed through graph adversarial learning. In a cold-start situation, where user–item interaction data is lacking, recommender systems leverage the data trained from non-cold-start scenarios as the proxy for cold-start user–item information. However, this approach can leak the privacy of non-cold-started user–items. Wei et al. [79] suggest training risk-free, user-approved private data, and then making privacy-preserving fair recommendations to cold-start users.

6.4. Connections with others

Discrimination occurs in an untrustworthy recommender system. Mansoury et al. [179] propose three different user profile features, and analyze the possible connection between these features and the different behaviors of the recommender system for different genders. They introduce the unfairness of the recommender system caused by gender discrimination, and find that women get less accurate recommendations than males based on their experiments. This phenomenon indicates that the recommendation algorithm is unfair to different genders.

An ethical recommendation system cannot discriminate against vulnerable groups. In addition to protecting user privacy from the standpoint of ethical and moral norms, we also need to consider the requirements of relatively underprivileged groups. A number of measures are suggested by Leonhardt et al. [180] as ways to quantify the influence

Table 6
A comprehensive overview of recommender systems that are both fairness-aware and incorporate trustworthiness features. The horizontal axis represents the recommendation methods promoting fairness, while the vertical axis corresponds to the trustworthiness features. The **blank** areas indicate an absence of related research in those particular domains. These unexplored areas also present opportunities for future research directions.

	Pre-processing methods	Regularization	Causal inference	Adversarial learning	Reinforcement learning	Ranking	Post-processing methods
Explainability				[174], [132]		[139]	
Robustness	[19], [87]	[177]		[176]			
Privacy			[21]	[79]			[51]
Others		[178]					[13], [26]

of fairness-aware pre-processing techniques on user prejudice. A re-ranking method is developed by Gomez et al. [13] in order to reduce bias caused by the discrimination against teachers' geographic location. By using black feminist and critical race theory, Schelenz [181] attempts to lessen the unfairness of the user's political and social environment.

In addition to the problem of discrimination, contemporary recommender systems use big data to conduct in-depth and detailed mining of users' historical behaviors, personal characteristics, and other data. They take the learned user's preference as the main standard and provide precise "Act According to Actual Circumstances" recommendations for the user. While these recommender systems excessively collect users' private data, they limit the possibility of ordinary people exploring a variety of new fields.

Another form of fairness that some studies suggest is diversity [181]. Diversity can alleviate the ethical issues caused by the precise recommendation of recommender systems. Mansoury et al. [26] solve the problem of unfairness by transforming it into a maximum flow problem, which improves the overall diversity and fair distribution of recommended items. Sacharidis et al. [178] propose a regularization for social recommendations that allows friends to be similar. However, within a community, it generally forces members to be more diverse, which results in fairer recommendations.

7. Future directions

In recent years, attention from the academy and industry communities has been paid to improving fair recommender systems. However, thoroughly building fair recommender systems that can be trusted still faces the following challenges. Table 5 outlines current works' shortcomings, unresolved issues, and potential benefits of resolving them.

Concepts for fairness. In different recommendation scenarios, the fairness goals that people pursue are also different. For example, in a recommender system with multiple stakeholders, the goal of fairness is to balance the interests of multiple stakeholders. However, the goal of fairness in a time-aware job recommender system is to balance exposure frequency with old and new job information [59]. The concepts of fairness in different scenarios have both mutually inclusive and different parts. For example, both long-term fairness and dynamic fairness are fairness affected by time. A statically fair recommender system may include individual fairness and group fairness. Therefore, it is difficult to have a unified and accurate concept to define the fairness of recommender systems. Obtaining a common concept for different definitions of fairness is an important challenge. In addition, there may be some different fairness issues in a recommendation scenario, and these fairness issues may be conflicting. A potentially promising approach is to consider the prioritization of fairness issues in a recommender system. Few works consider the importance of fairness issues. This will also be an important challenge in the future.

General frameworks for fairness. The application scenarios of fairness-aware recommender systems are broad, including education, society, health care, et al. However, each recommendation scenario has different fairness-aware recommendation methods, such as regularization-based methods and re-ranking methods. At present, there is no work to analyze which type of fairness method is applicable to a certain type of fairness issue from the perspective of recommendation scenarios. In addition, due to the diversity of fairness issues, building a unified recommendation framework to solve all fairness issues can simplify the analysis of various fairness issues and also can be quickly applied to new scenarios and unknown fairness issues. There is no work yet to solve the above problem. We look forward to building a general recommendation framework to address different fairness issues in the future.

Trade-off between fairness and performance of RSs. As a means of ensuring the fairness of the recommender system, a system needs to reduce the bias in the recommendation output. However, it is challenging to combine fairness with accuracy in recommender systems, mainly because the goals of fairness and accuracy are inconsistent and the trade-off between them is substantial. In recommender systems, accuracy is determined by the system's capacity to accurately anticipate and meet the needs and interests of users. Taking the example of a multi-stakeholder recommender system (such as suppliers and consumers), users want the recommendation system to recommend products that meet their preferences, and suppliers want to make the items they provide as fair as possible to recommend to users. If the fairness of item exposure is maintained, it may lead to a decrease in the accuracy of recommendations. Therefore, controlling the trade-off between accuracy and fairness becomes critical.

Mutual promotion with trustworthy properties. At present, building a trustworthy recommendation system is the development trend of artificial intelligence. Credibility includes multiple properties, and fairness is one of them. Most of the current work focuses on the association between fairness and trustworthiness, and does not study the law of mutual influence between these properties. For example, the work of Fu et al. [139] introduces achieving fairness on an explainable recommender system. This type of work mainly addresses the issue of fairness without improving explainability. In other words, these approaches to fairness do not promote explainability. We hope that future works can focus on the intrinsic interaction between fairness and other trustworthy properties in recommender systems. In addition to explainability, there is also little work that considers the positive and negative effects between robustness and fairness.

Most of the current fairness work coexists with only one fairness property, and no work comprehensively considers all fairness properties. Another challenge in the future is how to integrate fairness and other credible properties in a recommender system, establish internal correlations between fairness and properties, reduce conflicts between properties, and build a credible recommender system. Table 6 shows the current state of research on integrating fairness methods and trustworthiness properties. It is evident that numerous areas have yet to be explored, such as the application of reinforcement learning methodologies to develop explainability and fairness recommender systems. There remains significant research potential in exploring the integration of trustworthiness and fairness.

8. Conclusion

Recommender systems provide basic artificial intelligence services to facilitate our daily lives. To ensure that fairness-aware recommender systems are crucial for people to enjoy trustworthy recommendations, this survey reviews current efforts on fair recommender systems with the aim of facilitating their implementation and future research. Firstly, we establish a foundation by introducing fundamental concepts related to recommender systems and fairness. This introduction will later aid us in defining fairness in various contexts. Then, we introduce different types of biases that cause unfairness at each stage of the recommender system's lifecycle. Finally, we introduce the manifestations of fairness in different recommendation scenarios, which paves the way for the introduction of subsequent methods. For fairness-aware recommendation methods, we introduce three processing stages: pre-processing, in-processing, and post-processing. We summarize the categories of methods and divide them in detail in each processing stage. For the evaluation of the fairness-aware recommender system, we introduce the data sets suitable for different recommendation scenarios and the fairness measurement metrics. For the application of fairness-aware recommender systems in the real world, we describe how fairness-aware recommender systems maintain fairness in areas closely related to people's lives, such as e-commerce, education, and et al. Regarding the promotion of human trust in fairness-aware recommender systems, we start with several trustworthy properties and introduce the correlation between fairness and these trustworthy properties. We conclude this survey and discuss the future directions of fairness-aware recommender systems from several novel perspectives, including the development direction of fair recommender systems, as well as the development of fair and trustworthy recommender systems.

CRediT authorship contribution statement

Di Jin: Conceptualization, Project administration, Writing – original draft. **Luzhi Wang:** Writing – original draft. **He Zhang:** Writing – original draft. **Yizhen Zheng:** Writing – original draft. **Weiping Ding:** Writing – review & editing. **Feng Xia:** Writing – review & editing. **Shirui Pan:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Di Jin reports financial support was provided by National Science Foundation. Shirui Pan reports financial support was provided by Australian Research Council.

Data availability

No data was used for the research described in the article.

References

[1] F. Ricci, L. Rokach, B. Shapira, Recommender systems: Techniques, applications, and challenges, in: Recommender Systems Handbook, 2022, pp. 1–35.

[2] B. Kersbergen, S. Schelter, Learnings from a retail recommendation system on billions of interactions at bol.com, in: 2021 IEEE 37th International Conference on Data Engineering, ICDE, IEEE, 2021, pp. 2447–2452.

[3] Y. Gu, W. Bao, D. Ou, X. Li, B. Cui, B. Ma, H. Huang, Q. Liu, X. Zeng, Self-supervised learning on users' spontaneous behaviors for multi-scenario ranking in e-commerce, in: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, 2021, pp. 3828–3837.

[4] X. Guo, S. Wang, H. Zhao, S. Diao, J. Chen, Z. Ding, Z. He, J. Lu, Y. Xiao, B. Long, et al., Intelligent online selling point extraction for e-commerce recommendation, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2022, pp. 12360–12368.

[5] J. Yang, J. Li, S. Liu, A novel technique applied to the economic investigation of recommender system, Multimedia Tools Appl. 77 (2018) 4237–4252.

[6] D. Toquica, K. Agbossou, R. Malhamé, N. Henao, S. Kelouwani, M. Fournier, A recommender system for predictive control of heating systems in economic demand response programs, IEEE Open J. Ind. Appl. 3 (2022) 79–89.

[7] M. Bogaert, J. Lootens, D. Van den Poel, M. Ballings, Evaluating multi-label classifiers and recommender systems in the financial service sector, European J. Oper. Res. 279 (2019) 620–634.

[8] T. Hassan, B. Edmison, T. Stelter, D.S. McCrickard, Learning to trust: Understanding editorial authority and trust in recommender systems for education, in: Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, 2021, pp. 24–32.

[9] A.C. Chuang, N.F. Huang, J.W. Tzeng, C.A. Lee, Y.X. Huang, H.H. Huang, Moocers: Exercise recommender system in moocs based on reinforcement learning algorithm, in: 2021 8th International Conference on Soft Computing & Machine Intelligence, ISCMI, IEEE, 2021, pp. 186–190.

[10] A. Agarwal, D.S. Mishra, S.V. Kolekar, Knowledge-based recommendation system using semantic web rules based on learning styles for moocs, Cogent Eng. 9 (2022) 2022568.

[11] Z. Qin, S.J. Chen, D. Metzler, Y. Noh, J. Qin, X. Wang, Attribute-based propensity for unbiased learning in recommender systems: Algorithm and case studies, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020, pp. 2359–2367.

[12] C. Musto, P. Lops, G. Semeraro, Fairness and popularity bias in recommender systems: an empirical evaluation, 2021.

[13] E. Gómez, C. Shui Zhang, L. Boratto, M. Salamó, The winner takes it all: Geographic imbalance and provider (un)fairness in educational recommender systems, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 1808–1812.

[14] Y. Wang, W. Ma, M. Zhang*, Y. Liu, S. Ma, A survey on the fairness of recommender systems, ACM Trans. Inf. Syst. (2022) http://dx.doi.org/10.1145/3547333.

[15] Y. Li, H. Chen, Z. Fu, Y. Ge, Y. Zhang, User-oriented fairness in recommendation, in: WWW, ACM / IW3C2, 2021, pp. 624–632.

[16] H. Liu, Y. Wang, W. Fan, X. Liu, Y. Li, S. Jain, A.K. Jain, J. Tang, Trustworthy AI: A computational perspective, 2021, CoRR abs/2107.06641.

[17] H. Zhang, B. Wu, X. Yuan, S. Pan, H. Tong, J. Pei, Trustworthy graph neural networks: Aspects, methods and trends, 2022, arXiv preprint arXiv:2205.07424.

[18] W. Fan, X. Zhao, X. Chen, J. Su, J. Gao, L. Wang, Q. Liu, Y. Wang, H. Xu, L. Chen, Q. Li, A comprehensive survey on trustworthy recommender systems, 2022, CoRR abs/2209.10117.

[19] M. Fang, J. Liu, M. Momma, Y. Sun, Fairroad: Achieving fairness for recommender systems with optimized antidote data, in: SACMAT, ACM, 2022, pp. 173–184.

[20] O. Lesota, A. Melchiorre, N. Rekabsaz, S. Brandl, D. Kowald, E. Lex, M. Schedl, Analyzing item popularity bias of music recommender systems: Are different genders equally affected? in: Fifteenth ACM Conference on Recommender Systems, 2021, pp. 601–606.

[21] L. Wu, L. Chen, P. Shao, R. Hong, X. Wang, M. Wang, Learning fair representations for recommendation: A graph-based perspective, in: Proceedings of the Web Conference 2021, 2021, pp. 2198–2208.

[22] E. Holmes, Anti-discrimination rights without equality, Mod. Law Rev. 68 (2005) 175–194.

[23] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, A. Galstyan, A survey on bias and fairness in machine learning, ACM Comput. Surv. 54 (2021) 115:1–115:35.

[24] A.J. Biega, K.P. Gummadi, G. Weikum, Equity of attention: Amortizing individual fairness in rankings, in: SIGIR, ACM, 2018, pp. 405–414.

[25] D. Pedreschi, S. Ruggieri, F. Turini, Measuring discrimination in socially-sensitive decision records, in: SDM, SIAM, 2009, pp. 581–592.

[26] M. Mansoury, H. Abdollahpouri, M. Pechenizkiy, B. Mobasher, R. Burke, A graph-based approach for mitigating multi-sided exposure bias in recommender systems, ACM Trans. Inf. Syst. 40 (2022) 32:1–32:31.

[27] Y. Ge, S. Liu, R. Gao, Y. Xian, Y. Li, X. Zhao, C. Pei, F. Sun, J. Ge, W. Ou, et al., Towards long-term fairness in recommendation, in: Proceedings of the 14th ACM International Conference on Web Search and Data Mining, 2021, pp. 445–453.

[28] R. Dorfman, A formula for the gini coefficient, Rev. Econ. Stat. (1979) 146–149.

[29] H. Zhang, X. Yuan, Q.V.H. Nguyen, S. Pan, On the interaction between node fairness and edge privacy in graph neural networks, 2023, CoRR abs/2301.12951.

[30] S. Wu, F. Sun, W. Zhang, X. Xie, B. Cui, Graph neural networks in recommender systems: a survey, in: ACM Computing Surveys, CSUR, 2022, http://dx.doi.org/10.1145/3535101.

[31] Y. Deldjoo, T.D. Noia, F.A. Merra, A survey on adversarial recommender systems: From attack/defense strategies to generative adversarial networks, ACM Comput. Surv. 54 (2021) 35:1–35:38.

[32] L. Wu, X. He, X. Wang, K. Zhang, M. Wang, A survey on accuracy-oriented neural recommendation: From collaborative filtering to information-rich recommendation, IEEE Trans. Knowl. Data Eng. (2022).

[33] M. Zehlike, K. Yang, J. Stoyanovich, Fairness in ranking, part ii: Learning-to-rank and recommender systems, in: ACM Computing Surveys, CSUR, 2022.

[34] J. Chen, H. Dong, X. Wang, F. Feng, M. Wang, X. He, Bias and debias in recommender system: A survey and future directions, 2020, CoRR abs/2010.03240.

[35] Y. Ge, S. Liu, Z. Fu, J. Tan, Z. Li, S. Xu, Y. Li, Y. Xian, Y. Zhang, A survey on trustworthy recommender systems, 2022, CoRR abs/2207.12515.

[36] S. Wang, X. Zhang, Y. Wang, H. Liu, F. Ricci, Trustworthy recommender systems, 2022, CoRR abs/2208.06265.

[37] M.N. Freire, L.N. de Castro, E-recruitment recommender systems: a systematic review, Knowl. Inf. Syst. 63 (2021) 1–20.

[38] J. Lu, D. Wu, M. Mao, W. Wang, G. Zhang, Recommender system application developments: A survey, Decis. Support Syst. 74 (2015) 12–32.

[39] D. Jin, L. Wang, Y. Zheng, G. Song, F. Jiang, X. Li, W. Lin, S. Pan, Dual intent enhanced graph neural network for session-based new item recommendation, in: Proceedings of the ACM Web Conference 2023, WWW 2023, Austin, TX, USA, 30 April 2023-4 May 2023, 2023, pp. 684–693, http://dx.doi.org/10.1145/3543507.3583526.

[40] C. Gao, W. Lei, X. He, M.de. Rijke, T. Chua, Advances and challenges in conversational recommender systems: A survey, AI Open 2 (2021) 100–126.

[41] K. Lamertz, The social construction of fairness: Social influence and sense making in organizations, J. Organ. Behav. 23 (2002) 19–37.

[42] J. Konow, A positive theory of economic fairness, J. Econ. Behav. Organ. 31 (1996) 13–35.

[43] F.I. Michelman, Property, utility, and fairness: comments on the ethical foundations of just compensation law, in: Constitutional Protection of Private Property and Freedom of Contract, Routledge, 2013, pp. 117–210.

[44] S. Salloum, D. Rajamanthri, Implementation and evaluation of movie recommender systems using collaborative filtering, J. Adv. Inf. Technol. 12 (2021).

[45] D. Lian, Y. Wu, Y. Ge, X. Xie, E. Chen, Geography-aware sequential location recommendation, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020, pp. 2009–2019.

[46] A. Ferraro, X. Serra, C. Bauer, Break the loop: Gender imbalance in music recommenders, in: Proceedings of the 2021 Conference on Human Information Interaction and Retrieval, 2021, pp. 249–254.

[47] A. Giabelli, L. Malandri, F. Mercorio, M. Mezzanzanica, A. Seveso, Skills2job: A recommender system that encodes job offer embeddings on graph databases, Appl. Soft Comput. 101 (2021) 107049.

[48] Y. Wu, R. Xie, Y. Zhu, F. Zhuang, A. Xiang, X. Zhang, L. Lin, Q. He, Selective fairness in recommendation via prompts, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 2657–2662.

[49] A. Datta, M.C. Tschantz, A. Datta, Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination, 2014, CoRR abs/1408.6491.

[50] B. Imana, A. Korolova, J.S. Heidemann, Auditing for discrimination in algorithms delivering job ads, in: WWW, ACM / IW3C2, 2021, pp. 3767–3778.

[51] V. Do, S. Corbett-Davies, J. Atif, N. Usunier, Online certification of preference-based fairness for personalized recommender systems, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2022, pp. 6532–6540.

[52] Y. Zhao, W. Ma, Y. Jiang, J. Zhan, A moocs recommender system based on user's knowledge background, in: KSEM, Springer, 2021, pp. 140–153.

[53] Z. Zheng, Z. Qiu, T. Xu, X. Wu, X. Zhao, E. Chen, H. Xiong, Cbr: Context bias aware recommendation for debiasing user modeling and click prediction, in: Proceedings of the ACM Web Conference 2022, 2022, pp. 2268–2276.

[54] G. Jawaheer, M. Szomszor, P. Kostkova, Comparison of implicit and explicit feedback from an online music recommendation service, in: HetRec@RecSys, ACM, 2010, pp. 47–51.

[55] F. Liang, W. Pan, Z. Ming, Fedrec++: Lossless federated recommendation with explicit feedback, in: AAAI, AAAI Press, 2021, pp. 4224–4231.

[56] X. Zhang, S. Dai, J. Xu, Z. Dong, Q. Dai, J.R. Wen, Counteracting user attention bias in music streaming recommendation via reward modification, in: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022, pp. 2504–2514.

[57] S. Gupta, H. Wang, Z.C. Lipton, Y. Wang, Correcting exposure bias for link recommendation, in: ICML, PMLR, 2021, pp. 3953–3963.

[58] M. Cheng, F. Yuan, Q. Liu, S. Ge, Z. Li, R. Yu, D. Lian, S. Yuan, E. Chen, Learning recommender systems with implicit feedback via soft target enhancement, in: SIGIR, ACM., 2021, pp. 575–584.

[59] R.C. Chen, Q. Ai, G. Jayasinghe, W.B. Croft, Correcting for recency bias in job recommendation, 2019, pp. 2185–2188.

[60] D. Cohen, M. Aharon, Y. Koren, O. Somekh, R. Nissim, Expediting exploration by attribute-to-feature mapping for cold-start recommendations, in: RecSys, ACM, 2017, pp. 184–192.

[61] M. Vartak, A. Thiagarajan, C. Miranda, J. Bratman, H. Larochelle, A meta-learning perspective on cold-start recommendations for items, in: Advances in Neural Information Processing Systems, Vol. 30, 2017.

[62] Z. Zhu, J. Kim, T. Nguyen, A. Fenton, J. Caverlee, Fairness among new items in cold start recommender systems, in: SIGIR, ACM, 2021, pp. 767–776.

[63] Q. Wan, X. He, X. Wang, J. Wu, W. Guo, R. Tang, Cross pairwise ranking for unbiased item recommendation, in: Proceedings of the ACM Web Conference 2022, 2022, pp. 2370–2378.

[64] Z. Zhu, J. Wang, J. Caverlee, Measuring and mitigating item under-recommendation bias in personalized ranking systems, in: SIGIR, ACM, 2020, pp. 449–458.

[65] H. Abdollahpouri, M. Mansoury, R. Burke, B. Mobasher, The unfairness of popularity bias in recommendation, in: RMSE@RecSys, CEUR-WS.org, 2019.

[66] M. Naghiaei, H.A. Rahmani, M. Dehghan, The unfairness of popularity bias in book recommendation, in: BIAS, Springer, 2022, pp. 69–81.

[67] M. Marras, L. Boratto, G. Ramos, G. Fenu, Equality of learning opportunity via individual fairness in personalized recommendations, Int. J. Artif. Intell. Educ. 32 (2022) 636–684.

[68] D. Zhang, J. Wang, Recommendation fairness: From static to dynamic, 2021, CoRR abs/2109.03150.

[69] Z. Wang, Q. Xu, Z. Yang, X. Cao, Q. Huang, Implicit feedbacks are not always favorable: Iterative relabeled one-class collaborative filtering against noisy interactions, in: ACM Multimedia, ACM, 2021, pp. 3070–3078.

[70] F. Kamiran, T. Calders, Data preprocessing techniques for classification without discrimination, Knowl. Inf. Syst. 33 (2011) 1–33.

[71] A. Estabrooks, T. Jo, N. Japkowicz, A multiple resampling method for learning from imbalanced data sets, Comput. Intell. 20 (2004) 18–36.

[72] M. Montanari, C. Bernardis, P. Cremonesi, On the impact of data sampling on hyper-parameter optimisation of recommendation algorithms, in: SAC, ACM, 2022, pp. 1399–1402.

[73] J. Ding, Y. Quan, X. He, Y. Li, D. Jin, Reinforced negative sampling for recommendation with exposure data, in: IJCAI, ijcai.org, 2019, pp. 2230–2236.

[74] L.E. Celis, A. Deshpande, T. Kathuria, N.K. Vishnoi, How to be fair and diverse?, 2016, arXiv preprint arXiv:1610.07183.

[75] N. Sachdeva, J.J. McAuley, How useful are reviews for recommendation? A critical review and potential improvements, in: SIGIR, ACM, 2020, pp. 1845–1848.

[76] J. Wang, Y. Yang, S. Wang, J. Hu, Q. Wang, Context- and fairness-aware in-process crowdworker recommendation, ACM Trans. Softw. Eng. Methodol. 31 (2022) 35:1–35:31.

[77] X. Luo, M. Zhou, Y. Xia, Q. Zhu, An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems, IEEE Trans. Ind. Inform. 10 (2014) 1273–1284.

[78] Y. Yuan, X. Luo, M. Shang, Effects of preprocessing and training biases in latent factor models for recommender systems, Neurocomputing 275 (2018) 2019–2030.

[79] T. Wei, J. He, Comprehensive fair meta-learned recommender system, in: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022, pp. 1989–1999.

[80] F. Prost, H. Qian, Q. Chen, E.H. Chi, J. Chen, A. Beutel, Toward a better trade-off between performance and fairness with kernel-based distribution matching, 2019, CoRR abs/1910.11779.

[81] A. Beutel, J. Chen, T. Doshi, H. Qian, L. Wei, Y. Wu, L. Heldt, Z. Zhao, L. Hong, E.H. Chi, C. Goodrow, Fairness in recommendation ranking through pairwise comparisons, in: KDD, ACM, 2019, pp. 2212–2220.

[82] M. Wang, X. Zheng, Y. Yang, K. Zhang, Collaborative filtering with social exposure: A modular approach to social recommendation, in: AAAI, AAAI Press, 2018, pp. 2516–2523.

[83] J. Wasilewski, N. Hurley, Incorporating diversity in a learning to rank recommender system, in: FLAIRS Conference, AAAI Press, 2016, pp. 572–578.

[84] D. Kiswanto, D. Nurjanah, R. Rismala, Fairness aware regularization on a learning-to-rank recommender system for controlling popularity bias in e-commerce domain, in: 2018 International Conference on Information Technology Systems and Innovation, ICITSI, IEEE, 2018, pp. 16–21.

[85] H. Abdollahpouri, R. Burke, B. Mobasher, Controlling popularity bias in learning-to-rank recommendation, in: RecSys, ACM, 2017, pp. 42–46.

[86] Y. Hu, Y. Koren, C. Volinsky, Collaborative filtering for implicit feedback datasets, in: ICDM, IEEE Computer Society, 2008, pp. 263–272.

[87] B. Rastegarpanah, K.P. Gummadi, M. Crovella, Fighting fire with fire: Using antidote data to improve polarization and fairness of recommender systems, in: WSDM, ACM, 2019, pp. 231–239.

[88] R. Burke, N. Sonboli, A. Ordonez-Gauger, Balanced neighborhoods for multi-sided fairness in recommendation, in: FAT, PMLR, 2018, pp. 202–214.

[89] B. Edizel, F. Bonchi, S. Hajian, A. Panisson, T. Tassa, Fairecsys: mitigating algorithmic bias in recommender systems, Int. J. Data Sci. Anal. 9 (2020) 197–213.

[90] R. Borges, K. Stefanidis, F2VAE: a framework for mitigating user unfairness in recommendation systems, in: SAC, ACM, 2022, pp. 1391–1398.

[91] C. Wu, F. Wu, X. Wang, Y. Huang, X. Xie, Fairness-aware news recommendation with decomposed adversarial learning, in: AAAI, AAAI Press, 2021, pp. 4462–4469.

[92] L. Boratto, G. Fenu, M. Marras, Interplay between upsampling and regularization for provider fairness in recommender systems, User Model. User Adapt. Interact. 31 (2021) 421–455.

[93] Z. Zhu, X. Hu, J. Caverlee, Fairness-aware tensor-based recommendation, in: CIKM, ACM, 2018, pp. 1153–1162.

[94] W. Sun, S. Khenissi, O. Nasraoui, P. Shafto, Debiasing the human-recommender system feedback loop in collaborative filtering, in: WWW (Companion Volume), ACM, 2019, pp. 645–651.

[95] Z. Qin, S.J. Chen, D. Metzler, Y. Noh, J. Qin, X. Wang, Attribute-based propensity for unbiased learning in recommender systems: Algorithm and case studies, 2020, pp. 2359–2367.

[96] M. Zhao, L. Wu, Y. Liang, L. Chen, J. Zhang, Q. Deng, K. Wang, X. Shen, T. Lv, R. Wu, Investigating accuracy-novelty performance for graph-based collaborative filtering, in: SIGIR, ACM, 2022, pp. 50–59.

[97] Y. Zhang, F. Feng, X. He, T. Wei, C. Song, G. Ling, Y. Zhang, Causal intervention for leveraging popularity bias in recommendation, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 11–20.

[98] T. Wei, F. Feng, J. Chen, Z. Wu, J. Yi, X. He, Model-agnostic counterfactual reasoning for eliminating popularity bias in recommender system, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021, pp. 1791–1800.

[99] J. Huang, H. Oosterhuis, M. de Rijke, It is different when items are older: Debiasing recommendations when selection bias and user preferences are dynamic, in: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, 2022, pp. 381–389.

[100] R. Acharyya, S. Das, A. Chattoraj, M.I. Tanveer, Fairyted: A fair rating predictor for ted talk data, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2020, pp. 338–345.

[101] X. Wang, R. Zhang, Y. Sun, J. Qi, Combating selection biases in recommender systems with a few unbiased ratings, in: WSDM, ACM, 2021, pp. 427–435.

[102] L. Xia, C. Huang, C. Zhang, Self-supervised hypergraph transformer for recommender systems, in: KDD, ACM, 2022, pp. 2100–2109.

[103] T. Schnabel, A. Swaminathan, A. Singh, N. Chandak, T. Joachims, Recommendations as treatments: Debiasing learning and evaluation, in: International Conference on Machine Learning, PMLR, 2016, pp. 1670–1679.

[104] Y. Li, H. Chen, S. Xu, Y. Ge, Y. Zhang, Towards personalized fairness based on causal notion, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 1054–1063.

[105] X. Qian, Y. Xu, F. Lv, S. Zhang, Z. Jiang, Q. Liu, X. Zeng, T. Chua, F. Wu, Intelligent request strategy design in recommender system, in: KDD, ACM, 2022, pp. 3772–3782.

[106] Z. Wang, Y. He, J. Liu, W. Zou, P.S. Yu, P. Cui, Invariant preference learning for general debiasing in recommendation, in: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022, pp. 1969–1978.

[107] S. Rajanala, A. Pal, M. Singh, R.C.W. Phan, K. Wong, Descover: Debiased semantic context prior for venue recommendation, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 2456–2461.

[108] H. Liu, N. Zhao, X. Zhang, H. Lin, L. Yang, B. Xu, Y. Lin, W. Fan, Dual constraints and adversarial learning for fair recommenders, Knowl.-Based Syst. 239 (2022) 108058.

[109] C. Rus, J. Luppes, H. Oosterhuis, G.H. Schoenmacker, Closing the gender wage gap: Adversarial fairness in job recommendation, 2022, CoRR abs/2209.09592.

[110] Y. Ge, X. Zhao, L. Yu, S. Paul, D. Hu, C.C. Hsieh, Y. Zhang, Toward pareto efficient fairness-utility trade-off in recommendation through reinforcement learning, in: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, 2022, pp. 316–324.

[111] Z. Fu, Y. Xian, S. Geng, G. de Melo, Y. Zhang, Popcorn: Human-in-the-loop popularity debiasing in conversational recommender systems, in: CIKM, 2021, pp. 494–503, http://dx.doi.org/10.1145/3459637.3482461.

[112] H. Liu, D. Tang, J. Yang, X. Zhao, H. Liu, J. Tang, Y. Cheng, Rating distribution calibration for selection bias mitigation in recommendations, in: Proceedings of the ACM Web Conference 2022, 2022, pp. 2048–2057.

[113] C. Zhou, J. Ma, J. Zhang, J. Zhou, H. Yang, Contrastive learning for debiased candidate generation in large-scale recommender systems, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021, pp. 3985–3995.

[114] Q. Shen, W. Tao, J. Zhang, H. Wen, Z. Chen, Q. Lu, Sar-net: A scenario-aware ranking network for personalized fair recommendation in hundreds of travel scenarios, 2021, CoRR abs/2110.06475.

[115] C.T. Li, C. Hsu, Y. Zhang, Fairsr: Fairness-aware sequential recommendation through multi-task learning with preference graph embeddings, in: ACM Transactions on Intelligent Systems and Technology, TIST, 2022, pp. 131–121.

[116] M. Naghiaei, H.A. Rahmani, Y. Deldjoo, Cpfair: Personalized consumer and producer fairness re-ranking for recommender systems, 2022, arXiv preprint arXiv:2204.08085.

[117] M. Mansoury, H. Abdollahpouri, M. Pechenizkiy, B. Mobasher, R. Burke, Fairmatch: A graph-based approach for improving aggregate diversity in recommender systems, in: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, 2020, pp. 154–162.

[118] Y. Wu, J. Cao, G. Xu, Y. Tan, TFROM: A two-sided fairness-aware recommendation model for both customers and providers, 2021, CoRR abs/2104.09024.

[119] C. Dickens, R. Singh, L. Getoor, Hyperfair: A soft approach to integrating fairness criteria, 2020, CoRR abs/2009.08952.

[120] R. Mehrotra, J. McInerney, H. Bouchard, M. Lalmas, F. Diaz, Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems, in: CIKM, ACM, 2018, pp. 2243–2251.

[121] A. Singh, T. Joachims, Fairness of exposure in rankings, KDD (2018) 2219–2228.

[122] K. Yang, J. Stoyanovich, Measuring fairness in ranked outputs, in: SSDBM, ACM, 2017, pp. 221–226.

[123] S.C. Geyik, S. Ambler, K. Kenthapadi, Fairness-aware ranking in search & recommendation systems with application to linkedin talent search, 2019, CoRR abs/1905.01989.

[124] M. Tavakol, Fair classification with counterfactual learning, in: SIGIR, ACM, 2020, pp. 2073–2076.

[125] N. Akpinar, C. DiCiccio, P. Nandy, K. Basu, Long-term dynamics of fairness intervention in connection recommender systems, in: AIES, ACM, 2022, pp. 22–35.

[126] M. Mladenov, E. Creager, K. Ben-Porat, R.S. Zemel, C. Boutilier, Optimizing long-term social welfare in recommender systems: A constrained matching approach, in: ICML, PMLR, 2020, pp. 6987–6998.

[127] V. Do, S. Corbett-Davies, J. Atif, N. Usunier, Online certification of preference-based fairness for personalized recommender systems, in: AAAI, AAAI Press, 2022, pp. 6532–6540.

[128] E.H. Simpson, The interpretation of interaction in contingency tables, J. R. Stat. Soc. Ser. B Stat. Methodol. 13 (1951) 238–241.

[129] T. Xiao, S. Wang, Towards unbiased and robust causal ranking for recommender systems, in: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, 2022, pp. 1158–1167.

[130] Z. Wang, S. Shen, Z. Wang, B. Chen, X. Chen, J.R. Wen, Unbiased sequential recommendation with latent confounders, in: Proceedings of the ACM Web Conference 2022, 2022, pp. 2195–2204.

[131] W. Wang, F. Feng, X. He, X. Wang, T. Chua, Deconfounded recommendation for alleviating bias amplification, 2021, CoRR abs/2105.10648.

[132] Y. Ge, J. Tan, Y. Zhu, Y. Xia, J. Luo, S. Liu, Z. Fu, S. Geng, Z. Li, Y. Zhang, Explainable fairness in recommendation, in: E. Amigó, P. Castells, J. Gonzalo, B. Carterette, J.S. Culpepper, G. Kazai (Eds.), SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11-15 2022, ACM, 2022, pp. 681–691, http://dx.doi.org/10.1145/3477495.3531973.

[133] H. Liu, Y. Wang, H. Lin, B. Xu, N. Zhao, Mitigating sensitive data exposure with adversarial learning for fairness recommendation systems, Neural Comput. Appl. (2022) 1–15.

[134] J. Li, Y. Ren, K. Deng, Fairgan: Gans-based fairness-aware learning for recommendations with implicit feedback, in: Proceedings of the ACM Web Conference 2022, 2022, pp. 297–307.

[135] Y. Li, H. Chen, S. Xu, Y. Ge, J. Tan, S. Liu, Y. Zhang, Fairness in recommendation: A survey, 2022, CoRR abs/2205.13619.

[136] W. Liu, F. Liu, R. Tang, B. Liao, G. Chen, P.A. Heng, Balancing accuracy and fairness for interactive recommendation with reinforcement learning, 2021, arXiv preprint arXiv:2106.13386.

[137] Z. Deng, L. Huang, C. Wang, J. Lai, P.S. Yu, Deepcf: A unified framework of representation learning and matching function learning in recommender system, in: AAAI, AAAI Press, 2019, pp. 61–68.

[138] S. Rendle, C. Freudenthaler, Z. Gantner, L. Schmidt-Thieme, BPR: bayesian personalized ranking from implicit feedback, in: UAI, AUAI Press, 2009, pp. 452–461.

[139] Z. Fu, Y. Xian, R. Gao, J. Zhao, Q. Huang, Y. Ge, S. Xu, S. Geng, C. Shah, Y. Zhang, et al., Fairness-aware explainable recommendation over knowledge graphs, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 69–78.

[140] G.K. Patro, A. Biswas, N. Ganguly, K.P. Gummadi, A. Chakraborty, Fairrec: Two-sided fairness for personalized recommendations in two-sided platforms, in: WWW, ACM / IW3C2, 2020, pp. 1194–1204.

[141] V.W. Anelli, T. Di Noia, F.A. Merra, The idiosyncratic effects of adversarial training on bias in personalized recommendation learning, in: Fifteenth ACM Conference on Recommender Systems, 2021, pp. 730–735.

[142] Z. Zhu, J. Caverlee, Fighting mainstream bias in recommender systems via local fine tuning, in: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, 2022, pp. 1497–1506.

[143] Z. Zhu, Y. He, X. Zhao, Y. Zhang, J. Wang, J. Caverlee, Popularity-opportunity bias in collaborative filtering, in: Proceedings of the 14th ACM International Conference on Web Search and Data Mining, 2021, pp. 85–93.

[144] R. Sato, Enumerating fair packages for group recommendations, in: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, 2022, pp. 870–878.

[145] H. Wu, B. Mitra, C. Ma, F. Diaz, X. Liu, Joint multisided exposure fairness for recommendation, 2022, arXiv preprint arXiv:2205.00048.

[146] C. Zhou, J. Ma, J. Zhang, J. Zhou, H. Yang, Contrastive learning for debiased candidate generation at scale, 2020, CoRR abs/2005.12964.
[147] E. Mena-Maldonado, P. Cañamares, Y. Ren, M. Sanderson, Popularity bias in false-positive metrics for recommender systems evaluation, ACM Trans. Inf. Syst. 39 (2021) 1–43.
[148] H. Abdollahpouri, M. Mansoury, R. Burke, B. Mobasher, The connection between popularity bias, calibration, and fairness in recommendation, in: Fourteenth ACM Conference on Recommender Systems, 2020, pp. 726–731.
[149] R. Islam, K.N. Keya, Z. Zeng, S. Pan, J. Foulds, Debiasing career recommendations with neural fair collaborative filtering, in: Proceedings of the Web Conference 2021, 2021, pp. 3779–3790.
[150] D. Liu, P. Cheng, H. Zhu, Z. Dong, X. He, W. Pan, Z. Ming, Mitigating confounding bias in recommendation via information bottleneck, in: Fifteenth ACM Conference on Recommender Systems, 2021, pp. 351–360.
[151] Y. Saito, S. Yaginuma, Y. Nishino, H. Sakata, K. Nakata, Unbiased recommender learning from missing-not-at-random implicit feedback, in: Proceedings of the 13th International Conference on Web Search and Data Mining, 2020, pp. 501–509.
[152] J. Huang, H. Oosterhuis, M. de Rijke, H. van Hoof, Keeping dataset biases out of the simulation: A debiased simulator for reinforcement learning based recommender systems, in: Fourteenth ACM Conference on Recommender Systems, 2020, pp. 190–199.
[153] C. Yang, Q. Wu, J. Jin, X. Gao, J. Pan, G. Chen, Trading hard negatives and true negatives: A debiased contrastive collaborative filtering approach, in: IJCAI, ijcai.org, 2022, pp. 2355–2361.
[154] C. Yang, Q. Wu, J. Jin, X. Gao, J. Pan, G. Chen, Trading hard negatives and true negatives: A debiased contrastive collaborative filtering approach, 2022, arXiv preprint arXiv:2204.11752.
[155] T. Qi, F. Wu, C. Wu, P. Sun, L. Wu, X. Wang, Y. Huang, X. Xie, Profairrec: Provider fairness-aware news recommendation, 2022, arXiv preprint arXiv:2204.04724.
[156] Y. Sun, J. Pan, A. Zhang, A. Flores, FM2: field-matrixed factorization machines for recommender systems, in: WWW, ACM/ IW3C2, 2021, pp. 2828–2837.
[157] D. Antognini, B. Faltings, Fast multi-step critiquing for vae-based recommender systems, in: RecSys, ACM, 2021, pp. 209–219.
[158] J. Zou, E. Kanoulas, P. Ren, Z. Ren, A. Sun, C. Long, Improving conversational recommender systems via transformer-based sequential modelling, in: SIGIR, ACM, 2022, pp. 2319–2324.
[159] W. Krichene, S. Rendle, On sampled metrics for item recommendation, Commun. ACM 65 (2022) 75–83.
[160] Y. Wang, L. Wang, Y. Li, D. He, T. Liu, A theoretical analysis of NDCG type ranking measures, in: COLT, JMLR.org, 2013, pp. 25–54.
[161] D.C. da Silva, M.G. Manzato, F.A. Durão, Exploiting personalized calibration and metrics for fairness recommendation, Expert Syst. Appl. 181 (2021) 115112.
[162] X. Wu, H. Chen, J. Zhao, L. He, D. Yin, Y. Chang, Unbiased learning to rank in feeds recommendation, in: Proceedings of the 14th ACM International Conference on Web Search and Data Mining, 2021, pp. 490–498.
[163] Gómez E., C.S. Zhang, L. Boratto, M. Salamó, G. Ramos, Enabling cross-continent provider fairness in educational recommender systems, Future Gener. Comput. Syst. 127 (2022) 435–447.
[164] M. Marras, L. Boratto, G. Ramos, G. Fenu, Equality of learning opportunity via individual fairness in personalized recommendations, Int. J. Artif. Intell. Educ. (2021) 1–49.
[165] L. Boratto, G. Fenu, M. Marras, The effect of algorithmic bias on recommender systems for massive open online courses, in: European Conference on Information Retrieval, Springer, 2019, pp. 457–472.
[166] T. Qi, F. Wu, C. Wu, Y. Huang, Pp-rec: News recommendation with personalized user interest and time-aware news popularity, 2021, arXiv preprint arXiv:2106.01300.
[167] B. Kirdemir, J. Kready, E. Mead, M.N. Hussain, N. Agarwal, D. Adjeroh, Assessing bias in youtube's video recommendation algorithm in a cross-lingual and cross-topical context, in: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, Springer, 2021, pp. 71–80.
[168] R. Zhan, C. Pei, Q. Su, J. Wen, X. Wang, G. Mu, D. Zheng, P. Jiang, K. Gai, Deconfounding duration bias in watch-time prediction for video recommendation, in: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022, pp. 4472–4481.
[169] C. Wu, F. Wu, T. Qi, Y. Huang, Fairrank: Fairness-aware single-tower ranking framework for news recommendation, 2022, arXiv preprint arXiv:2204.00541.
[170] A. Brek, Z. Boufaïda, Semantic approaches survey for job recommender systems, in: RIF, CEUR-WS.org, 2022, pp. 101–111.
[171] J. Yi, F. Wu, C. Wu, Q. Li, G. Sun, X. Xie, Debiasedrec: Bias-aware user modeling and click prediction for personalized news recommendation, 2021, arXiv preprint arXiv:2104.07360.
[172] D. Shakespeare, L. Porcaro, C. Gómez, Exploring artist gender bias in music recommendation, 2020, arXiv preprint arXiv:2009.01715.
[173] A.B. Melchiorre, E. Zangerle, M. Schedl, Personality bias of music recommendation algorithms, in: Fourteenth ACM Conference on Recommender Systems, 2020, pp. 533–538.
[174] S. Xu, Y. Li, S. Liu, Z. Fu, Y. Zhang, Learning post-hoc causal explanations for recommendation, 2020, CoRR abs/2006.16977.
[175] G. Cornacchia, F. Narducci, A. Ragone, A general model for fair and explainable recommendation in the loan domain (short paper), in: KaRS/ComplexRec@RecSys, CEUR-WS.org, 2021.
[176] P. Dokoupil, L. Peska, Robustness against polarity bias in decoupled group recommendations evaluation, in: Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, 2022, pp. 302–307.
[177] Z. Zhu, S. Si, J. Wang, J. Xiao, Cali3f: Calibrated fast fair federated recommendation system, 2022, arXiv preprint arXiv:2205.13121.
[178] D. Sacharidis, C.P. Mukamakuza, H. Werthner, Fairness and diversity in social-based recommender systems, in: UMAP (Adjunct Publication), ACM, 2020, pp. 83–88.
[179] M. Mansoury, H. Abdollahpouri, J. Smith, A. Dehpanah, M. Pechenizkiy, B. Mobasher, Investigating potential factors associated with gender discrimination in collaborative recommender systems, in: FLAIRS Conference, AAAI Press, 2020, pp. 193–196.
[180] J. Leonhardt, A. Anand, M. Khosla, User fairness in recommender systems, in: WWW (Companion Volume), ACM, 2018, pp. 101–102.
[181] L. Schelenz, Diversity-aware recommendations for social justice? exploring user diversity and fairness in recommender systems, in: UMAP (Adjunct Publication), ACM, 2021, pp. 404–410.

Di Jin received the Ph.D. degree in computer science from Jilin University, Changchun, China, in 2012. He was a research scholar in DMG at UIUC from 2019 to 2020. He is currently a Professor at the College of Intelligence and Computing, Tianjin University, Tianjin, China. His research interests include graph data mining and graph machine learning, especially on community detection, network embedding, and GNNs. To date, he has published more than 100 research papers in top-tier journals and conferences, including the TKDE, TNNLS, TYCB, AAAI, IJCAI, NeurIPS, and WWW. He was the recipient of the Best Paper Award Runner-up of WWW 2021, Best Student Paper Award Runner-up of ICDM 2021, and Rising Star Award of ACM Tianjin in 2018.

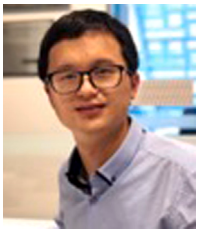
Luzhi Wang is currently a third-year Ph.D. student at Tianjin University, under the supervision of Prof. Di Jin and Prof. Shirui Pan. Her primary research interests revolve around graph representation learning, self-supervised learning, and graph neural networks. She has published numerous papers in top-tier venues and journals, including the IJCAI, WWW, and TWEB, etc.

He Zhang is currently a Ph.D. student with the Faculty of Information Technology, Monash University, Australia. His research interests include graph neural networks, graph attack and defence, and trustworthy graph learning.

Yizhen Zheng is a second-year Ph.D. student at Monash University, supervised by A/Prof. Vincent CS Lee and Prof. Shirui Pan. His research interests lie in graph representation learning, self-supervised learning and graph neural networks. He received the M.Bus.Info.Sys degree from Monash University as a Top 1 student in 2020, M.Bus degree from Monash University in 2018 and B.B.A degree from Shenzhen University in 2016. He published papers on top venues and journals such as ICML, NeurIPS, WWW, AAAI, TNNLS, ICDM, IJCAI, etc. He is invited to review conferences and journals including TPAMI, TNNLS, ECAI, ICDM, TWEB, PR, etc.

Weiping Ding (M'16-SM'19) received the Ph.D. degree in Computer Science, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2013. In 2016, He was a Visiting Scholar at National University of Singapore, Singapore. From 2017 to 2018, he was a Visiting Professor at University of Technology Sydney, Australia. He is a Full Professor with the School of Information Science and Technology, Nantong University, Nantong, China. His main research directions involve deep neural networks, multimodal machine learning, and medical images analysis. He has published over 200 articles, including over 90 IEEE Transactions papers. His fifteen authored/co-authored papers have been selected as ESI Highly Cited Papers. He serves as an Associate Editor/Editorial Board member of IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Fuzzy Systems, IEEE/CAA Journal of Automatica Sinica, IEEE Transactions on Intelligent Transportation Systems, IEEE Transactions on Intelligent Vehicles, IEEE Transactions on Emerging Topics in Computational Intelligence, IEEE Transactions on Artificial Intelligence, Information Fusion, Information Sciences, Neurocomputing, Applied Soft Computing. He is the Leading Guest Editor of Special Issues in several prestigious journals, including IEEE Transactions on Evolutionary Computation, IEEE Transactions on Fuzzy Systems, and Information Fusion.

Dr. Feng Xia is a Professor in School of Computing Technologies, RMIT University, Australia. He is/was on the Editorial Boards of over 10 int'l journals. He has served as the General Chair, PC Chair, Workshop Chair, or Publicity Chair of over 30 int'l conferences and workshops, and PC Member of over 90 conferences. Dr. Xia has authored/co-authored two books and over 300 scientific papers in int'l journals and conferences (such as IEEE TAI, TKDE, TNNLS, TC, TMC, TPDS, TBD, TCSS, TNSE, TETCI, TETC, THMS, TVT, TITS, TASE, ACM TKDD, TIST, TWEB, TOMM, WWW, AAAI, SIGIR, WSDM, CIKM, JCDL, EMNLP, and INFOCOM). He was recognized as a Highly Cited Researcher (2019). Dr. Xia received a number of prestigious awards, including IEEE DSS 2021 Best Paper Award, IEEE Vehicular Technology Society 2020 Best Land Transportation Paper Award, ACM/IEEE JCDL 2020 The Vannevar Bush Best Paper Honorable Mention, IEEE CSDE 2020 Best Paper Award, WWW 2017 Best Demo Award, IEEE DataCom 2017 Best Paper Award, IEEE UIC 2013 Best Paper Award, and IEEE Access Outstanding Associate Editor. He has been invited as Keynote Speaker at seven int'l conferences, and delivered a number of Invited Talks at int'l conferences and many universities worldwide. His research interests include data science, artificial intelligence, graph learning, and systems engineering. He is a Senior Member of IEEE and ACM, and an ACM Distinguished Speaker.



Shirui Pan received a Ph.D. in computer science from the University of Technology Sydney (UTS), Ultimo, NSW, Australia. He is an ARC Future Fellow and a professor at the School of Information and Communication Technology, Griffith University, Australia. His research interests include data mining and machine learning. To date, Dr Pan has published over 150 research papers in top-tier journals and conferences, including the IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), IEEE Transactions on Knowledge and Data Engineering (TKDE), IEEE Transactions on Neural Networks and Learning Systems (TNNLS), ICML, NeurIPS, KDD, AAAI, IJCAI, WWW, CVPR, and ICDM. His paper received the Best Student Paper Award of IEEE ICDM 2020. He is recognized as one of the AI 2000 AAAI/IJCAI Most Influential Scholars in Australia (2022, 2021).