

Chapter 2

Recommender Systems and Discrimination



Susanne Lilian Gössl

Abstract The following article deals with the topic of discrimination “by” a recommender system. Several reasons can create discriminating recommendations, especially the lack of diversity in training data, bias in training data or errors in the underlying modelling algorithm. The legal frame is still not sufficient to nudge developers or users to effectively avoid those discriminations, especially data protection law as enshrined in the EU General Data Protection Regulation (GDPR) is not feasible to fight discrimination. The same applies for the EU Unfair Competition Law, that at least contains first considerations to allow an autonomous decision of the subjects involved to know about possible forms of discrimination. Furthermore, with the Digital Service Act (DSA) and the AI Act (AIA) there are first steps into a direction that can inter alia tackle the problem. Most effectively seems a combination of regular monitoring and audit obligations and the development of an information model, supported by information by legal design, that allows an autonomous decision of all individuals using a recommender system.

Keywords Algorithmic discrimination · Fairness · Digital Services Act · AI Act proposal

The following article deals with the topic of discrimination “by” a recommender system that is based on incomplete or biased data or algorithms. After a short introduction (I.), I will describe the main reasons why discrimination by such a recommender system can happen (II.). Afterwards I will describe the current legal frame (III.) and conclude on how the future legal frame could look like and how the legal situation might be further improved (IV.).

S. L. Gössl (✉)
University of Bonn, Bonn, Germany
e-mail: sgoessl@uni-bonn.de

© The Author(s) 2023
S. Genovesi et al. (eds.), *Recommender Systems: Legal and Ethical Issues*, The International Library of Ethics, Law and Technology 40,
https://doi.org/10.1007/978-3-031-34804-4_2

2.1 Introduction

A recommender system gives recommendations based on an algorithm, often a machine learning algorithm (Projektgruppe Wirtschaft, Arbeit, Green IT 2013, 20). A machine learning algorithm basically works in the way that it takes a set of data and tries to find correlations between different data sets. If it finds enough correlations, it might derive a rule from those correlations. Based on the rule, the algorithm then makes a prediction about how a similar input might be handled in the future. Based on the prediction, the recommendation is made. For example, a machine learning algorithm that is supposed to classify cats, is trained on a certain number of pictures of cats and other animals. The algorithm then finds a correlation regarding the shape and size of ears, the tail, and whiskers. When novel pictures are used as inputs it checks for these features to conclude whether the picture shows a cat or not.

All these steps, be it the data set or data gathering, the finding of the correlations or, consequently, of the rules and predictions, can contain biases. As a result, the recommendation can contain those biases as well, which might lead to a discriminatory recommendation, e.g. that a machine gives a recommendation that is more favorable towards men than women or towards persons from a privileged social background than persons from another background (Alpaydın 2016, 16 ff.; Gerards and Xenidis 2021, 32 ff.; Kelleher 2019, 7 ff.; Kim and Routledge 2022, 75–102, 77 ff.; Vöneky 2020, 9–22, 21).

While some recommendations, e.g. the ranking of proposed items on a shopping website, (Wachter 2020, 367–430, 369 ff.) can be of a lesser fundamental rights relevance (Speicher et al. 2018), some recommender systems can be extremely relevant for the well-being of a person. E.g. a website can match employers and employees. If the website does not propose a possible employee for a job even though s/he would have been well-suited (Lambrecht and Tucker 2020, 2966–2981), that is not only a question of a bad-functioning algorithm but can touch the professional existence of the person left out (see, e.g., recital 71 of the Data Protection Directive – General Data Protection Regulation (GDPR 2016, 1)). Similarly, rankings of professionals (doctors, lawyers etc.) for somebody looking for the relevant service, are highly important as only the first few candidates have a chance to be chosen.¹

2.2 Reasons for Discriminating Recommendations

There are several reasons why a recommendation can be discriminating. They can basically be distinguished into three categories: The data set from which the machine learning algorithm is trained and adjusted can lack the relevant diversity (1), the training data can contain conscious or unconscious bias of the people creating the

¹ See e.g. with a focus on scoring (Gerberding and Wagner 2019, 116–119, 188).

data (2) and, finally, the underlying algorithm can be modelled in a way that it enhances discriminations (3) (von Ungern-Sternberg [forthcoming](#)).

2.2.1 Lack of Diversity in Training Data

The level of diversity in training data is paramount for the outcome of the concrete recommendation. One famous example where the lack of diversity lead to discrimination of women, was the Amazon hiring tool (Gershgorn [2018](#)): The hiring tool was supposed to make objective predictions of the quality and suitability of applying job candidates. Problematic was that the algorithm was “fed” by application data of the last decade – which included a significant higher proportion of male (and probably white) candidates. The training data, therefore, lacked diversity regarding women. As a consequence, the hiring tool “concluded” that women were less qualified for the job, resulting in discriminatory recommendations (Gershgorn [2018](#)). Similarly, whenever training data is only taken from reality and not created artificially, there is a high probability that it will lack diversity – especially in jobs that have typically a higher number of men (as STEM areas - Science, Technology, Engineering, and Mathematics (Wikipedia [2022](#))) or women (as care and social work) or lack – so far – People of Colour (PoC) or candidates with an immigration, LGBTIAQ* or disability background, as in these jobs the representation of these groups might be extraordinarily higher or lower than the one of other groups (Reiners [2021](#); Sheltzer and Smith [2014](#), 10107–10112). The effect of missing diversity in training data was also shown in face recognition software using machine learning algorithms: Face recognition software that was trained mainly with photos from white and male people, afterwards had stronger problems to identify black or female and especially female black persons (Buolamwini and Gebru [2018](#)).

2.2.2 (Unconscious) Bias in Training Data

The second, very influential reason why recommender systems often show discriminating results is the fact that the training data very often contains data from real life people and therefore, also reflects their conscious or unconscious bias. For example, there has been a study of the University of Bonn regarding “Gender Differences in Financial Advice” (Bucher-Koenen et al. [2021](#); Cavalluzzo and Cavalluzzo [1998](#), 771–792) analysing the recommendations financial advisors gave to different people looking for advice. The study shows that usually the recommendations women receive are more expensive than those of male candidates. There are several explanations, e.g. the fact that women very often are more risk adverse, resulting in more expensive but also safer investments. Another possible reason was that men often look for advice to get a “second opinion”, while women do not consult other

advisors and lack the information male candidates may already have (Bucher-Koenen et al. 2021, 12 ff., 14 ff.). An algorithm that “learns” from this data might conclude that women always should get the more expensive recommendations without looking at the concrete woman applying. The whole problem can be enhanced by data labelling practises. Training data usually gets labelled as “correct” or “incorrect” (or “good” or “bad”) to enable the learning process of the algorithm. Whenever the decision whether a résumé or a person’s performance is “good” is not only based on the hiring decision, but furthermore, a separate person labels it as “good” or “bad”, the labelling decision can contain an additional (unconscious) bias of the labelling person (Calders and Žliobaitė 2013, 48 ff.). For example, there are algorithms that recommend professionals or professional services, often based on users’ recommendations. Very often a ranking is made with those receiving the highest recommendation coming first (thus having the label “good”).² This can discriminate e.g. women or members of minority groups: There is research that people rate members of these groups or women typically less favourable than a man not belonging to a minority group even though the performance is the same. Research shows, e.g., that equal résumés with a male or female name on it are evaluated differently, usually the female one less favourable (by male and female evaluators equally) (Moss-Racusin et al. 2012, 16474–16479; Handley et al. 2015, 13201–13206). The same applies to teaching materials in law schools (Özgümüs et al. 2020, 1074).

If a machine learning algorithm, thus, ranks the recommendation of professionals or professional services based on these user evaluations, the probability is high that these (unconscious) biases that led to a less favorable rating in the first place, will also lead to a lower ranking in the recommendation – with negative influence e.g., on the income and career of the professional. For instance, a study looking at the company Uber that based the ranking of its drivers on consumers’ ratings, shows these biases clearly (Rosenblat et al. 2017, 256–279).

2.2.3 *Modelling Algorithm*

Finally, the algorithm can be modelled in a way that it enhances biases already contained in the training data. One reason can of course be the selection of the relevant features the algorithm uses for the selection – e.g., if a personalized ad algorithm filters ads only according to the gender of the user, the result might be that women always receive recommendations for sexy dresses and make-up while men might always receive recommendations for adventure trips, barbeque and home building tools (Ali et al. 2019, 1–30).

Flaws in the modelling can also have a tremendous impact depending on the area where they are used. Another example of discriminating results with regard to the programming of the algorithm could be found in the famous COMPAS program

²E.g. the page <https://www.jameda.de/> for physicians in Germany.

used by several US-States (Angwin et al. 2016; Flores et al. 2016, 38–46; Martini 2019, 2 ff.). This program aimed at recommendations regarding the re-offense probability of criminal offenders. A high risk for future crime would lead to a less favorable treatment in detention – e.g. a higher bail or an exclusion of the possibility to be bailed out. The program reflected the unconscious bias the judges had towards Afro-American candidates, assuming their re-offense risk would be higher and towards Caucasian offenders, assuming their re-offense risk would be lower. These two flaws from the real world could have been at least mitigated by a calibration of the algorithm allocating different error rates to different groups within the training data, making the algorithm “learn” to avoid the same bias. This is important whenever different groups have different base rates, i.e., rates of positive or negative outcomes.

Therefore, one problem in the modelling algorithm was that the allocation of error rates analyzing the existing data was equal towards both groups, even though it should have included the fact that a Caucasian person had two biases in his/her favor (no assumption of higher re-offence risk and assumption of lower re-offence risk) while the Afro-American person had only one against him (the assumption of a higher re-offence risk), thus, different base rates. The probability that the outcome regarding of an Afro-American person would be a false (and negative) prediction, therefore, was higher. This should have been reflected in the error rate.

Therefore, an equal allocation of error rates even enhanced the biases already contained in the training data (Chouldechova 2017, 153–163; Rahman 2020; Barocas et al. 2018, 23, 31, 68). Problematic, on the other hand, is that different error rates assume that there are differences in groups, thus making a distinction even though a distinction was supposed to be avoided (Barocas et al. 2018, 47 ff.).

2.2.4 *Interim Conclusion and Thoughts*

Recommender systems can discriminate as they can reinforce and deepen stereotypes and biases already found in our society. Several problems can lead or enhance those outcomes: First, the data used from experience, is always from the past, thus reflecting biases and difficulties from the past. Thus, the person selecting the training data must always have in mind that the past has no perfect data to reflect the diversity of our society. So-called “counterfactuals” that have to be created artificially can help to avoid the lack of diversity (Cofone 2019, 1389–1443; Mothilal et al. 2020; Oosterhuis and de Rijke 2020).³ Counterfactuals refer to artificially created data sets that can counterbalance the aforementioned lack of diversity in data stemming from reality – e.g., the aforementioned lack of female résumés in the STEM areas can be counterbalanced by introducing artificially created female résumés.⁴

³ See also the solution proposed by (Blass 2019, 415–468).

⁴ Regarding the use of counterfactuals and its risks e.g. see (Kasirzadeh and Smart 2021, 228 ff.).

Second, while it is easy to avoid differentiation features that are obviously discriminatory, such as “race” or “gender”, the compilation of data can have similar effects as such direct discriminating features (Ali et al. 2019, 1–30; Buolamwini and Gebru 2018, 12). For example, the postal code of a person in many countries is highly correlated with ethnicity or social background, thus, if an algorithm “learns” that résumés from a certain area are usually “bad”, this indirectly leads to discrimination based on the social or ethical background (Calders and Žliobaitė 2013, 4–49).

Third, because of these effects caused by certain data compilations that are difficult to predict *ex ante*, especially if the algorithm is self-learning, it is also difficult to predict under which circumstances discrimination will be caused by which reason. This unpredictability makes it necessary to monitor and adjust such algorithms on a regular basis.

2.3 Legal Frame

So far, there is no coherent legal frame to tackle discrimination by recommender systems. Nevertheless, certain approaches can be derived from the existing legal frame: Existing solutions are either based on agreement (1.), information (2.), or a combination of both approaches. Finally, the general rules of anti-discrimination law apply (3.).

2.3.1 Agreement – Data Protection Law

The first approach that is based on user agreement can be found in data protection law, especially Article 22 para. 1 GDPR (2016, 1). According to that rule, “[t]he data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.”

Recital 71 gives more specifications on the Article and makes clear that the controller of the algorithm should “prevent, *inter alia*, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or processing that results in measures having such an effect.”

While this rule at first glance sounds like a clear prohibition to create recommendations based on exclusively algorithmic decisions, there are several problems in the application of the rule that make it questionable whether it is sufficient to resolve the problem. First, one can question in general whether data protection law is the proper venue to prevent discriminatory results. Data protection law primarily is intended to protect the personal data of natural persons and to give them control on how this data is used. It aims at the protection of the personality rights of such a person. Anti-discrimination law, on the other hand, tackles certain inequalities that

exist in society, and protects the individual from discrimination – independently of the data used or concerned. While, of course, discrimination can also lead to the infringement of a personality right, the protective function is a different one.

Furthermore, literature disagrees under which circumstances there is a “decision” in the sense of Article 22 GDPR regarding recommender systems. While recital 71 clarifies that such a “decision” is the case when we have a refusal, e.g., “of an online credit application or a recruiting practice without any human intervention”, the case becomes less clear when the algorithm only proposes a certain job opportunity (or not) (Lambrecht and Tucker 2020, 2966–2981) or ranking that afterwards will be subjected to the decision of a person. While some voices regard such a preliminary recommendation as excluded from the scope of Article 22 DGPR (German government 2000, 37; EU Commission 2012, 26 et seq.; Martini 2019, 173; see also OLG Frankfurt/M. 2015, 137), others limit the notion of decision to the exclusion of a person (from e.g. a ranking).⁵

Nevertheless, even if we apply Article 22 to all recommendations, a justification is possible if the controller uses the algorithm, *inter alia*, “to ensure the security and reliability of a service provided by the controller, or necessary for the entering or performance of a contract between the data subject and a controller, or when the data subject has given his or her explicit consent” (Recital 71, also Article 22 para. 2). Para. 3 then introduces some procedural safeguards for the protection of the personality rights of the person concerned. Nevertheless, the basic rule is that whenever the data subject has given the explicit consent for the processing of the data, the infringement within the meaning of Article 22 para. 1 is justified under Article 22 para. 2 GDPR (Vöneky 2020, 9–22, 13; Martini 2019, 171 ff.). This is problematic as research shows that the majority of internet users are willing to give their consent to proceed on a website without really dealing with the content of the agreement (Carolan 2016, 462–473; in detail see also Machuletz and Böhme 2020, 481–498). If an agreement is easily given without a conscious choice, Article 22 GDPR does not provide a very stable protection against discriminatory results.

2.3.2 Information – Unfair Competition Law

The second approach can be called an information-centered approach. The main measure consists in giving information to the user about the available ranking parameters and the reasons for the relative importance of certain parameters to others. We can see that approach on the Business-to-Business (B2B) level in Article 5 P2B Regulation (Regulation (EU) 2019/1150. 2019, 57 ff.) regarding online providers and businesses using their platforms. A similar rule has also been introduced into the UCP Directive (Directive 2005/29/EC 2005, 22) regarding the

⁵E.g. (von Lewinski 2021, para. 16, unclear at 16.1). To the whole discussion see von Ungern-Sternberg, Discriminatory AI and the Law: Legal Standards for Algorithmic Profiling. In *Responsible AI*, ed. Silja Vöneky et al., forthcoming. II. 2. b).

Business-to-Consumer (B2C) level in its 2019 amendment (Article 3 Nr. 4 lit. b) (Directive (EU) 2019/2161 2019, 7 ff.). Article 7 para. 4a of the UCP Directive provides that whenever a consumer can search for products offered by different traders or by general consumers “information [...] on the main parameters determining the ranking of products presented to the consumer as a result of the search query and the relative importance of those parameters, as opposed to other parameters, shall be regarded as material,” meaning that this information has to be part of the general information obligations towards the consumer. The effectiveness of these measures to combat discriminatory recommendations is doubtful.

First, both rules only contain information obligations, meaning that the effectiveness mainly depends on the attention of the user and his or her willingness to read the information, understand what the “relative importance of certain parameters” means for his or her concrete use of the platform and act upon that knowledge. Even if a trader or intermediary indirectly gives the information that the recommendation can be discriminatory, in most cases the platform or search possibility will most probably still be used as the majority of users will not notice it (Martini 2019, 188; Bergam et al. 2020). Furthermore, the information necessary to understand the logic of a discriminatory recommender system might not be part of the information that is part of the information obligation. The limit will most probably lie behind the protection of trade secrets of the provider of the algorithm – including the algorithm or at least some of its features. So, discrimination caused by a certain algorithm model will probably stay undetected despite the information obligation.

2.3.3 *General Anti-discrimination Law*

Specific rules regarding recommender systems or algorithms do not seem sufficient to tackle discriminatory recommendations. Nevertheless, they are not exhaustive in that area – also the general anti-discrimination rules apply and might sufficiently prevent discriminatory recommendations.

These rules, usually, on the national or EU level, e.g., forbid an unjustified unequal treatment according to certain personal features such as gender, race, disability, sexual orientation, age, social origin, nationality, faith or political opinion (list not exhaustive, depending on country or entity) (TFEU (EU) 2007, Art. 19; CFR (EU) 2012, Art. 21; Fundamental Law (Ger) 1949, Art. 3 para. 3; AGG(Ger) 2006 sec. 1). While many important features, therefore, are included, there is no general prohibition to treat people differently, e.g., for the region they live in or the dialect they speak or the color of their hair (Martini 2019, 238; Wachter [forthcoming](#)). Of course, those features can accumulate to features protected by anti-discrimination law, e.g., the region and the dialect of a person can allow conclusions regarding the ethical or social background (see above, Sect. 2.4.). But the general rule remains that discrimination is allowed as long as an explicitly mentioned feature is not the reason.

Applying anti-discrimination rules to the relationship between the provider of a recommender system and a user raises some further issues. First, those anti-discrimination rules primarily were drafted to protect the citizens against the State. If a public agency, for instance, uses a recommender system as a recruiting tool, anti-discrimination law applies directly.⁶ On the other hand, the effect of these rules in private legal relationships, where the majority of recommender systems is used, is less easy to establish and highly disputed (Knebel 2018, 33 ff.; Perner 2013, 143 ff.; Schaaf 2021, 249; Neuner 2020, 1851–1855). Additionally, recommender systems are often used without the conclusion of a contract, thus, they move in the pre-contractual area where the parties' responsibility is traditionally harder to establish. Nevertheless, a tendency can be observed that the prohibition of discrimination slowly creeps into private relationships, especially contract law and employment law, at least in the EU (AGG (Ger) 2006, sec. 2, 7 para. 2, 21 para. 4; Hellgardt 2018, 901; Perner 2013, 145 ff.). Several EU anti-discrimination directives (Directive 2000/43/EC 2000, 22; Directive 2000/78/EC 2000, 16; Directive 2002/73/EC 2002, 15; Directive 2004/113/EC 2004, 37) as well as a constant flow of case law from the CJEU have enhanced this process and extended it to the pre-contractual level as well (CJEU 1976 Defrenne/SABENA, para 39; CJEU 2011 Test-Achats; Perner 2013, 157 ff.; Grüneberger and Reinelt 2020, 19 ff.). However, whether and to whom a provider of a recommender system is responsible if the recommender system is discriminatory, is unclear.⁷

Furthermore, there is the problem of indirect discrimination. As mentioned above, it will be easy to detect discrimination if the modelling algorithm uses a forbidden differentiation criterion. Nevertheless, a combination of other, not directly forbidden criteria, can lead to the same result (Ali et al. 2019, 1–30; Buolamwini and Gebru 2018, 12). Recruiting tools, for example, have often regarded résumés with longer periods without gainful employment as a sign of a weaker working performance. However, these periods can also be caused by breaks such as parental leaves or additional care obligations, typically involving more women than men. Thus, differentiating regarding that criterion, in consequence, can lead to the discrimination of women.

In anti-discrimination law it has been recognized that indirect discrimination can be forbidden as well (see Sect. 3 para. 2 AGG). The difference can become relevant for the requirements for the justification of unequal treatment. Unequal treatment can be justified if there are equally weighing values or interests on the other side to makeup the differentiation. This leads to a balancing of interests and risks of the people involved. Usually, direct discrimination weights more heavily and is almost impossible to justify, compared to an indirect one is (von Ungern-Sternberg forthcoming). Of course, the result also depends on the area of life where the

⁶E.g. Public Job Services, (see e.g. Allhutter et al. 2020).

⁷See, e.g. to the application of German Anti-discrimination law in the context of insurance recommendations Martini 2019, 234; see also to the problem of the scope of application of anti-discrimination law Hacker 2021 at fn. 88 to 98.

recommender system is used in. Thus, personalized ads are not as risky and relevant for the person involved as, for example, a job proposal or the exclusion of a job proposal.

Finally, the chain of responsibilities can be difficult. Often the recommender system is used by a platform but programmed by another business while the contract in question will be concluded between a user of the platform (e.g., an employer) and another user (e.g., the job seeker). Anti-discrimination law usually only has effects between the latter two, meaning that afterwards the possible employer must seek compensation from the platform provider who, in return, can seek compensation by the programmer. To ensure that the person of business finally responsible for the discriminatory algorithm is really forced to compensate the other parties, and, consequently, has an incentive to change the algorithm, is difficult in this way. Additionally, a justification might be possible if the functioning of the algorithm was not predictable to him or her as especially a self-learning algorithm is difficult to control regarding the data input and the improvement of the algorithm (black box problem).

2.3.4 Interim Conclusion

The legal frame only partly deals with discrimination by algorithms and is not sufficient to efficiently tackle it. Furthermore, the existing anti-discrimination law bears several uncertainties for all the parties involved.

2.4 Outlook

From these first conclusions, the next question is what should be done.

2.4.1 Extreme Solutions

One extreme possibility could be the prohibition to use machine learning algorithms in recommender systems at all. This would, of course, stop discriminations by recommendations, but also impede any progress regarding the use of machine learning algorithms or the development of recommender systems.

The other extreme solution could be a hands-off-approach and to leave it to the market powers to regulate the use of recommender systems. This approach also does not seem feasible as the past has shown that the mere play of market powers has been unable to prevent discrimination.⁸

⁸ See e.g. regarding gender discrimination (Ekin 2018).

2.4.2 Further Development of the Information Approach

One possible solution between those two extreme positions could be a further development of the already existing information approach (Martini 2019, 187). Providers of recommender systems should also provide the necessary information for users to foresee and understand the risks of discrimination by a certain system in combination with an opt-out or opt-in possibility, meaning that they should not only have the choice to use the system or not, but also to use the system with the possible discrimination but also with alternatives. Furthermore, providers should be obliged to use a legal design that ensures that the people involved really read and understand the information (Martini 2019, 189; Kim and Routledge 2022, 97 ff.).⁹

This approach has also been chosen by a recent EU regulation, the Digital Service Act (DSA 2022/2065 (EU)). Article 27 para. 1 DSA states an explicit obligation for recommender systems provided by “online platforms” (not including “micro and small enterprises”, Art. 19 DSA) to “set out in their terms and conditions, in plain and intelligible language, the main parameters used in their recommender systems, as well as any options for the recipients of the service to modify or influence those main parameters”. Furthermore, according to Article 38 DSA, providers of very large online platforms that use recommender systems “shall provide at least one option for each of their recommender systems which is not based on profiling”. Moreover, another proposed EU Act, the Artificial Intelligence Act (AIA 2021 (EU)), foresees that “AI” must be transparent and explainable for the user in Article 13 of the Commission Proposal (Kalbhenn 2021, 668).

This approach, in general, is a good step in the right direction. However, it has two flaws. First, Article 38 DSA only address “very large online platforms”, platforms with more than 45 million recipients each month and designated as such by the Commission (Article 33 para. 1, 4 DSA). Recommender systems can, nevertheless, also be used in certain niche areas and be of high importance for the live of the parties involved, e.g., in certain job branches where highly specific people are recruited or searched. The AIA does not have this restriction. Besides, Article 38 only provides an “opt-out”, meaning that users actively must choose not to use the proposed algorithm. The AIA does not provide any comparable consequences. Studies show that most users do not read the information but only continue to click to progress with the process they visited a certain platform for (Bergram et al. 2020; Martini 2019, 188). An opt-out possibility, therefore, is less efficient than an opt-in and nudges the users to just use what is already provided.

⁹Regarding the importance of design see e.g. (Machuletz and Böhme 2020, 481–498); see also the proposal to introduce counterfactual explanations as a complete information by (Wachter et al. 2018, 841–887).

2.4.3 *Monitoring and Audit Obligations*

The DSA also provides another interesting feature to control very large online platforms by establishing an obligation for regular audits (Article 37 DSA) to ensure that certain standards are met (Kalbhenn 2021, 671). Unfortunately, the audit obligation does not include recommender systems and possible discriminatory outcomes as mentioned in Articles 27, 38 DSA. An audit obligation, however, could be extended to possible discriminations, especially in areas where such discrimination can have massive effects on the life of the person involved, e.g. in questions of employment or job evaluation (Buolamwini and Gebru 2018, 12).

Therefore, it is no coincidence that another proposal for an EU Act, the Artificial Intelligence Act (AIA), also establishes an audit obligation for AI that is used in “high risk” areas, referring to areas that bear a high risk for the involved subjects. Contrary to the DSA, it applies no matter how many users a platform or provider has.

A similar approach can also be seen in other countries: The “Automated Employment Decision Tools” Bill by New York City (Law No. 2021/144 (Int 1894–2020) (NYC)) only allows the use of algorithms in employment decisions if the algorithm is subjected to a yearly audit. The advantage of such an audit is that the algorithm can be analyzed by specialists and nudge businesses to improve them (Raji and Buolamwini 2019). On the other hand, businesses only have to hand over trade sensitive information to those auditors, thus, their trade secrets can be respected and protected as well.

2.4.4 *Interim Conclusion and Thoughts*

To conclude, both (proposed) approaches of DSA and AIA, information/transparency and a regular audit obligation, should be combined for the use of recommender systems, at least in highly risky/sensitive areas for the person involved. An information obligation together with an opt-in possibility (rather than the opt-out-option provided in the DSA) and not limited to “very large platforms” would be feasible in those areas. Furthermore, a regular audit should be obligatory to ensure that possible discriminations in recommender systems can be found by the auditors and countered by them or others.

2.5 **Conclusions**

1. Recommender Systems based on algorithms can cause discrimination.
2. The existing legal framework is not sufficient to combat those discriminations. It is limited to certain information obligations and general non-discrimination rules that cannot provide the necessary legal certainty.

3. Information about the consequences of using a certain recommender system should be available for the people involved and phrased in a way that the users can understand it. Also, similar to Articles 27 para. 1, 38 DSA, at least an “opt-out” possibility should be provided, even though an opt-in possibility would be preferable.
4. A regular audit should be required, at least in areas that are highly sensitive to discrimination. This audit would allow the analysis by experts to find the reasons for discriminatory recommendations without endangering the trade secrets of the provider of the algorithm.

References

- Act on Equal Treatment (*Allgemeines Gleichbehandlungsgesetz – AGG*). 2006. Available online: <https://www.gesetze-im-internet.de/agg/index.html>. Accessed on 06.09.2022.
- Ali, Muhammad, Sapiezynski, Piotr, Bogen, Miranda, Korolova, Aleksandra, Mislove, Alan, and Rieke, Aaron. 2019. Discrimination Through Optimization. In Proceedings of the ACM on Human-Computer Interaction 3, CSCW 2019, 1–30.
- Allhutter, Doris, Mager, Astrid, Cech, Florian, Fischer, Fabian, and Grill, Gabriel. 2020. Der AMS-Algorithmus: Eine Soziotechnische Analyse Des Arbeitsmarktchancen-Assistenz-Systems (AMAS). ITA-Projektbericht Nr.: 2020-02 2020.
- Alpaydm, Ethem. 2016. *Machine Learning: The New AI. The MIT Press Essential Knowledge Series*. Cambridge, MA/London: MIT Press.
- Angwin, Julia, Larson, Jeff, Mattu, Surya, and Lauren Kirchner. 2016. Machine Bias. *ProPublica*, May 23, Available online: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Accessed on 07.09.2022.
- Automated Employment Decision Tools Bill by New York City Law No. 2021/144 (Int 1894-2020) 2021 (New York City). Available online: <https://www.assembly.ny.gov/leg/?bn=A07244&term=&Summary=Y&Actions=Y&Votes=Y&Memo=Y&Text=Y>. Accessed on 06.09.2022.
- Barocas, Solon, Hardt, Moritz, and Narayanan, Arvind. 2018. Fairness and Machine Learning: Limitations and Opportunities. 2018 (last update 2022). Online book available at <https://fairml-book.org/pdf/fairmlbook.pdf>. Accessed on 07.09.2022.
- Bergram, Kristoffer, Bezençon, Valéry, Maingot, Paul, Gjerlufsen, Tony, and Holzer, Adrian. 2020. Digital Nudges for Privacy Awareness: From Consent to Informed Consent?. In Proceedings of the 28th European Conference on Information Systems (ECIS), An Online AIS Conference, June 15–17, 2020. Available online https://aisel.aisnet.org/ecis2020_rp/64. Accessed on 06.09.2022.
- Blass, Joseph. 2019. Algorithmic Advertising Discrimination. *Northwestern University Law Review* 114 (2): 415–468.
- Bucher-Koenen, Tabea, Hackethal, Andreas, Koenen, Johannes, and Laudenbach, Christine. 2021. Gender Differences in Financial Advice. ECONtribute Discussion Paper no. 095 2021.
- Buolamwini, Joy, and Gebru, Timnit. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In Proceedings of Machine Learning Research, vol. 81, issue 1.
- Calders, Toon, and Indrè Žliobaitė. 2013. Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures. In *Discrimination and Privacy in the Information Society*, ed. Bart Custers et al., vol. 3, 43–57. Berlin/Heidelberg: Springer.
- Carolan, Eoin. 2016. The Continuing Problems with Online Consent Under the EU’s Emerging Data Protection Principles. *Computer Law & Security Review* 32 (3): 462–473.

- Cavalluzzo, Ken S., and Cavalluzzo, Linda C. 1998. Market Structure and Discrimination: The Case of Small Businesses. *Journal of Money, Credit and Banking* 30(4): 771–792. Available online: <https://EconPapers.repec.org/RePEc:mcb:jmoncb:v:30:y:1998:i:4:p:771-92>. Accessed on 06.09.2022.
- Chouldechova, Alexandra. 2017. Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. *Big Data* 5 (2): 153–163.
- Court of Justice of the European Union (CJEU). Defrenne/SABENA. ECLI:EU:C:1976:56.
- . Test-Achats. ECLI:EU:C:2011:100.
- Cofone, Ignacio N. 2019. Algorithmic Discrimination is an Information Problem. *Hastings Law Journal* 70: 1389–1443.
- Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin, OJ L 180, 19.7.2000.
- Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation, OJ L 303, 2.12.2000.
- Directive 2002/73/EC of 23 September 2002 Amending Council Directive 76/207/EEC on the implementation of the principle of equal treatment for men and women as regards access to employment, vocational training and promotion, and working conditions, OJ L 269, 5.10.2002.
- Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services, OJ L 373, 21.12.2004.
- Directive 2005/29/EC of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/6/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive'), OJ L 149, 11.6.2005: 22.
- Directive (EU) 2019/2161 of 27 November 2019 amending Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU as regards the better enforcement and modernisation of Union consumer protection rules, OJ L 328, 18.12.2019.
- Ekin, Annette. 2018. Quotas get more women on boards and stir change from within. *Horizon*. The EU Research and Innovation Magazine. Available online: <https://ec.europa.eu/research-and-innovation/en/horizon-magazine/quotas-get-more-women-boards-and-stir-change-within>.
- EU Commission. 2012. EU Commission regarding the predecessor rule Article 15 DPD, COM(92) 422 final – SYN 287.
- Flores, Anthony W., Kristin Bechtel, and Christopher T. Lowenkamp. 2016. False Positives, False Negatives, and False Analyses: A Rejoinder to Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks. *Federal Probation* 80 (2): 38–46.
- Fundamental Law (*Grundgesetz* – German Constitution) 1949 (Germany). Available online: <https://www.gesetze-im-internet.de/gg/index.html>. Accessed on 06.09.2022.
- Gerards, Janneke, and Raphaële Xenidis. 2021. *Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non-Discrimination Law: A Special Report*. Luxembourg: Publications Office of the European Union.
- Gerberding, Johannes, and Gert G. Wagner. 2019. Qualitätssicherung Für "Predictive Analytics" Durch Digitale Algorithmen. *Zeitschrift für Rechtspolitik* 2019: 116–119.
- German Government. 2000. Gesetzentwurf der Bundesregierung. Entwurf eines Gesetzes zur Änderung des Bundesdatenschutzgesetzes und anderer Gesetze. Bundestags-Drucksache 14/4329. Available online: <https://dsserver.bundestag.de/btd/14/043/1404329.pdf>. Accessed on 06.09.2022.
- Gershgorn, Dave. 2018. Amazons "holy grail" recruiting tool was actually just biased against women. *Quartz*, October 10, Available online: <https://qz.com/1419228/amazons-ai-powered-recruiting-tool-was-biased-against-women>. Accessed on 07.09.2022.

- Grünberger, Michael, and André Reinelt. 2020. *Konfliktlinien im Nichtdiskriminierungsrecht: Das Rechtsdurchsetzungsregime aus Sicht soziologischer Jurisprudenz*, 2020. Tübingen: Mohr Siebeck.
- Hacker, Philipp. 2021. A Legal Framework for AI Training Data – From First Principles to the Artificial Intelligence Act. *Law, Innovation and Technology* 13 (2): 257–301.
- Handley, Ian M., Elizabeth R. Brown, Corinne A. Moss-Racusin, and Jessi L. Smith. 2015. Quality of Evidence Revealing Subtle Gender Biases in Science is in the Eye of the Beholder. *Proceedings of the National Academy of Sciences of the United States of America* 112 (43): 13201–13206.
- Hellgardt, Alexander. 2018. Wer Hat Angst Vor Der Unmittelbaren Drittwirkung? *Juristen Zeitung* 73 (19): 901.
- Kalbhenn, Jan C. 2021. Designvorgaben Für Chatbots, Deepfakes Und Emotionserkennungssysteme: Der Vorschlag Der Europäischen Kommission Zu Einer KI-VO Als Erweiterung Der Medienrechtlichen Plattformregulierung. *Zeitschrift für Urheber- und Medienrecht* 2021: 663–674.
- Kasirzadeh, A., and A. Smart. 2021. The Use and Misuse of Counterfactuals in Ethical Machine Learning. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 228–236. <https://doi.org/10.1145/3442188.3445886>
- Kelleher, John D. 2019. *Deep Learning. The MIT Press Essential Knowledge Series*. Cambridge, MA/London: MIT Press.
- Kim, Tae W., and Bryan R. Routledge. 2022. Why a Right to an Explanation of Algorithmic Decision-Making Should Exist: A Trust-Based Approach. *Business Ethics Quarterly* 32 (1): 75–102.
- Knebel, Sophie V. 2018. *Die Drittwirkung Der Grundrechte Und -Freiheiten Gegenüber Privaten*, 2018. Baden-Baden: Nomos.
- Lambrecht, Anja, and Catherine Tucker. 2020. Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *Management Science* 65 (7): 2966–2981.
- Machuletz, Dominique, and Böhme, Rainer. 2020. Multiple Purposes, Multiple Problems: A User Study of Consent Dialogs After GDPR. In *Proceedings on Privacy Enhancing Technologies*, 2, 481–498. Available online: <http://arxiv.org/pdf/1908.10048v2>. Accessed on 06.09.2022.
- Martini, Mario. 2019. *Blackbox Algorithmus – Grundfragen Einer Regulierung Künstlicher Intelligenz*. Berlin/Heidelberg: Springer.
- Moss-Racusin, Corinne A., John F. Dovidio, Victoria L. Brescoll, Mark J. Graham, and Jo Handelsman. 2012. Science Faculty’s Subtle Gender Biases Favor Male Students. *Proceedings of the National Academy of Sciences of the United States of America* 109 (41): 16474–16479.
- Mothilal, Ramaravind K., Amit Sharma, and Chenhao Tan. 2020. Explaining Machine Learning Classifiers Through Diverse Counterfactual Explanations. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, ed. Mireille Hildebrandt et al., 607–617. New York: ACM.
- Neuner, Jörg. 2020. Das BVerfG Im Labyrinth Der Drittwirkung. *Neue Juristische Wochenschrift* 26: 1851–1855.
- OLG Frankfurt/M. 2015. ZD.
- Oosterhuis, Harrie, and Maarten de Rijke. 2020. Taking the Counterfactual Online: Efficient and Unbiased Online Evaluation for Ranking. In *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*, ed. Krisztian Balog et al., 137–144. New York: ACM.
- Özgümüş, Asri, Holger A. Rau, Stefan T. Trautmann, and Christian König-Kersting. 2020. Gender Bias in the Evaluation of Teaching Materials. *Frontiers in Psychology* 11: 1074.
- Perner, Stefan. 2013. Grundfreiheiten, Grundrechte-Charta und Privatrecht. Beiträge zum ausländischen und internationalen Privatrecht 98. Tübingen: Mohr Siebeck.

- Projektgruppe Wirtschaft, Arbeit, Green IT der Enquete-Kommission Internet und digitale Gesellschaft. 2013. Achter Zwischenbericht der Enquete-Kommission “Internet und digitale Gesellschaft” Wirtschaft, Arbeit, Green IT. Bundestagsdrucksache 17/12505. Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, COM(2020) 825 final; consolidated text as adopted by the European Parliament P9_TC1-COD(2020)036. Available online: https://eur-lex.europa.eu/procedure/EN/2020_361. Accessed on 06.09.2022.
- Rahman, Farhan. 2020. COMPAS Case Study: Fairness of a Machine Learning Model. *Towards Data Science*, September 7. Available online: <https://towardsdatascience.com/compas-case-study-fairness-of-a-machine-learning-model-f0f804108751>. Accessed on 31.05.2022
- Raji, Inioluwa D., and Buolamwini, Joy. 2019. Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. AIES-119 Paper N0. 223. Available online: https://dam-prod.media.mit.edu/x/2019/01/24/AIES-19_paper_223.pdf. Accessed on 06.09.2022.
- Regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, General Data Protection Regulation (GDPR) (EU) 2016, OJ L 119, 4 May 2016.
- Regulation (EU) 2019/1150 of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services, OJ L 186, 11.7.2019 OJ L 186, 11.7.2019.
- Reiners, Bailey. 2021. 57 Diversity in the Workplace Statistics You Should Know, *BuiltIn*, October 20. Available online <https://builtin.com/diversity-inclusion/diversity-in-the-workplace-statistics>. Accessed 07.09.2022.
- Rosenblat, Alex; Levy, Karen E. C.; Barocas, Solon; Hwang, Tim. 2017. Discriminating Tastes: Uber’s Customer Ratings as Vehicles for Workplace Discrimination. *Policy & Internet* 9 (3): 256–279.
- Schaaf, Henning. 2021. Drittwirkung Der Grundrechte – Dogmatik Und Fallbearbeitung. *JURA – Juristische Ausbildung* 43 (3): 249–257.
- Sheltzer, Jason M., and Joan C. Smith. 2014. Elite Male Faculty in the Life Sciences Employ Fewer Women. *Proceedings of the National Academy of Sciences of the United States of America* 111 (28): 10107–10112.
- Speicher, Till, Muhammad Ali, Giridhari Venkatadri, Filipe N. Ribeiro, George Arvanitakis, Fabricio Benevenuto, Krishna P. Gummadi, Patrick Loiseau, and Alan Mislove. 2018. Potential for Discrimination in Online Targeted Advertising. *Proceedings of Machine Learning Research* 81 (1).
- von Ungern-Sternberg, Antje. forthcoming. Discriminatory AI and the Law: Legal Standards for Algorithmic Profiling. In *Responsible AI*, ed. Silja Vöneky et al.
- Vöneky, Silja. 2020. Key Elements of Responsible Artificial Intelligence – Disruptive Technologies, *Dynamic Law. Ordnung der Wissenschaft* 1: 9–22.
- Wachter, Sandra. 2020. Affinity Profiling and Discrimination by Association in Online Behavioral Advertising. *Berkeley Technology Law Journal* 35 (2): 367–430.
- . forthcoming. The Theory of Artificial Immutability: Protecting Algorithmic Groups under Anti-Discrimination Law. *Tulane Law Review* 97: 2022–2023.
- Wachter, Sandra, Brent Mittelstadt, and Chris Russell. 2018. Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR. *Harvard Journal of Law & Technology* 31 (2): 841–887.
- Wikipedia. 2022. Science, technology, engineering, and mathematics. Available online: https://en.wikipedia.org/wiki/Science,_technology,_engineering,_and_mathematics. Accessed on 05.09.2022.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

