AI and Ethics
https://doi.org/10.1007/s43681-021-00107-7

REVIEW

# Towards responsible media recommendation

**Mehdi Elahi[1] · Dietmar Jannach[2] · Lars Skjærven[3] · Erik Knudsen[1] · Helle Sjøvaag[4] · Kristian Tolonen[5] · Øyvind Holmstad[5] · Igor Pipkin[6] · Eivind Throndsen[7] · Agnes Stenbom[7] · Eivind Fiskerud[8] · Adrian Oesch[8] · Loek Vredenberg[9] · Christoph Trattner[1]**

Received: 7 July 2021 / Accepted: 4 October 2021


**Abstract**
Reading or viewing recommendations are a common feature on modern media sites. What is shown to consumers as recommendations is nowadays often automatically determined by AI algorithms, typically with the goal of helping consumers discover relevant content more easily. However, the highlighting or filtering of information that comes with such recommendations may lead to undesired effects on consumers or even society, for example, when an algorithm leads to the creation of filter bubbles or amplifies the spread of misinformation. These well-documented phenomena create a need for improved mechanisms for *responsible* media recommendation, which avoid such negative effects of recommender systems. In this research note, we review the threats and challenges that may result from the use of automated media recommendation technology, and we outline possible steps to mitigate such undesired societal effects in the future.

**Keywords** Recommender systems · Societal impact · Biases

✉ Mehdi Elahi
mehdi.elahi@uib.no

Dietmar Jannach
dietmar.jannach@aau.at

Lars Skjærven
lars.skjerven@tv2.no

Erik Knudsen
erik.knudsen@uib.no

Helle Sjøvaag
helle.sjovaag@uis.no

Kristian Tolonen
kristian.tolonen@nrk.no

Øyvind Holmstad
oyvind.holmstad@nrk.no

Igor Pipkin
igor.pipkin@amedia.no

Eivind Throndsen
eivind.throndsen@schibsted.com

Agnes Stenbom
agnes.stenbom@schibsted.com

Eivind Fiskerud
eivind.fiskerud@bt.no

Adrian Oesch
adrian.oesch@schibsted.com

Loek Vredenberg
loek.vredenberg@no.ibm.com

Christoph Trattner
christoph.trattner@uib.no

[1] University of Bergen, Bergen, Norway
[2] University of Klagenfurt, Klagenfurt, Austria
[3] TV2, Bergen, Norway
[4] University of Stavanger, Stavanger, Norway
[5] NRK, Oslo, Norway
[6] Amedia, Oslo, Norway
[7] Schibsted, Oslo, Norway
[8] Bergens Tidende, Bergen, Norway
[9] IBM, Oslo, Norway

Published online: 02 November 2021


**Fig. 1** Snapshot of the mobile app of *Bergens Tidende*, one of the largest newspapers in Norway, showing news recommendations

## 1 Introduction

Many modern media sites nowadays provide content recommendations for their online consumers, e.g., additional news stories to read or related videos to watch (see Fig. 1). The selection of the content to be presented to the users is increasingly automated and done with the help of machine learning algorithms. Such *recommender systems*, which typically rely both on individual user interests and collective preference patterns in a community, are commonly designed to make it easier for consumers to discover relevant content. At the same time, personalized recommendations can also create value for the media providers, e.g., in terms of increased user retention or ad revenue, see [47] for an overview. Kirshenbaum et al. [53] and Garcin et al. [39] for example both report that recommendations increased the click-through rates on their news sites by more than 30%. In the online streaming domain [41], furthermore, discuss the various ways recommendations can create business value at Netflix, e.g., in terms of customer retention.

However, the use of recommendation technology may also lead to certain undesired effects, some of which only manifest themselves over time. Probably the best known example is the phenomenon of the "filter bubble" [77]. Such a bubble can emerge when the algorithms learn about user interests and opinions over time, and then start to solely present content that matches these assumed interests and opinions. Ultimately, this can lead to self-reinforcing feedback loops which may then result in undesired societal effects, such as opinion polarization or the increased spread of one-sided information [21].

A common argument is that the emergence of such phenomena is often a result of how the underlying algorithms work or what they are optimized for. For example, when the goal is to maximize user interaction—and thus clicks and ad impressions—an algorithm may learn that the best choice is to recommend what the consumer liked in the past or what is generally popular or trending [4]. Recommendation algorithms focused on such optimization goals can further lead to addicting users to social media platforms [8, 84, 101]. Furthermore, we cannot rule out that there are cases, where recommendations providers do not view this to be problematic, e.g., due to their goal to maximize short-term profitability [103]. Following a specific political agenda can also be a motivation, e.g., in the infamous case of Cambridge Analytica, who employed mechanisms of Facebook to target voters in 2014 and 2015 [74]. In many other cases, however, organizations may have an interest to avoid negative effects through more *responsible recommendations*. Public broadcasters in Europe, for example, often have the explicit mission to provide unbiased political information or to deliver content that is diverse in nature. As an example, the Council of Europe has established standards for public broadcasters to produce programmes that reflect the cultural and linguistic diversity auf the audience [26]. Furthermore, the British Broadcasting Corporation (BBC) has formulated a set of principles on its own for the provision of news and TV programmes, with the goals, e.g., of representing the different cultures of their audience and to represent alternative opinions [11]. Another example is Norway, where the diversity of opinions is reflected in the official Norwegian media policy [72] and anchored in Article 100 in the

Norwegian constitution. This mission should then also be reflected in the recommendations, which often have a major influence of what users consume online. But also private organizations might be interested in avoiding one-sided or unbalanced recommendations, as this might contradict their corporate mission or might simply hurt their public reputation in the long run.

Next, in Sect. 2, we review possible threats and undesired side effects of recommendations and we shed some light on the underlying reasons for the emergence of this effects. Afterwards, in Sect. 3, we discuss a selected set of existing approaches to deal with these challenges and to deliver responsible recommendations.

## 2 Undesired effects and underlying causes

Prior research has identified a number of undesired effects that can be unintentionally caused or intensified by recommender systems. Some of these effects can be mainly attributed to characteristics of the algorithms that generate the recommendations. Other effects, in contrast, largely stem from particularities of the data that is used by the algorithms [24, 25], such as the history of recorded user interactions. Next, we review a number of such negative effects in some more depth before we summarize the potential underlying reasons.

### 2.1 Description of undesired effects

*Filter bubbles*,[1] as mentioned above, are one of the most frequently discussed potential effects of personalization and recommendation, which assumedly may pose serious threats to individuals and societies. A filter bubble refers to a social environment that lacks the exposure to diverse beliefs and viewpoints. It can occur in undiversified communities, where people are *encapsulated* within a stream of media content (e.g., videos or news articles) that is optimized to match their specific preferences [70]. This effect can be created or reinforced by recommender systems, by over-personalizing the media content based on the users' interests, and consequently, trapping them within an unchanging environment [77]. While "good" personalization helps users to obtain relevant information and hence addresses information overload [71], overdoing it can lead the users to only view what they individually *want* and keeping them inside a closed world, cut out of the outside (or: diverse) world [96]. In the long term, this can harm users when they become isolated from outside of the "bubble" and create additional negative effects, such as partial information blindness [77]. As a result, it becomes unlikely that users will receive recommendations of less attractive but important content. Instead, they will be surrounded by the viewpoints of like-minded users, and protected from surprising information, or information that challenges their opinions. This may ultimately weaken their creativity, as well as their thinking and learning capabilities [70].[2]

*Echo chambers*—another potential effect of recommendations—refer to a polarized environment, where only certain viewpoints, information, and beliefs are shared via communication. In such an environment, the users' viewpoints are repeatedly amplified through recurring exposure to similar media content [40]. This situation is more likely to occur within closed communities, where people will only share opinions that they are in high agreement, without free circulation of information with the outside world [29, 46]. Echo chambers can be seen as an inevitable effect in social media networks, due to their particular characteristics, which can easily result in the formation of homogeneous and segregated communities [40, 82]. Members of such polarized communities tend to ignore information that is conflicting with their beliefs and ideas [45, 55]. Recommender systems can even intensify the echo chamber effect by suggesting media content to users that reconfirms their background beliefs and existing viewpoints, and hence, decrease their exposure to more diverse opinions.

The reinforced *spread of misinformation*, i.e., the communication and circulation of false and misleading information, is another potential negative effect of recommendations. This information that is spread is, however, not necessarily meant to deceive people. *Disinformation*, in contrast, refers to false information that is created and communicated in order to deceive people [56]. Recommender systems can unintentionally contribute to both of these undesired effects, thus posing serious threats to communities. A notable example is the spread of misinformation on the Swine Flu on Twitter [69]. Despite the lack of concrete evidence, it is commonly believed that the Twitter recommendation algorithm has facilitated and reinforced the spread of that misinformation

[1] A filter bubble can be considered as a stage of a bigger effect, *Information Polarization*, which occurs when the individuals have no or limited access to diverse media content, and hence, are exposed to a narrow range of information sources [66].

[2] We acknowledge that a growing body of literature suggests that today's technology and use of recommender systems actually may have *not* isolated large segments of the audience into bubbles to a large extent (e.g., [12, 37, 38, 42, 68], or that filter bubbles are rather mainly formed in our heads [15, 16]. In addition, the threats of creating filter bubbles might be much more pronounced for large content aggregators such as Google and Facebook than for more traditional media sites that mainly provide curated content. Similar considerations apply for echo chambers. It is nonetheless important to highlight that the situation can quickly change when technology improves and their use increases [105].

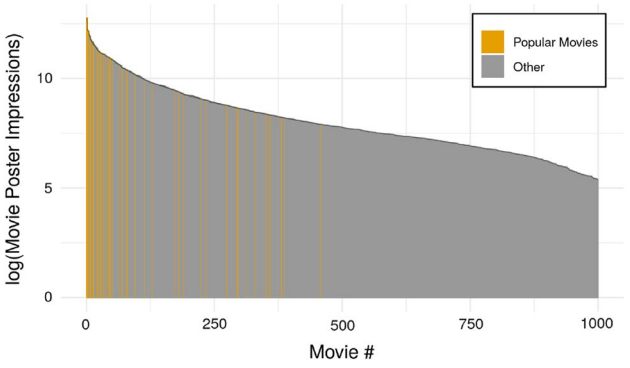


**Fig. 2** Impressions of movies recommendations on the front page of TV 2 Play, *plotted in logarithmic scale*. The yellow bars represent the top 50 popular movies based on the number of playbacks made by users (colour figure online)

so that it has reached a very large user community and consequently caused panic in parts of the population [33]. Online social platforms are a primary medium for the spread of such misinformation, often due to the lack of editorial control. As a result, these platforms are often considered as unreliable and untrustworthy sources of news [18].

*Popularity bias*, i.e., the tendency of a recommender system to focus on popular items, is an effect that often originates from the characteristics of the data that is used to generate the recommendations. In real-world data collections, a large fraction of the contained information is often related to a small set of popular ("blockbuster") items, known as the *short head*. The rest of the data, in contrast, is related to the *long tail* of average or niche items [1, 2, 13, 32]. For example, Fig. 2 shows recent data from TV 2 Play,[3] one of the largest movie streaming platforms in Norway. The numbers clearly indicate that a small fraction of the movies that are recommended on the front page accounts for a large number of the recorded page impressions. Interestingly, while recommender systems are often considered as means to increase sales in the long tail, the concentration on the short head can in fact be increased by a recommender system [36, 48, 60]. A popularity bias can be amplified by a recommender system when it learns from the recorded data to recommend popular items more frequently than less popular items. Not all recommendation algorithms exhibit such tendencies to the same extent, as discussed in Ref. [48]. In general, while recommending popular items is often considered a safe strategy in practice, it is not beneficial for the discovery of fresh or niche items. Moreover, such a strategy also leads to a limited level of personalization and can push the choices and consumption behavior of users towards the *mainstream*, which is not always a desired effect. Several works have studied popularity biases and reported the existence of such effects for various online platforms that serve their users with some forms of recommendation, see, e.g., [64] for the case of Spotify.

*Discrimination* is another potential side effect of recommendations that may harm individuals or certain groups in a society. Discrimination can be defined as the unfair or unequal treatment of individuals, groups, classes or social categories according to certain characteristics, e.g., gender, age, income, education, religion, or ethnicity race [35, 79]. When discrimination is the result of using an intelligent system like a recommender system, the phenomenon is often referred to as "digital discrimination". Nowadays, digital discrimination is becoming more prevalent. Today, it is considered a serious challenge due to the increasing number of decisions that are either made automatically by such systems or due to human decisions that are based on the output of algorithms. Discrimination through recommendation can occur in different forms and can affect certain individuals or groups within a social environment. As an example, it has been shown that collaborative filtering algorithms—which are among the most popular recommendation techniques in the media industry—may intensify existing gender biases that are inherited from the input data [85]. When such algorithms are used, it may become much less likely that female artists are pushed compared to male artists [34].

Finally, *unfairness* (or: the lack of fairness) is among the most important challenges that may result as a side effect of automated recommendations. Research on fairness and unfairness can be traced back to well over 50 years [44], and it has received renewed attention in the most recent years, in particular also in the areas of machine learning or artificial intelligence in general. Informally speaking, unfairness refers to a social environment, where individuals perceive a severe lack of fairness. Fairness, in turn, may be characterized as the absence of any bias, prejudice, favoritism, mistreatment toward individuals, group, classes, or social categories based on their inherent or acquired characteristics [25]. While such a characterization of fairness is certainly helpful, the notion of fairness often remains vague and no common definition has been established within the relevant literature. Prior studies on algorithmic fairness reported that the perception of fairness may strongly vary across individuals. For instance, [95] found that the perception of fairness largely differed across hundreds of participants of an experiment. This can be due to the complexity of the topic or the highly sensitive, contextual, and subjective nature of fairness [57].

In the context of recommender systems, different forms of fairness can be defined. [3] recently proposed a taxonomy of different classes of fairness in recommender systems according to the various stakeholders. They defined *C-fairness*, where the focus is on the perspective of those

[3] https://play.tv2.no.

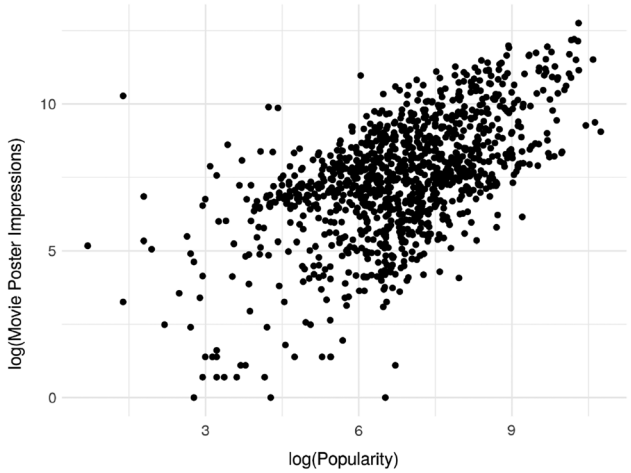


**Fig. 3** Comparing the popularity of movies recommended on the front page of TV 2 Play and the impressions (views) they received from the users

who receive the recommendations (consumers); *P-Fairness* for those who provide the items or content (providers); and *S-fairness* for those who neither receive nor provide the recommendations yet are influenced by the recommendations as side stakeholders.

## 2.2 Discussion of underlying reasons

Various of the discussed phenomena, including unfairness and discrimination, can be caused by different forms of bias in the data. The potential undesired effects of recommendations, therefore, often enter the system through the data, since the system learns from the data to replicate preexisting biases [30]. As an example, it is estimated that 68.5% of Twitter users are male, where only 31.5% are female [91]. It is, therefore, easy to imagine that a system may recommend disproportionately more male users to follow than female users. As another example, according to data by TV 2 Play shown in Fig. 2, a narrow range of movies (44 out of about 1000 movies) recommended on the front page received almost half of the impressions. This may indicate that the implemented recommendation algorithms have a tendency to recommend already popular items, e.g., movies that have been watched frequently previously. The popularity of movies and the number of impressions through the recommendations on TV 2 Play are plotted in Fig. 3, where a correlation can be clearly observed.

Technically, bias can be defined as a deviation from the standard, indicating the existence of some form of statistical patterns in the data [27, 35]. Bias in the data can come from how the data are collected. *Selection bias*, for example, refers to an ill-designed process of data collection and sampling, where the data have been obtained from subgroups of a population through a *specific* form of process (e.g., a non-random process). As a result of such a selection bias, the trends estimated for a population cannot generalize to data collected from a new population [63]. For example, consider a dataset collected by a social media company surveying the video tastes of its users. If the website is mainly used by experts users, with a degree in art or cinema, the elicited preferences will be biased, and hence, not representative of the entire society. Another type of bias is the *population bias*, which refers to situations, where statistics, demographics, representatives, and user characteristics are different in the user population represented in the data set from the target population [75]. As an example, this bias can arise when relying on data collected from a social network (e.g., Snapchat, which is mostly used by younger individuals) to make recommendations for a population with different demographics (e.g., forum users on Reddit).

But not only biases in the data can contribute to the creation or intensification of the described phenomena. In particular, the recommendation algorithms can be another root cause for several of them, as discussed. As an example, it has been found that the number of friends for a Facebook user does not only reflect the popularity of a user, but is also dependent on the bias of the recommendation algorithm of Facebook [94]. Algorithms can also amplify already existing biases. For example, recommendation algorithms that are trained on the MovieLens data set[4]—a very popular data set in the research community— were found to strongly intensify the preexisting popularity bias they inherited from the data set [7, 60].

A bias amplification tendency of certain algorithms is often rooted in their specific optimization goal (or technically, their objective function). Real-world recommender systems typically focus on optimizing the underlying algorithm according to the given Key Performance Indicators, e.g., to increase sales by converting visitors into buyers. Therefore, the goal is to create recommendation lists that maximize the probability that a customer makes an order, i.e., to create lists that increase the *conversion rate* [89]. While the conversion rate is a common metric in different business sectors, it has been shown that it can create a severe case of popularity bias [97]. Hence, the negative effects of recommendations can go beyond the users (consumers of items) and also damage the business (suppliers of items).

In a different study [22], it has been shown that recommendation algorithms, biased towards popular items, might undermine the consumption (or sales) of unpopular items (long tail), hence, preventing such items to ever gain visibility and become popular. This may not be a big challenge for companies, where the majority of the revenue comes from

[4] https://grouplens.org/datasets/movielens/.

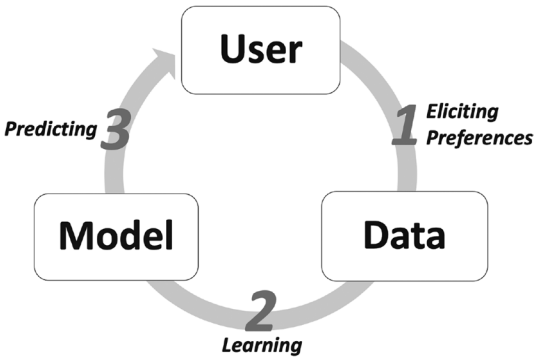


**Fig. 4** Feedback loop of recommender systems

a few popular items. However, it may represent a problem if the main business of a company is based on selling from the long tail of less-popular items. Such biased algorithm can thus negatively impact the revenue of such a company and cause significant damage [10].

Recent research has addressed this problem, for example by defining novel optimization objectives that also consider the diversity of the recommendations [9], or are able to balance the popularity, novelty, or diversity of the items that are recommended to users [1, 50, 88].

## 3 Mitigating the undesired effects

There has been a rising attention paid by the research community on the threats posed by recommender systems and their potential impact on individuals or societies [76]. Correspondingly, numerous research works have focused on designing solutions as countermeasures to mitigate these threats.

Each of the solutions may target a specific key component of the environment of a recommender systems, i.e., *user*, *data*, and *model*, and may be applied at different stages of the feedback loop of recommender systems shown in Fig. 4 [25]:

- user → data refers to the preference elicitation stage, where the users provide data to the system as explicit or implicit feedback;
- data → model refers to the learning stage, where machine learning algorithms are exploited to build user models based on the elicited data;
- model → user refers to the process of predicting preferences of users based on the elicited information and to the process of generating recommendations accordingly.

In this section, we briefly describe various solution approaches from the literature, categorized into data-driven approaches, algorithmic approaches, and user-centric approaches.

### 3.1 Data-oriented approaches

#### 3.1.1 Data de-biasing

Several techniques have been proposed to de-bias the data. As an example, various techniques addressed the *selection bias* in recommender systems, which typically impacts the evaluation phase [25]. The primary reason for this bias can be that the available data often is not a representative sample of the user preferences, as discussed above. This is in parts due to the fact that users are free to decide for which items they provide their feedback (e.g., in the form of ratings). A potential solution to mitigate the selection bias is to redefine the prediction model to learn to predict which data is missing. Hence, in addition to predicting the relevance of items for a target user, a second task is to predict the likelihood that an item is chosen by the user to rate. The assumption is that the chance of choosing an item by a user to rate will depend on the value of the rating the user will provide [25]. Technically, the probability of observing a user-item interaction can for example be modelled by mixture of multinomials [62], logit [61] or matrix factorization models [43].

Another known phenomenon is the *conformity bias*, which happens when users are influenced by the opinions of others (e.g., on social media) and when their expressed preferences deviate from their true preferences. An example solution can be to treat the observed preferences of users as a synthetic outcome of combining the true preferences with social influence. As a result, social influence is taken directly into account in the recommendation model [25].

Data biases and de-biasing approaches are highly relevant in practice. At *Bergens Tidende*, for example, an age-related bias is often observed in the data, resulting from the demographics of the subscriber population, which has a high proportion of readers above the age of 50. This bias may make it difficult to appropriately serve younger audiences, which is, however, desirable both from a societal and commercial perspective. Hence, it can be important to apply methods to mitigate these types of biases when serving recommendations, e.g., by incorporating additional user features (e.g., age) within the user profile [59]. However, incorporating such extended features needs further considerations regarding fairness in user modelling, as discussed in the next section.

#### 3.1.2 Fair user modeling

User modeling refers to the process of creating and modifying a conceptual representation of the users, and it deals with the personalization and adaptation of systems according

the specific preferences and needs of the user [58]. In this context, fairness can refer to a modeling process that does not create any unfair discrimination or unjust consequences [99]. Accordingly, fairness in user modeling describes the condition, where the model, built on top of the user data, can fairly represent the values of the users.

Various approaches have addressed fairness in user modeling. Existing recommender systems tend to collect user data in high volumes and large varieties. It is often believed that every single action of online media users is carefully monitored and precisely recorded. While some of the recorded data can be essential, others may not necessarily be needed or may expose sensitive information about the users. Building user models on top of such data can cause serious issues of user *privacy*. To address such issues, some approaches proposed to eliminate sensitive attributes of the users (e.g., gender, religion, or race) when building models [23, 100]. While this can be effective in avoiding unfairness, it may fail to work properly in certain cases, e.g., when the sensitive attributes are highly correlated with other attributes [52]. In addition to that, eliminating certain attributes of users can reduce the recommendation quality. In order to address this, some approaches utilized embedding techniques to encode the attributes before building the models. Consequently, the resulting user models do not directly measure the sensitive attributes and instead compute latent features for describing the users [100].

## 3.2 Algorithmic countermeasures

A number of algorithmic approaches were proposed to deal with the potential undesired effects and biases of recommender systems [25, 90]. Increasing the *diversity* of the returned recommendations is often a central approach [71]. Technically, some existing works in this direction enhance the diversity of the recommendation output by modifying the core recommendation algorithms. Others rely on re-ranking the output of an existing recommendation algorithm [2, 6, 50, 93]. In the former case, the rating prediction model is extended with additional terms aiming to improve the fairness of the system, e.g., by reducing the bias. In the latter case, re-ranking techniques are applied on top of the existing recommendation algorithm, e.g., to post-process the recommendation output and to build a more diversified list [4]. In this section, we briefly review such approaches.

### 3.2.1 Enhancing diversity, novelty, and serendipity

Over the past years, several approaches were proposed to enhance the diversity of the recommendations created by a system. One of the early works on diversity is [14], where an algorithm based on *Bounded Random Selection* was proposed. Another example for an early work on diversity is [104], where the authors developed an algorithmic framework focusing on topic diversification. In addition, a dissimilarity metric was proposed to measure the level of diversification and the effectiveness of the underlying algorithm. Another notable work is [78], where the authors proposed a technique that can positively enhance the diversity of a recommender system for different stakeholders, i.e., users (consumers of items) and business (suppliers of items). The technique sets a minimum threshold for the exposure of different items, ensuring that a wider range of suppliers are listed in the recommendations generated for users. Various other techniques, however, exist as well, which were proposed earlier but are not directly tied to the problems of responsible recommendation, see, e.g., [6, 51, 50].

Serendipity is another important dimension in recommender systems which can contribute to the perceived fairness of a system. Serendipity as a concept typically is considered to reflect the *surprise element* of recommendations. Recommending serendipitous items can also be considered as an attempt to reduce potential biases and hence improve the fairness of a recommender system. Remember that the continuous recommendation of items that are already known to users may reinforce the recommendation of popular items, hence intensifying the popularity bias. Emphasizing serendipity and novelty can help to promote items that have not had many chances to receive user feedback. It has been shown that a certain lack of novelty and serendipity in recommender systems can contribute to an overall dissatisfaction of users [102]. However, introducing higher serendipity levels has to be done carefully as some users are more engaged when surprising recommendations are suggested to them, while others may become disengaged, even dissatisfied, with such recommendations. Different research works exist which focus on designing recommender algorithms that can deliver relevant recommendations while including new items that the users might be less familiar with [5]. This is in fact a crucial capability, since there is a known trade-off between relevance, serendipity and novelty within recommender systems. Ultimately, it is important that the the system fairly deals with different types of user with different attitudes towards novel and serendipitous content.

### 3.2.2 Technical approaches for enhancing recommendations

Looking at technical approaches, the problem of *fair rankings* has traditionally been dealt with from the perspective of search engines and the results they provide to users. In this context, fairness refers to the condition, where the generated ranking contains a sufficiently diverse set of items that reflects the interest of different groups of the society (e.g., underrepresented groups) and avoids statistical discrimination against such groups [20].

In the context of recommender systems, *re-ranking* algorithms have been typically employed to post-process the output of recommender systems to achieve a certain goal (e.g., improving fairness). An example is the work of [2], where the authors propose a post-processing technique (dubbed *xQuAD*) in order to balance the exposure of the different items in the catalog. This approach empowers the systems to tune the output towards the generation of fair recommendations. Another example of a re-ranking techniques is called *ReGularization (RG)*, which aims to improve fairness through balancing the ratio of popular and less popular items in the recommendation output. Technically, this is done by extending the objective function with an additional regularization term. Accordingly, recommendations containing more popular items are penalized in order to better create a balance between popular and unpopular items [1]. A similar technical approach was proposed in Ref. [88] for the problem of news recommendation. In their approach, the goal was, however, to balance item novelty with relevance, allowing the system, for example, to promote novel content that is not yet too popular. In addition to the above described approaches, there are some works that adopt techniques to perform multi-objective optimization by simultaneously optimizing both accuracy and diversity [19, 90]. Another example is the work by [80], which utilizes genetic algorithms capable of balancing accuracy, diversity, and novelty when generating recommendations for users.

## 3.3 User-centric approaches

Algorithmic fairness, as described in the previous section, can play an important role in mitigating the negative impacts of the noted phenomena. However, it could also be too simplistic to believe that this type of fairness can solve the entire problem. Hence, one should not ignore the other, more user-centric aspects of fairness, which can play an important role as well.

### 3.3.1 Dimensions of user-centric fairness

User-centric fairness can be studied along different dimensions, including in particular *Engagement*, *Representation*, and *Action and Expression* [31]. Engagement refers to how different users are engaged with the recommender system and interact with the provided recommendations. A wide range of factors can impact the engagement of users, e.g., culture, beliefs, personal characteristics, ethnicity, or education. The Representation dimension refers to the adoption of different means when presenting recommendations to users. This will enable different groups of users, with different characteristics (e.g., with different abilities), to properly comprehend the presented recommendations. Providing explanations for the recommendation or summarizing the key features of the recommendations, e.g., through supporting materials, are examples to improve the fairness of a recommender system from the representation point of view. The Action and Expression dimensions refers to supporting users in expressing their feedback on the recommendations through different channels. This may be required due to the fact that different users may prefer different ways of interacting with a recommender system and expressing their opinion. Some may prefer to provide their feedback to the recommendations via pressing a button and some via writing it down. The system should offer different ways to give feedback and hence allow users to gain a certain level of control on the provided recommendations.

### 3.3.2 Creating transparency

Research on transparency dates back to nearly 40 years ago, where early works on expert systems proposed basic forms of explanations and justifications for the advice made by these systems [17]. Later on, research works on search engines indicated that transparency may largely improve the performance of a search engine from the users' perspective, often leading to higher satisfaction with the system [54].

While no common definition can be found within the relevant literature for the concept of transparency, some works provide a generic description of transparency as an information exchange between a *subject* and an *object*, where the subject receives the information describing a service or a system that the object is responsible for [65, 98]. Other works characterize transparency as a set of best practices regarding how users should be provided with insights about a system, hence, enabling them to understand *why* and *how* it works [83].

In the context of recommender systems, the need for transparency has been articulated more frequently in recent years, see [73]. Traditionally, users of recommender systems mainly expect the recommendations to be *accurately* personalized. In future fairness-aware and responsible recommender systems, however, users may more often expect the recommendation to be communicated and presented transparently. In the literature, different forms of transparency are discussed. One predominant form of establishing transparency is to provide explanations about how the system works. Accordingly, the system should be able to provide sufficient information on the relationship between the input of the system (e.g., user preferences) and the mechanism that led to its output (i.e., the recommendations). This information helps users to gain a better understanding of the recommendation process, thereby enabling them to revise their input in order to improve the recommendations. An example of such research work is given in Ref. [86], where the authors conducted a user study comparing five music recommender systems. The results of the study showed that users felt more

confident when the recommendation process is perceived by them as being transparent.

#### 3.3.3 Increasing awareness

One of the key areas of interest in user-centric approaches is *user awareness*, and a number of studies investigated the potential impact of this factor. Some of the studies focused only on raising the awareness of users towards the potential threats in recommender systems by providing some form of explanations to them on *why* and *how* the system is acting responsibly [87, 92]. An example can be a news recommender systems that notifies users that some of the articles might be disputed and may need careful attention by the user [67]. Other studies have gone beyond such a simple approach and devised tools and methods that can further support users to act properly in problematic situations, e.g., methods that can automatically detect fake news in recommended news articles and inform the users appropriately on how to get rid of them [28, 81]. The argument for such approaches is that while raising the awareness of users regarding fairness issues is an essential objective, offering solutions to address these issues is another equally important objective. Addressing both objectives can better help users to gain knowledge, and at the same time, support them to find and use the concrete countermeasures provided by the system. Such countermeasures are often among the more hidden features of the system and their functionalities may not always very clear to the users [49]. An example of a work in this area is [87], where the authors conducted an exploratory study to investigate the user perception of fairness and fairness-aware objectives in a recommender system. The study concluded with three important suggestions:

- Recommender systems should offer explanations to describe the fairness objective of the system for the users.
- Recommender systems should not provide explanations in order to nudge users into making a choice, although the goal might be fairness.
- Recommender systems should explain the motivation for considering fairness as an objective of the system.

Regardless of the goals and methodology, any form of transparency as discussed above may be beneficial for users and improve the perceived fairness of the system.

## 4 Conclusion

Algorithm-generated recommendations are nowadays ubiquitous on the Web, in particular on media sites, where recommender systems are used to suggest news content or videos to watch for users. While there are many industry reports on the benefits of recommender systems, such personalized systems may also lead to undesired effects on individuals, communities, or a society as a whole. In this paper, we have reviewed the corresponding challenges and threats and outlined existing approaches to mitigate problems such as biases or the lack of fairness. Overall, while there is an increasing awareness in the community of these problems, more research is still needed to develop future techniques for responsible media recommendation.

A coordinated effort to address these problems is currently made in the recently established MediaFutures Research Centre for Responsible Media Technology & Innovation.[5] The centre involves a number of partners from academia and industry, including the most important players in media in the Nordic region as well as a partner from the global tech industry, such as TV2 and NRK (two main TV broadcasters in Norway); Schibsted, including Bergens Tidende (BT), and Amedia as the two largest news media houses in Norway; and the global tech and media player IBM. One main objective of the project is to study and tackle negative effects of recommendation technologies and to develop a new generation of responsible media technology by leveraging state-of-the-art AI technology for the media sector.

**Funding** Open access funding provided by University of Bergen (incl Haukeland University Hospital). This work was supported by industry partners and the Research Council of Norway with funding to MediaFutures: Research Centre for Responsible Media Technology and Innovation, through the Centres for Research-based Innovation scheme, project number 309339.

## Declarations

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.



[5] https://mediafutures.no

## References


1. Abdollahpouri, H., Burke, R., Mobasher, B.: Controlling popularity bias in learning-to-rank recommendation. In: Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys '17, pp. 42–46 (2017)
2. Abdollahpouri, H., Burke, R., Mobasher, B.: Managing popularity bias in recommender systems with personalized re-ranking. In: Proceedings of the Thirty-Second International Florida Artificial Intelligence Research Society Conference (FLAIRS '19), pp. 413–418 (2019)
3. Abdollahpouri, H., Adomavicius, G., Burke, R., Guy, I., Jannach, D., Kamishima, T., Krasnodebski, J., Pizzato, L.: Multistakeholder recommendation: survey and research directions. User Model. User Adapt. Interact. **30**(1), 127–158 (2020)
4. Abdollahpouri, H., Mansoury, M., Burke, R., Mobasher, B., Malthouse, E.: User-centered evaluation of popularity bias in recommender systems. In: Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '21, pp. 119–129 (2021)
5. Adamopoulos, P., Tuzhilin, A.: On unexpectedness in recommender systems: or how to better expect the unexpected. ACM Trans. Intell. Syst. Technol. **5**(4), 1–32 (2014)
6. Adomavicius, G., Kwon, Y.: Improving aggregate recommendation diversity using ranking-based techniques. IEEE Trans. Knowl. Data Eng. **24**(5), 896–911 (2012)
7. Adomavicius, G., Bockstedt, J., Curley, S., Zhang, J.: Debiasing user preference ratings in recommender systems. In: Proceedings of the Workshop on Interfaces and Human Decision Making for Recommender Systems (IntRS 2014), pp. 2–9 (2014)
8. Andersson, H.: Social media apps are 'deliberately' addictive to users—bbc. https://www.bbc.com/news/technology-44640959 (2008). Accessed 1 Jun 2021
9. Antikacioglu, A., Ravi, R.: Post processing recommender systems for diversity. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 707–716(2017)
10. Baeza-Yates, R.: Bias in search and recommender systems. In: Proceedings of the Fourteenth ACM Conference on Recommender Systems (RecSys '20) (2020)
11. BBC: Mission, values and public purposes—about the BBC. https://www.bbc.com/aboutthebbc/governance/mission (2019). Accessed 1 Jun 2021
12. Bechmann, A., Nielbo, K.L.: Are we exposed to the same "news''in the news feed? An empirical analysis of filter bubbles as information similarity for Danish Facebook users. Digit. J. **6**(8), 990–1002 (2018)
13. Boratto, L., Fenu, G., Marras, M.: Connecting user and item perspectives in popularity debiasing for collaborative recommendation. Inf. Process. Manag. **58**(1), 102387 (2021)
14. Bradley, K., Smyth, B.: Improving recommendation diversity. In: Proceedings of the Twelfth Irish Conference on Artificial Intelligence and Cognitive Science, pp. 141–152 (2001)
15. Bruns, A.: Filter bubble. Internet Policy Rev. **8**(4) (2019)
16. Bruns, A.: It's not the technology, stupid: how the'echo chamber' and 'filter bubble' metaphors have failed us. International Association for Media and Communication Research (2019)
17. Buchanan, B.G., Shortliffe, E.H.: Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project. Addison-Wesley, Boston (1984)
18. Budak, C., Agrawal, D., El Abbadi, A.: Limiting the spread of misinformation in social networks. In: Proceedings of the 20th International Conference on World Wide Web, pp. 665–674 (2011)
19. Caldeira, J., Oliveira, RS., Marinho, L., Trattner, C.: Healthy menus recommendation: optimizing the use of the pantry. In: Proceedings of Health RecSys Workshop at ACM RecSys '18 (2018)
20. Castillo, C.: Fairness and transparency in ranking. ACM SIGIR Forum **52**(2), 64–71 (2019)
21. Celis, LE., Kapoor, S., Salehi, F., Vishnoi, N.: Controlling polarization in personalization: an algorithmic framework. In: Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19), pp. 160–169 (2019)
22. Chaney, AJB., Stewart, BM., Engelhardt, BE.: How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In: Proceedings of the 12th ACM Conference on Recommender Systems (RecSys '18), pp. 224–232(2018)
23. Chausson, O.: Who watches what? Assessing the impact of gender and personality on film preferences. Paper published online on the MyPersonality project website. http://www.mypersonalityorg/wiki/dokuphp (2010). Accessed 5 Jun 2021
24. Chen, J., Feng, Y., Ester, M., Zhou, S., Chen, C., Wang, C.: Modeling users' exposure with social knowledge influence and consumption influence for recommendation. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM '18), pp. 953–962 (2018)
25. Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., He, X.: Bias and debias in recommender system: a survey and future directions. CoRR arXiv:2010.03240 (2020)
26. Council of Europe, Commissioner: Public service broadcasting under threat in Europe. https://www.coe.int/en/web/commissioner/-/public-service-broadcasting-under-threat-in-europe (2017). Accessed 5 Jun 2021
27. Danks, D., London, AJ.: Algorithmic bias in autonomous systems. In: Procceedings International Joint Conference on Artificial Intelligence (IJCAI '17), vol. 17, pp. 4691–4697 (2017)
28. Della Vedova, ML., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M., de Alfaro, L.: Automatic online fake news detection combining content and social signals. In: Proceedings 22nd Conference of Open Innovations Association (FRUCT), pp. 272–279 (2018)
29. Dubois, E., Blank, G.: The echo chamber is overstated: the moderating effect of political interest and diverse media. Inf. Commun. Soc. **21**(5), 729–745 (2018)
30. Ekstrand, MD., Kluver, D.: Exploring author gender in book rating and recommendation. User Model. User Adapt. Interact. **31**, 377–420 (2021)
31. Elahi, M., Abdollahpouri, H., Mansoury, M., Torkamaan, H.: Beyond algorithmic fairness in recommender systems. In: Adjunct Proceedings of the ACM Conference on User Modeling, Adaptation and Personalization (UMAP '21) (2021)
32. Elahi, M., Kholgh, D.K., Kiarostami, M.S., Saghari, S., Rad, S.P., Tkalcic, M.: Investigating the impact of recommender systems on user-based and item-based popularity bias. Inf. Process. Manag. **58**, 102655 (2021)
33. Fernandez, M., Bellogin, A.: Recommender systems and misinformation: the problem or the solution? In: Proceedings of the Workshop on Online Misinformation- and Harm-Aware Recommender Systems at ACM RecSys '20, pp. 22–26(2020)
34. Ferraro, A., Serra, X., Bauer, C.: Break the loop: gender imbalance in music recommenders. In: Proceedings of the 2021 Conference on Human Information Interaction and Retrieval (CHIIR '21), pp. 249–254 (2021)
35. Ferrer, X., van Nuenen, T., Such, J.M., Coté, M., Criado, N.: Bias and discrimination in ai: a cross-disciplinary perspective. IEEE Technol. Soc. Mag. **40**(2), 72–80 (2021)

36. Fleder, D., Hosanagar, K.: Blockbuster culture's next rise or fall: the impact of recommender systems on sales diversity. Manag. Sci. **55**, 697–712 (2009)
37. Fletcher, R.: The truth behind filter bubbles: Bursting some myths. Reuters Institute for the Study of Journalism. https://www.reutersinstitute.politics.ox.ac.uk/risj-review/truth-behind-filter-bubbles-bursting-some-myths (2020). Accessed 1 Jun 2021
38. Fletcher, R., Nielsen, R.K.: Are news audiences increasingly fragmented? A cross-national comparative analysis of cross-platform news audience fragmentation and duplication. J. Commun. **67**(4), 476–498 (2017)
39. Garcin, F., Faltings, B., Donatsch, O., Alazzawi, A., Bruttin, C., Huber, A.: Offline and online evaluation of news recommender systems at swissinfo.ch. In: Proceedings of the 8th ACM Conference on Recommender Systems (RecSys '14), pp. 169–176 (2014)
40. Ge, Y., Zhao S., Zhou, H., Pei, C., Sun, F., Ou, W., Zhang, Y.: Understanding echo chambers in e-commerce recommender systems. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20), pp. 2261–2270 (2020)
41. Gomez-Uribe, C.A., Hunt, N.: The Netflix recommender system: algorithms, business value, and innovation. Trans. Manag. Inf. Syst. **6**(4), 13:1-13:19 (2015)
42. Haim, M., Graefe, A., Brosius, H.B.: Burst of the filter bubble? Effects of personalization on the diversity of Google News. Digit. J. **6**(3), 330–343 (2018)
43. Hernández-Lobato, JM., Houlsby, N., Ghahramani, Z.: Probabilistic matrix factorization with non-random missing data. In: International Conference on Machine Learning (ICML '14), pp. 1512–1520 (2014)
44. Hutchinson, B., Mitchell, M.: 50 years of test (un) fairness: Lessons for machine learning. In: Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19), pp. 49–58 (2019)
45. Iyengar, S., Hahn, K.S.: Red media, blue media: evidence of ideological selectivity in media use. J. Commun. **59**(1), 19–39 (2009)
46. Jamieson, K.H., Cappella, J.N.: Echo chamber: Rush Limbaugh and the conservative media establishment. Oxford University Press, Oxford (2008)
47. Jannach, D., Jugovac, M.: Measuring the business value of recommender systems. ACM Trans. Manag. Inf. Syst. **10**(4), pp. 1–23 (2019)
48. Jannach, D., Lerche, L., Kamehkhosh, I., Jugovac, M.: What recommenders recommend: an analysis of recommendation biases and possible countermeasures. User Model. User Adapt. Interact. **25**(5), 427–491 (2015)
49. Jannach, D., Naveed, S., Jugovac, M.: User control in recommender systems: overview and interaction challenges. In: Proceedings 17th International Conference on Electronic Commerce and Web Technologies (EC-Web 2016) (2016)
50. Jugovac, M., Jannach, D., Lerche, L.: Efficient optimization of multiple recommendation quality factors according to individual user tendencies. Expert Syst. Appl. **81**, 321–331 (2017)
51. Kaminskas, M., Bridge, D.: Diversity, serendipity, novelty, and coverage: a survey and empirical analysis of beyond-accuracy objectives in recommender systems. ACM Trans. Interact. Intell. Syst. **7**(1), pp. 1–42 (2016)
52. Kamishima, T., Akaho, S., Sakuma, J.: Fairness-aware learning through regularization approach. In: Proceedings 11th IEEE International Conference on Data Mining Workshops, pp. 643–650 (2011)
53. Kirshenbaum, E., Forman, G., Dugan, M.: A live comparison of methods for personalized article recommendation at Forbes. com. In: Machine Learning and Knowledge Discovery in Databases, pp. 51–66 (2012)
54. Koenemann, J., Belkin, NJ.: A case for interaction: a study of interactive information retrieval behavior and effectiveness. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 205–212 (1996)
55. Lawrence, E., Sides, J., Farrell, H.: Self-segregation or deliberation? Blog readership, participation, and polarization in American politics. Perspect. Polit. **8**(1), 141–157 (2010)
56. Lazer, D.M., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D., et al.: The science of fake news. Science **359**(6380), 1094–1096 (2018)
57. van Leeuwen, C., Smets, A., Jacobs, A.: Blind spots in AI: the role of serendipity and equity in algorithm-based decision-making. ACM SIGKDD Explor. Newsl **23**(1), 42–49 (2021)
58. Li, S., Zhao, H.: A survey on representation learning for user modeling. In: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI '20), pp. 4997–5003 (2020)
59. Luo, C., Zhang, Y., Lin, W., Wang, Y., Yu, W.: An enhanced factorized model based on user and item features. In: IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pp. 1–5 (2014)
60. Mansoury, M., Abdollahpouri, H., Pechenizkiy, M., Mobasher, B., Burke,R.: Feedback loop and bias amplification in recommender systems. In: Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20), pp. 2145–2148 (2020)
61. Marlin, BM., Zemel, RS.: Collaborative prediction and ranking with non-random missing data. In: Proceedings of the Third ACM Conference on Recommender Systems (RecSys '09), pp. 5–12 (2009)
62. Marlin, BM., Zemel, RS., Roweis, S., Slaney, M.: Collaborative filtering and the missing at random assumption. In: Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence (UAI '07), pp. 267–275 (2007)
63. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A.: A survey on bias and fairness in machine learning. CoRR arXiv:1908.09635 (2019)
64. Mehrotra, R., McInerney, J., Bouchard, H., Lalmas, M., Diaz, F.: Towards a fair marketplace: counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM '18), pp. 2243–2251 (2018)
65. Meijer, A.: Understanding the complex dynamics of transparency. Public Adm. Rev. **73**(3), 429–439 (2013)
66. Min, Y., Jiang, T., Jin, C., Li, Q., Jin, X.: Endogenetic structure of filter bubble in social networks. R. Soc. Open Sci. **6**(11), 190868 (2019)
67. Mohseni, S., Ragan, E., Hu, X.: Open issues in combating fake news: interpretability as an opportunity. CoRR arXiv:1904.03016 (2019)
68. Möller, J., Trilling, D., Helberger, N., van Es, B.: Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. Inf. Commun. Soc. **21**(7), 959–977 (2018)
69. Morozov, E.: Swine flu: Twitter's power to misinform. https://www.npr.org/templates/story/story.php?storyId=103562240 (2009). Accessed 2 Jun 2021
70. Nagulendra, S., Vassileva, J.: Understanding and controlling the filter bubble through interactive visualization: a user study. In: Proceedings of the 25th ACM Conference on Hypertext and Social Media (HT '14), pp. 107–115 (2014)

71. Nguyen, TT., Hui, PM., Harper, FM., Terveen, L., Konstan, JA.: Exploring the filter bubble: the effect of using recommender systems on content diversity. In: Proceedings of the 23rd International Conference on World Wide Web, pp. 677–686 (2014)
72. NOU: Det norske mediemangfoldet - en styrket mediepolitikk for borgerne [media pluralism in Norway—a strengthened media policy for citizens]. The Ministry of Culture (2017)
73. Nunes, I., Jannach, D.: A systematic review and taxonomy of explanations in decision support and recommender systems. User Model. User Adapt. Interact. **27**(3–5), 393–444 (2017)
74. Oddleifson, E.: The effects of modern data analytics in electoral politics: Cambridge Analytica's Suppression of Voter Agency and the implications for global politics. Polit. Sci. Undergrad. Rev. **5**(1), 46–52 (2020)
75. Olteanu, A., Castillo, C., Diaz, F., Kıcıma, E.: Social data: Biases, methodological pitfalls, and ethical boundaries. Front. Big Data **2**, 13 (2019)
76. Paraschakis, D.: Recommender systems from an industrial and ethical perspective. In: Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16), pp. 463–466(2016)
77. Pariser, E.: The Filter Bubble: What the Internet Is Hiding from You. The Penguin Group, New York (2011)
78. Patro, G.K., Biswas, A., Ganguly, N., Gummadi, K.P., Chakraborty, A.: Fairrec: two-sided fairness for personalized recommendations in two-sided platforms. Proc. Web Conf. **2020**, 1194–1204 (2020)
79. Pedreshi, D., Ruggieri, S., Turini, F.: Discrimination-aware data mining. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '08), pp. 560–568 (2008)
80. Ribeiro, MT., Lacerda, A., Veloso, A., Ziviani, N.: Pareto-efficient hybridization for multi-objective recommender systems. In: Proceedings of the Sixth ACM Conference on Recommender systems, pp. 19–26(2012)
81. Ruchansky, N., Seo, S., Liu, Y.: CSI: A hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM '17), pp. 797–806 (2017)
82. Sasahara, K., Chen, W., Peng, H., Ciampaglia, GL., Flammini, A., Menczer, F.: On the inevitability of online echo chambers. CoRR arXiv:1905.03919 (2019)
83. Schelenz, L., Segal, A., Gal, K.: Best practices for transparency in machine generated personalization. In: Adjunct Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '20), pp. 23–28 (2020)
84. Schwär, H.: How Instagram and Facebook are intentionally designed to mimic addictive painkillers. https://www.businessinsider.com/facebook-has-been-deliberately-designed-to-mimic-addictive-painkillers-2018-12 (2021). Accessed 5 Jun 2021
85. Shakespeare, D., Porcaro, L., Gómez, E., Castillo, C.: Exploring artist gender bias in music recommendation. In: Proceedings of the Workshops on Recommendation in Complex Scenarios and the Impact of Recommender Systems ComplexRec-ImpactRS 2020 (2020)
86. Sinha, R., Swearingen, K.: The role of transparency in recommender systems. In: CHI '02 Extended Abstracts on Human Factors in Computing Systems, pp. 830–831 (2002)
87. Sonboli, N., Smith, JJ., Cabral Berenfus, F,. Burke, R., Fiesler, C.: Fairness and transparency in recommendation: The users' perspective. In: Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization UMAP '21, pp. 274–279 (2021)
88. de Souza Pereira Moreira, G., Jannach, D., da Cunha, A.M.: Contextual hybrid session-based news recommendation with recurrent neural networks. IEEE Access **7**, 169185–169203 (2019)
89. Sun, Y., Zhang, Y.: Conversational recommender system. In: Proceedings 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '18), pp. 235–244(2018)
90. Sürer Ö, Burke, R., Malthouse, EC.: Multistakeholder recommendation with provider constraints. In: Proceedings of the 12th ACM Conference on Recommender Systems (RecSys '18), pp. 54–62 (2018)
91. Tankovska, H.: Global Twitter user distribution by gender. https://www.statista.com/statistics/828092/distribution-of-users-on-twitter-worldwide-gender/ (2021). Accessed 2 Jun 2021
92. Tintarev, N., Masthoff, J.: A survey of explanations in recommender systems. In: 2007 IEEE 23rd International Conference on Data Engineering Workshop, pp. 801–810 (2007)
93. Trattner, C., Elsweiler, D.: Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems. In: Proceedings of the 26th International Conference on World Wide Web, pp. 489–498 (2017)
94. Ugander, J., Karrer, B., Backstrom, L., Marlow, C.: The anatomy of the facebook social graph. CoRR arXiv:1111.4503 (2011)
95. Wang, R., Harper, FM., Zhu, H.: Factors influencing perceived fairness in algorithmic decision-making: Algorithm outcomes, development procedures, and individual differences. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2020)
96. Wang, X., Wang, Y., Hsu, D., Wang, Y.: Exploration in interactive personalized music recommendation: a reinforcement learning approach. ACM Trans. Multimed. Comput. Commun. Appl. (TOMM) **11**, 1–22 (2014)
97. Wang, Y., Ning, Y., Liu, I., Zhang, XX.: Food discovery with Uber Eats: recommending for the marketplace. https://www.eng.uber.com/uber-eats-recommending-marketplace (2021). Accessed 1 Jun 2021
98. Woudstra, F.: What does transparent AI mean? AI policy exchange. https://www.aipolicyexchange.org/2020/05/09/what-does-transparent-ai-mean/ (2021). Accessed 1 Jun 2021
99. Yang, K., Stoyanovich, J.: Measuring fairness in ranked outputs. In: Proceedings of the 29th International Conference on Scientific and Statistical Database Management, pp. 1–6 (2017)
100. Yao, S., Huang, B.: Beyond parity: fairness objectives for collaborative filtering. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 2925–2934 (2017)
101. Zakon, A.: Optimized for addiction: Extending product liability concepts to defectively designed social media algorithms and overcoming the communications decency act. Wis. Law Rev. **5**, 1107 (2020)
102. Zhang YC, Séaghdha DÓ, Quercia, D., Jambor, T.:Auralist: introducing serendipity into music recommendation. In: Proceedings of the fifth ACM International Conference on Web Search and Data Mining (WSDM '12), pp. 13–22 (2012)
103. Zheng, H., Wang, D., Zhang, Q., Li, H., Yang, T.: Do clicks measure recommendation relevancy? An empirical user study. In: Proceedings of the fourth ACM Conference on Recommender Systems, pp. 249–252 (2010)
104. Ziegler, CN., McNee, SM., Konstan, JA., Lausen, G.: Improving recommendation lists through topic diversification. In: Proceedings of the 14th International Conference on World Wide Web, pp. 22–32 (2005)
105. Zuiderveen, FB., Trilling, D., Moeller, J., Bodó, B., de Vreese, CH., Helberger, N.: Should we worry about filter bubbles? Internet Policy Rev. 5(1), pp. 1–16 (2016)