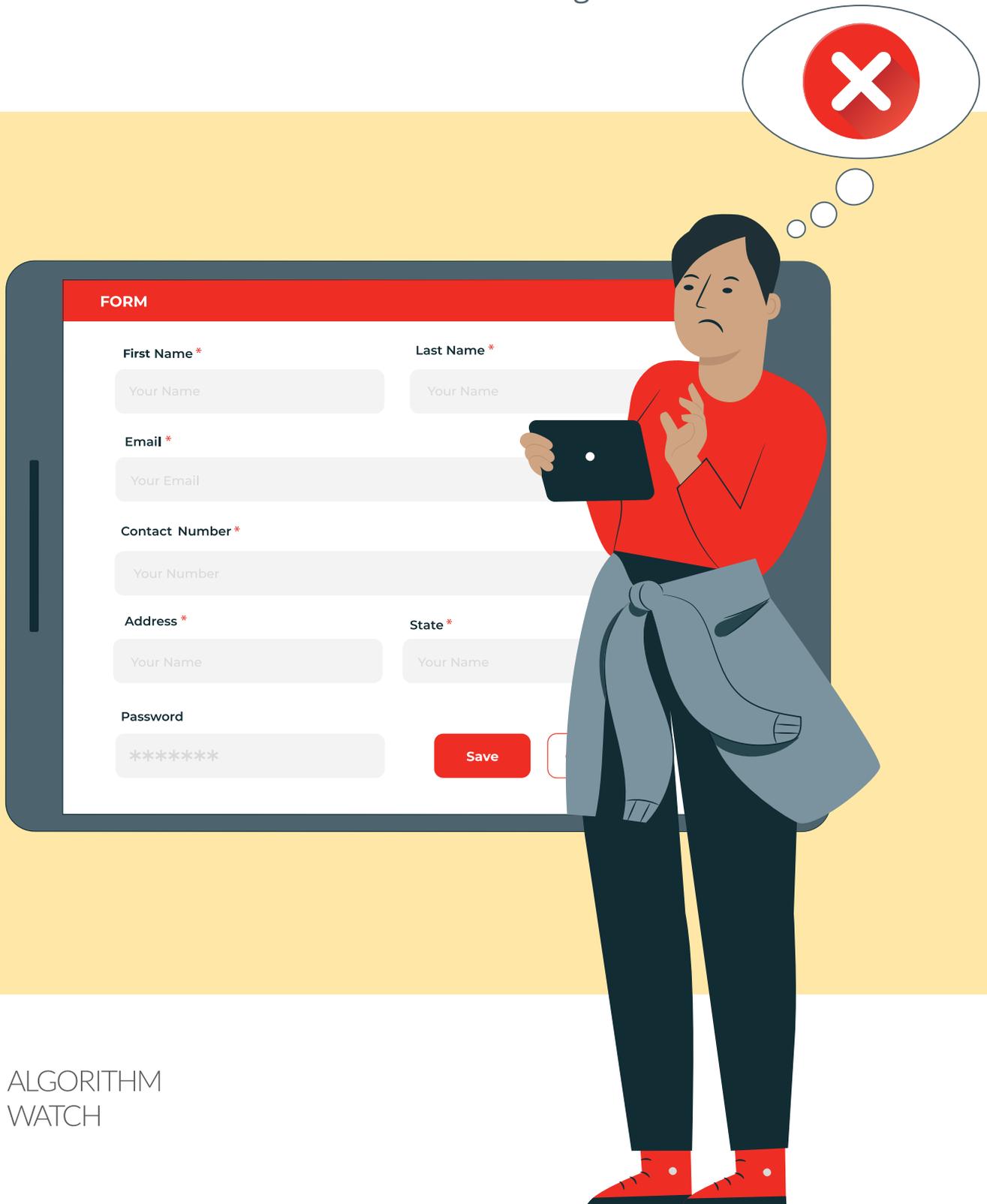


Automated Decision-Making Systems and Discrimination

Understanding causes, recognizing cases, supporting those affected

A guidebook for anti-discrimination counseling



Imprint

Automated Decision-Making Systems and Discrimination.
Understanding causes, recognizing cases, supporting those affected.
A guidebook for anti-discrimination counseling.

June 2022

Available online at: <https://algorithmwatch.org/en/autocheck/>

Publisher:

Matthias Spielkamp
AW AlgorithmWatch gGmbH
Linienstraße 13
10178 Berlin
Germany

Contact: info@algorithmwatch.org

Author and project manager “AutoCheck”

Jessica Wulf

Translation

Jessica Wulf

Layout and illustrations

Beate Autering

Copy editing

Graham Holliday

The “AutoCheck” project is funded by the Federal Anti-discrimination Agency Germany.



Antidiskriminierungsstelle
des Bundes



This publication is licensed under a Creative Commons Attribution 4.0 International License.
<https://creativecommons.org/licenses/by/4.0/legalcode>

▶ Introduction page 4

1 Basics and Key Terms

In this section we define key terms and explain important aspects of how automated decision-making (ADM) systems work. page 5

2 Case Studies

The case studies tell of unfair or discriminatory outcomes from the use of ADM systems. page 6

3 Causes

Charts illustrate how unfair or discriminatory outcomes are produced by using ADM systems. page 11

4 Fields of Application

With a focus on private services and access to goods, we illustrate how various ADM systems are used in the field. page 14

5 Recognize

A list of questions intended to help consultants identify indications of algorithms or ADM systems during case clarification in anti-discrimination counseling. page 16

6 Handling

Utilizing steps commonly used in anti-discrimination counseling, we show how counselors can deal with concerns about discrimination caused by ADM systems. page 18

7 Glossary

A clear explanation of the technical words and phrases used in this publication. page 22

8 Sources and Further Information

A list of recommended books, films, initiatives, and much more aimed to encourage further engagement with ADM systems. page 24

SYMBOLS IN THIS PUBLICATION



Race



Gender Identity



Age



Finance Sector



Insurance Sector



Online Trade



Public Administration



Labor Market



Housing Sector



Health Sector



Education Sector

▶ Introduction

“I am sure: risks of discrimination through automated decision-making systems will encompass all areas of life. It will be one of the most relevant issues in anti-discrimination work.” (Interviewee 02¹)

When our creditworthiness is assessed, when we are shown individualized prices online, or when we correspond with a chatbot² – these are examples of automated decision-making systems (ADM systems). We encounter these systems in more and more areas of our lives. However, ADM systems do not necessarily lead to more objective or neutral decisions, as the results are often unfair or even discriminatory.

The anti-discrimination expert quoted above assessed the discrimination risks associated with ADM systems as highly relevant to the future of anti-discrimination work. A study³ commissioned by the Federal Anti-Discrimination Agency (FADA) and a report by the European Network of Equality Bodies (Equinet) came to similar conclusions.⁴

However, what exactly are ADM systems? How does discrimination occur when they are used? And what can be done about it in counseling? We answer these and other questions in concrete and comprehensive instructions, diagrams, and text in this guidebook.

This publication was developed as a part of the project: *AutoCheck*⁵ – a Guide about Automated Decision-Making Systems for Equality Bodies. It is an introduction to the risks of discrimination when using ADM systems. For a deeper understanding of the topic, we developed training concepts within the AutoCheck project that we will publish and are offered as trainings for multipliers.

In this guidebook, we primarily address employees at anti-discrimination offices in Germany. By building up competencies in this increasingly important field, we hope to enable them to better recognize and assess risks to support those affected by discrimination. In addition, this publication is intended to give those affected by discrimination the opportunity to inform themselves about the topic.

The focus is on access to goods and private services, for example, discrimination in online commerce or when obtaining insurance. We also touch upon other areas of application of the General Equal Treatment Act.⁶

FADA funded this project, which was implemented by AlgorithmWatch. We would like to take this opportunity to thank both FADA for the funding and Nathalie Schlenzka, Niklas Hofmann, and Gudula Fritz for their support and superb collaboration.

1 As part of the AutoCheck project, 15 employees at anti-discrimination offices and anti-discrimination experts were interviewed. The main topics of the interviews were: Prior knowledge and points of contact with ADM systems in counseling, assessment of the relevance of the topic in an anti-discrimination context, and everyday tasks and topics in counseling work. The interviews were conducted in German and the quotes were translated by the author. All quotes in this publication, unless otherwise indicated, are statements from the interviews.

2 Terms relevant for this publication are grey and underlined in the text and explained in the glossary (p. 22f).

3 Orwat, Carsten (2019). Diskriminierungsrisiken durch Verwendung von Algorithmen: eine Studie, erstellt mit einer Zuwendung der Antidiskriminierungsstelle des Bundes. *Nomos*.

4 Allen QC, Robin & Masters, Dee (2020). Regulating For An Equal AI: A New Role For Equality Bodies. *An Equinet Publication*. https://equineteurope.org/wp-content/uploads/2020/06/ai_report_digital.pdf (14.04.2022).

5 <https://algorithmwatch.org/en/autocheck/>

6 The General Equal Treatment Act (in German: Allgemeines Gleichbehandlungsgesetz, AGG) “incorporates four Anti-Discrimination Directives of the EU into German law.” <https://www.antidiskriminierungsstelle.de/EN/about-discrimination/order-and-law/general-equal-treatment-act/general-equal-treatment-act-node> (03.06.2022).

1 Basics and Key Terms

*“I would like to have a general introduction to the topic, in order to know what we are talking about.”
(Interviewee 13)*

When you read the title of this guidebook “Automated Decision-Making Systems and Discrimination” you may wonder what is an automated decision-making system? We define this and other key terms used in this publication in this chapter. For further information on key terms⁷, please refer to the glossary (p. 22).

DEFINITION

When we talk about automated decisions made about humans, why do we use the term “automated decision-making system”? And not “artificial intelligence”?

First and foremost, *artificial intelligence* (AI) is a field of research in computer science that deals with machine learning and the automation of cognitive tasks. However, the term AI suggests proximity to human intelligence and, thus, to something resembling responsibility for decisions. This is misleading (see chapter 3). More appropriate is the term „automated decision-making system“, which we use instead.

ADM systems execute decision models or paths based on algorithms. ADM systems can support or prepare decisions by making recommendations or processing data. For example, an ADM system could be used to filter a list of candidates who meet certain criteria from a large number of applicants for an apartment. Based on this list, a human decides whom to invite to an apartment inspection. If the execution of a decision is completely delegated to an ADM system, then no human is involved in formulating the result.

ADM systems use algorithms, a specific form of instruction that leads to the solution of a mathematical problem. Algorithms describe “a solution path,

correctly interpretable by the computer, that computes the correct solution in finite time for each possible input defined by the mathematical problem.”⁸

The following two types of algorithms influence the functioning of ADM systems and need to be distinguished: rule-based and learning algorithms. In a rule-based algorithm, a programmer explicitly programs the instructions that the algorithm executes. An example of this is online forms, which you can often only submit if you assign yourself to one of the binary gender categories „male“ or „female“. A human programmed this condition in order to correctly fill out the form.

Learning algorithms, on the other hand, attempt to “learn” rules from pre-existing data that lead to a desired result in an initial phase. Humans can support this phase by annotating the data available for training, for example marking data that contains the desired result. In the next phase, the algorithm “applies” the learned rules to similar but unknown data to get a correct result. In this case, the data that has been used in the first phase has a great influence on the learned rules and thus the result.

Irrespective of the algorithms used, you need to consider the context in order to better understand ADM systems. It makes sense to view ADM systems as socio-technical systems, i.e., to include social components in the analysis in addition to technical components. The specific historical and societal context influence and define ADM systems. The systems, in turn, influence specific groups and individuals in that society as well as societal development as a whole. This influence can be negative and lead to disadvantages or discrimination, as the case studies in the following chapter show.

⁷ All key terms explained in the glossary are grey and underlined in the text.

⁸ AlgorithmWatch (2016). 1. Arbeitspapier: Was ist ein Algorithmus? <https://algorithmwatch.org/de/arbeitspapier-was-ist-ein-algorithmus/> (28.02.2022).

2 Case Studies

*“Case studies are always great. You can read a lot of theory, but you need examples to remember the theory.”
(Interviewee 06)*

The following five case studies illustrate how the introduction of ADM systems can lead to unfair or discriminatory outcomes. We chose these case studies on private services and access to goods in Germany to provide clues for consulting cases in which ADM systems might play a role. We also include an example from Italy and examples from other fields of application because the logic behind ADM systems and their application matters. The same logic applies to other contexts.



a) Do Age and Gender Affect Creditworthiness?

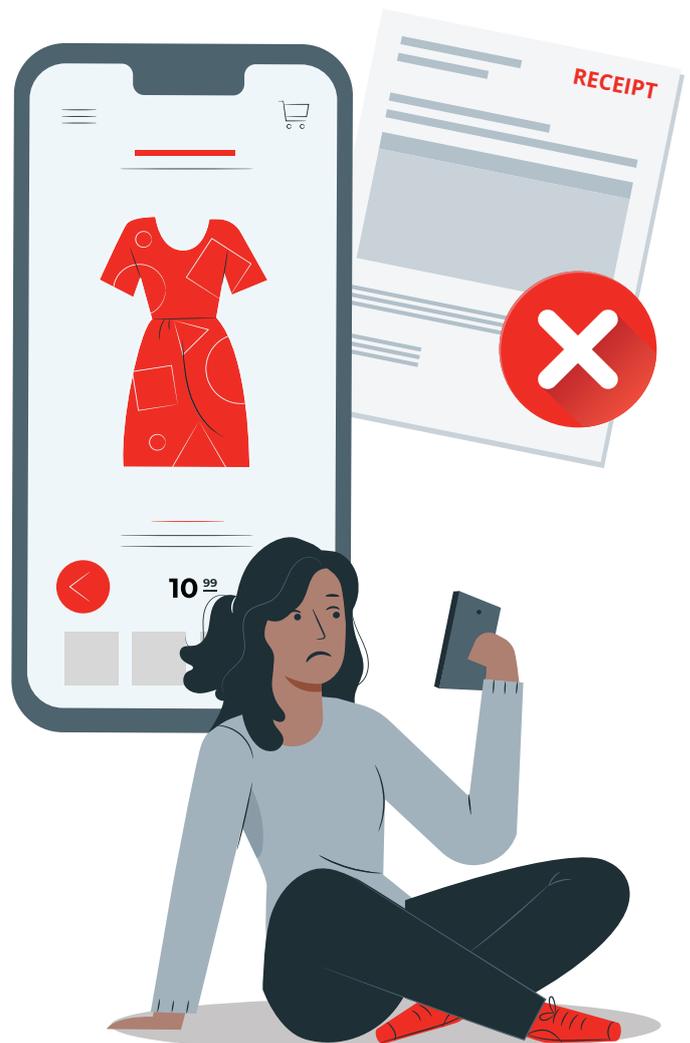
A woman cannot buy on account, presumably due to an unfavorable combination of age and gender for her creditworthiness.⁹

In 2021, a woman in Hesse, in central Germany, wanted to buy clothes from an online store by using the “purchase on account” option. However, the automated credit check (known as “scoring”); denied her the ability to complete the purchase this way. This irritated the woman as she had a good job, was around 40 years old and, to her knowledge, she had no negative entries with credit agencies.¹⁰

⁹ This case study is based on a counseling case. The counselor worked at an anti-discrimination agency and described this story during an interview. In Finland, a similar case was decided in a legally binding way by the Non-Discrimination and Equal Treatment Tribunal: <https://algorithmwatch.org/de/kreditscoring-urteil-aus-finnland-wirft-fragen-zur-diskriminierung-auf/> (24.03.2022).

¹⁰ The largest credit agencies in Germany are Schufa Holding AG (largest market share), Creditreform Boniversum, Crif Bürgel and Infoscore Consumer Data. For further information on credit agencies: <https://www.finanztip.de/schufa/auskunfteien/> (24.03.2022).

As a result, she contacted the online retailer who sent her the result of the check that was made by an external agency to assess her creditworthiness. The credit agency seemed to have used an ADM system that evaluated women of around 40 years old as not being creditworthy enough. The woman only learned the background to this automated decision when she decided to call the credit agency herself. During the call, an employee made assumptions about why she was rejected: Because „women her age were divorced and therefore destitute“ – in other words, the decision was due to a presumed connection between age and gender. In this case, a testing procedure (p. 19) could have worked well. However, it was too expensive and time-consuming for the affected woman and the anti-discrimination agency. In addition, if she had taken the case to court, the chance of a successful outcome in line with the General Equal Treatment Act was deemed too low in



relation to the cost.¹¹ Fortunately, the complaint led the online retailer to change their credit agency.

In this example, the model could have caused an insufficient credit score.¹² In that case, the assumption of the model would be that gender and age influence a person's ability to pay. You can find further information about this in [Chapter 3: Causes](#) (p. 11).

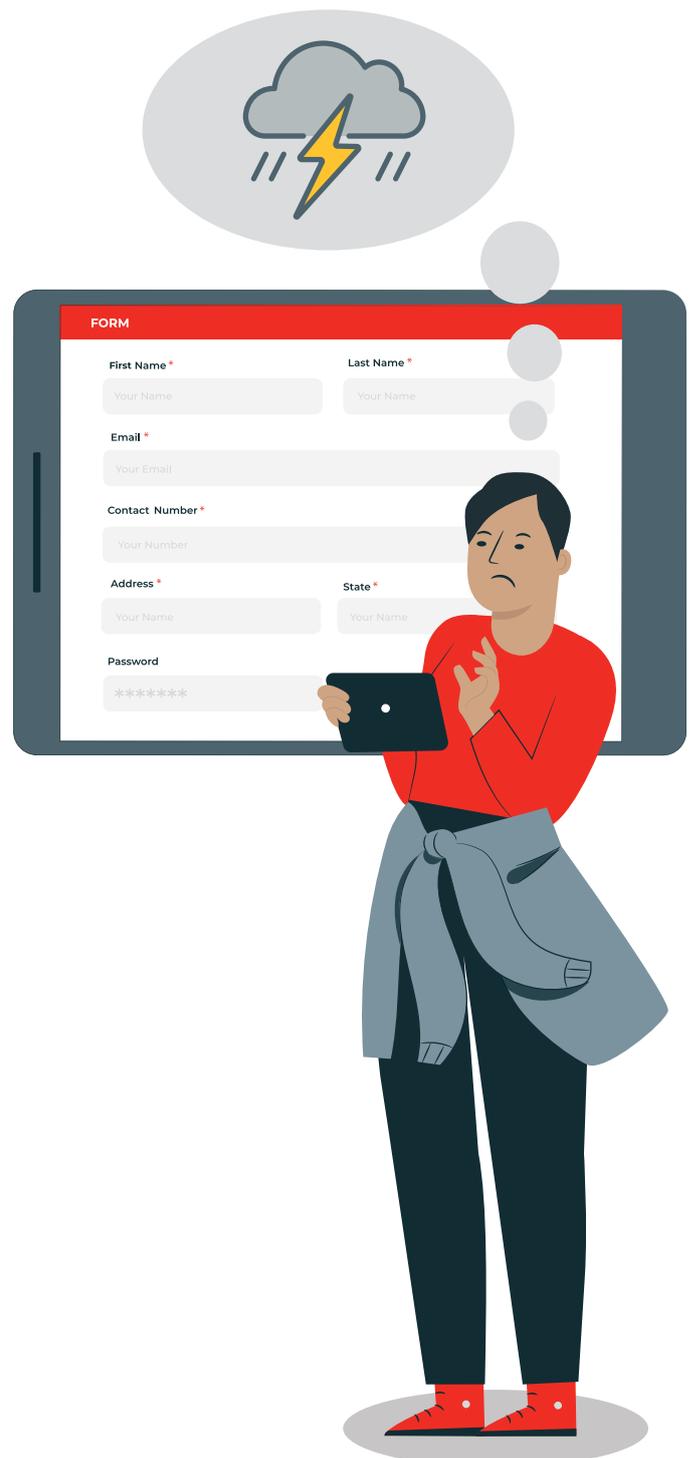


b) Online Forms

Online forms cannot be completed or submitted due to the shortness of the last name.¹³

In 2013, Clemens Li wanted to buy some Bitcoin cryptocurrency on the Bitcoin.de website. However, he was unsuccessful because the site's online registration form would not accept his last name. Mr. Li experienced the same problem on several other websites that use online forms. Often, an error message appeared whenever he filled in the name field as his surname, Li, was deemed 'too short'. Online forms are often programmed in such a way that names must contain three letters or more to complete the process. It would appear that people with less than three letters in their surname were not considered during the programming process.

Mr. Li complained to the operators on the Bitcoin.de website and, six months later, he received a response via email: The customer service representative asked him to add a dot to his name during the registration process, „because our system mandates three letters.“ Mr. Li refused and, as a result, he could not buy any cryptocurrency from the website.



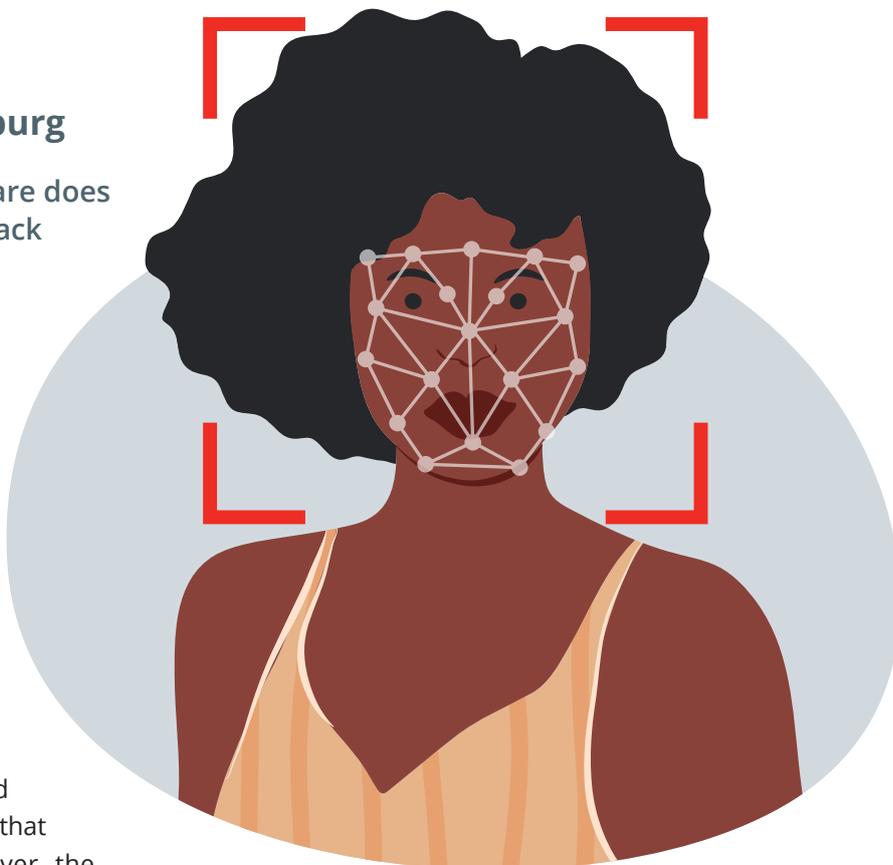
- 11 Among other things, the following report addresses the challenges for lawsuit proceedings regarding access to goods and services in connection with the General Equal Treatment Act: Beigang, Steffen et al. (2021). Möglichkeiten der Rechtsdurchsetzung des Diskriminierungsschutzes bei der Begründung, Durchführung und Beendigung zivilrechtlicher Schuldverhältnisse: Bestandsaufnahme, Alternativen und Weiterentwicklung. *Nomos*. https://www.anti-diskriminierungsstelle.de/SharedDocs/downloads/DE/publikationen/Expertisen/rechtsdurchsetzung_zivilrecht.pdf?__blob=publicationFile&v=2 (03.03.2022).
- 12 We do not know the actual reasons for the result, as we have no information from the credit agency about the case or the model used.
- 13 Lulamae, Josephine (2022). Fixing Online Forms Shouldn't Wait Until Retirement. *AlgorithmWatch*: <https://algorithm-watch.org/en/undoing-online-forms/> (03.03.2022).



c) Photo Booth in Hamburg

When face recognition software does not recognize the face of a black woman.¹⁴

In 2020, Audrey K. went to the State Office of Transportation in the northern port city of Hamburg to apply for an international driver's license. She brought all the necessary documents to her appointment apart from a biometric photo as she wanted to take the photo in the booth at the office. To take a biometric picture, you have to hold your face in a certain section of the camera, and only when the face is recognized in that section is the photo taken. However, the photo booth of the State Office of Transportation did not recognize her face. The face recognition software used in this photo booth apparently only recognized faces with light skin tones. The behavior of an employee at the office led Audrey K. to conclude that this was not the first time this had happened. She recalled the employee saying, „there might be a problem with your skin color.“ The photo booth belonged to the federal printing office. According to them, the problem was not software-related, as the latest technology was used in the photo booth. Instead, the office explained that the problem was due to insufficient lighting in the booth. However, research by Joy Buolamwini and Timnit Gebru



illustrated that the latest technology can include flaws that lead to racist and sexist results.¹⁵ In this study, the researchers examined the accuracy of AI applications, offered by the likes of Microsoft and IBM, that classify gender by faces. They concluded that the AI applications they examined showed a higher error rate for women with darker skin tones compared to men with lighter skin tones. A lack of diversity in data is cited as one of the reasons for this result. In the case study of Audrey K., the cause of the problem could also be due to the data used to develop the software.¹⁶

14 Schipkowski, Katharina (2020). Rassismus im Bild: Blind für schwarze Menschen. *Taz.de*: <https://taz.de/Rassismus-im-Bild/!5700872/> (03.03.2022). Similar cases: Achenbach, Cornelia (2019). Junge kann Fotoautomaten im Osnabrücker Bürgeramt nicht nutzen. *Neue Osnabrücker Zeitung*: <https://www.noz.de/lokales/osnabrueck/artikel/-20523365> (03.03.2022). And: Karangwa, Sandra (2015). Entlarvt: Schwarze zu dunkel für biometrische Passfotos. *MiGAZIN.de*: <https://www.migazin.de/2015/07/13/schwarze-sind-zu-dunkel-fuer-das-biometrische-passfoto/> (03.03.2022).

15 Buolamwini, Joy & Gebru, Timnit (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceeds of Machine Learning Research*: 81, 1-15. <http://gendershades.org/> (03.03.2022).

16 We do not know the actual reasons for the result.



d) Job Postings and Gender Stereotypical Ad Delivery

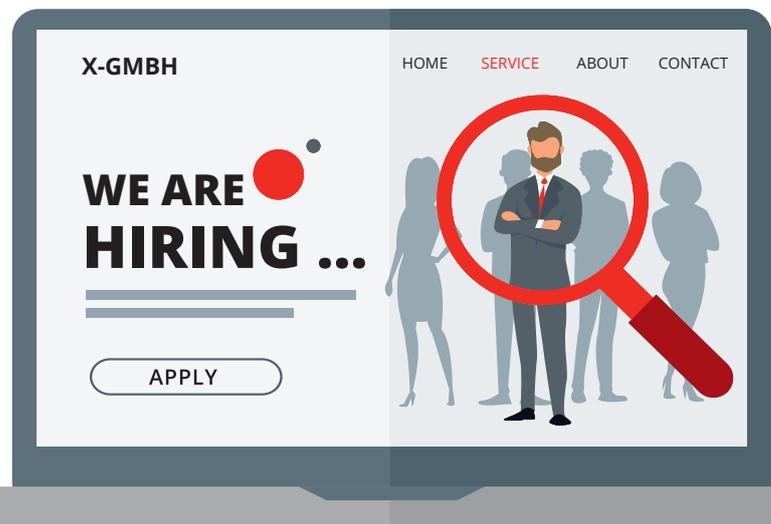
Job postings are distributed automatically in a gender stereotypical way.¹⁷

In 2020, as an experiment, AlgorithmWatch published two fictitious job postings. One vacancy was for a truck driver and the other was for an educator. Facebook displayed these job postings – which are linked to ads on a job platform – and when job seekers click on the ad, they are redirected to the job platform. Platforms like Facebook offer “targeting”: advertisers can specify a target group of people they want to reach with the ad. In the experiment, we deliberately chose not to use “targeting”. Once Facebook published the ads, they were primarily shown to people deemed most likely to click on the ads, based on a calculation by an algorithm. In our experiment, the ad for the truck driver was shown to 4,864 men, but only 386 women. Whereas the ad for the educator, which was posted at the same time, reached 6,456 women, but only 258 men. Thus, Facebook targeted the ads without asking permission or informing us about it. This form of targeting can be applied to other types of ads too, for example to ads for apartments. The main problem with this is that the people who are affected by it do not know that they are being shown something different compared to other users, nor on what criteria that is based.¹⁸

Algorithms used by these platforms mostly reflect existing structures and, therefore, can reproduce existing inequalities.

In this case study, gender stereotypical ad delivery could be due to the defined goal of the applied ADM system.¹⁹ The goal in this case might be to help maximize the number of clicks on job ads. For more information on the causes of discrimination by ADM systems, see [Chapter 3: Causes](#) (p. 11).

“The algorithms themselves are not the problem. The problem is that they mask discrimination and structural inequality. In the context of discrimination, it is not the intent that matters, but the effect. Discrimination is inscribed into the structure and into ourselves: We all learned this.” (Interviewee 04)



17 Kayser-Bril, Nicolas (2020). Automated Discrimination: Facebook uses gross stereotypes to optimize ad delivery. *AlgorithmWatch*: <https://algorithmwatch.org/en/automated-discrimination-facebook-google/> (03.03.2022).

18 An assessment of legal enforcement related to this example: Fröhlich, Wiebke (2020). Männer fahren LKW, Frauen erziehen Kinder – Diskriminierendes Gendertargeting durch Facebook. *Verfassungsblog*: <https://verfassungsblog.de/diskriminierende-facebook-algorithmen/> (03.03.2022).

19 We do not know the actual causes for these results, since we do not have any information from Facebook, e.g., which goals, models, or data were used.

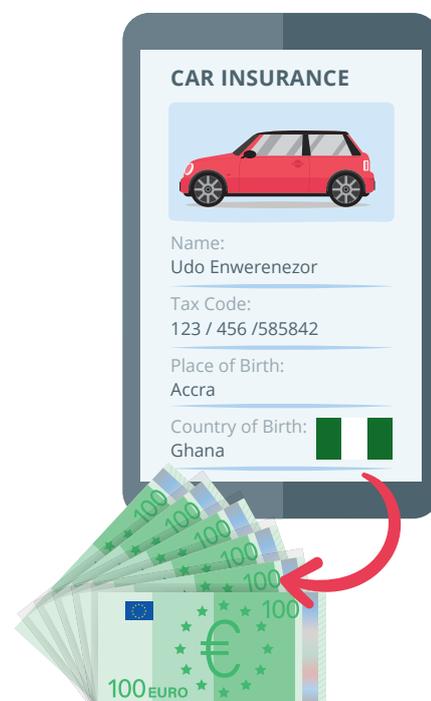


the price of car insurance varied greatly in some cases.²¹ The study showed that a driver born in Ghana might have to pay over €1,000 more than a driver from Milan with an otherwise identical profile.

e) The Cost of Car Insurance is Influenced by Your Nationality

Citizenship impacts car insurance prices.²⁰

Udo Enwerenezor lives in Italy and has dual Italian and Nigerian nationality. In the past, he had noticed how the price of car insurance differed depending on which nationality he indicated when asking for prices via telephone. If he indicated that he was Nigerian, the price was higher. However, since car insurance is now done almost exclusively online, he could not check whether his citizenship would affect the price. By asking for his tax code, the insurance companies already knew that he had dual citizenship. To reveal potential discrimination, the Universities of Padua, Udine, and Carnegie Mellon conducted research using comparison portals and fictitious profiles. Udo Enwerenezor's experience was confirmed: Depending on your place of birth, your citizenship, or both,



20 Boscarol, Francesco (2022). Costly Birthplace: discriminating insurance practice. *AlgorithmWatch*: <https://algorithmwatch.org/en/discriminating-insurance/> (03.03.2022).

21 Fabris, Alessandro et al. (2021). Algorithmic Audit of Italian Car Insurance: Evidence of Unfairness in Access and Pricing. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 458-468). http://www.dei.unipd.it/~silvello/papers/2021_aies2021.pdf (03.03.2022).

3 Causes

“Everything that is programmed, is only as good as the humans who programmed it.” (Interviewee 07)

The use of ADM systems can lead to unfair, even discriminatory, outcomes that can have far-reaching, negative effects on people’s lives. ADM systems are neither neutral nor objective. They must be understood as socio-technical systems – that is, as systems that emerge in a specific social context and are influenced by the structures and (power) relations that prevail within it.

“Automated decision-making systems are made by humans, who are subject to the same perceptions to which they are subject to in everyday life.” (Interviewee 11)

Unconscious biases, structural inequality, and discrimination that are common in societies also impact ADM systems because socio-technical systems reflect the assumptions, values, perspectives, and biases of the people who develop and apply them. This is the case because people make decisions at different points in the development and application of these systems. In the worst cases, these decisions lead to unwanted discrimination. In this chapter, we explain how misconceptions and biases can negatively affect how ADM systems work.

“Then the questions: Who programmed this? People who do not experience discrimination? Who do not deal with it? Who are not sensitive to it?” (Interviewee 07)

The following figures illustrate the moments people make decisions, how they introduce their assumptions and perspectives in those moments, and the questions we should ask about ADM systems.

DATA

Data can never represent reality exactly. Data is not neutral and does mirror relations in society. People decide what data is collected, when, and how. Data differs in quality, for example, in whether they are current or outdated and which groups are over- or under-represented in them. Data was collected in the past and thus reflects it. If data is used to predict the future, for example, this assumes that the future is similar to the past. This is especially so in the case of learning algorithms and machine learning – the data used affects how ADM systems work. Humans clean and annotate data so learning algorithms can process it.

The data used could, for example, have been the reason Audrey K.’s (case study c), p. 8) face was not recognized by the face recognition software used in the photo booth.²² If, in the learning phase, the algorithm

DATA

Does the data represent all groups on which ADM systems will have an effect?

Who chooses the data and why does the person choose that data (e.g., quality of the data, availability, costs)?



Who inspects the data? Is the data up-to-date and sufficient?

What questions can the data answer?

Who annotated the data and based on what criteria?



22 We do not know the actual reasons for the result.

was trained with data consisting predominantly of images of faces with light skin tones, then these were overrepresented and, in turn, faces with dark skin tones were under-represented. Consequently, ADM systems work badly, or not at all, for the under-represented group.

OBJECTIVE

People not only decide on the data to use, but also on the goal of using an ADM system. First, they identify a problem that needs to be solved, e.g., how to select the most suitable person for a job from a pool of applicants. The solution could be a recruiting process performed by humans. Alternatively, an ADM system might be seen as a viable way of solving the problem. The next step is to decide which objective the system itself should pursue. For example, to reduce costs, increase efficiency, etc. However, the objectives could also include the reduction of bias in a decision-making process or to prevent discrimination.

In case study d) (p. 9), Facebook displayed job ads based on gender stereotypes. The goal could have

been to maximize the number of clicks on job ads.²³ Based on historical data, Facebook could have calculated that female users are more likely to click on a job ad for an “educator”, while male users are more likely to click on an ad for a “truck driver”. If Facebook had defined a different goal, such as supporting gender equity in society, the ad distribution would probably have been different.

MODEL

A model describes a segment of reality; reality is reduced in complexity. By simplifying a situation, it can be transformed into a mathematical problem. People decide which criteria and aspects of the situation are relevant to the problem that the model is supposed to help solve. For example, when someone applies for a job, the ADM system should automatically select suitable applicants based on their CVs. However, people need to define what “suitable” means. This might include, work experience,

²³ We do not know the actual causes for these results, since we do not have any information from Facebook, e.g. which goals, models or data were used.

OBJECTIVE

What is the purpose of using the ADM system?

What (critical) questions are asked?
Are alternatives considered?

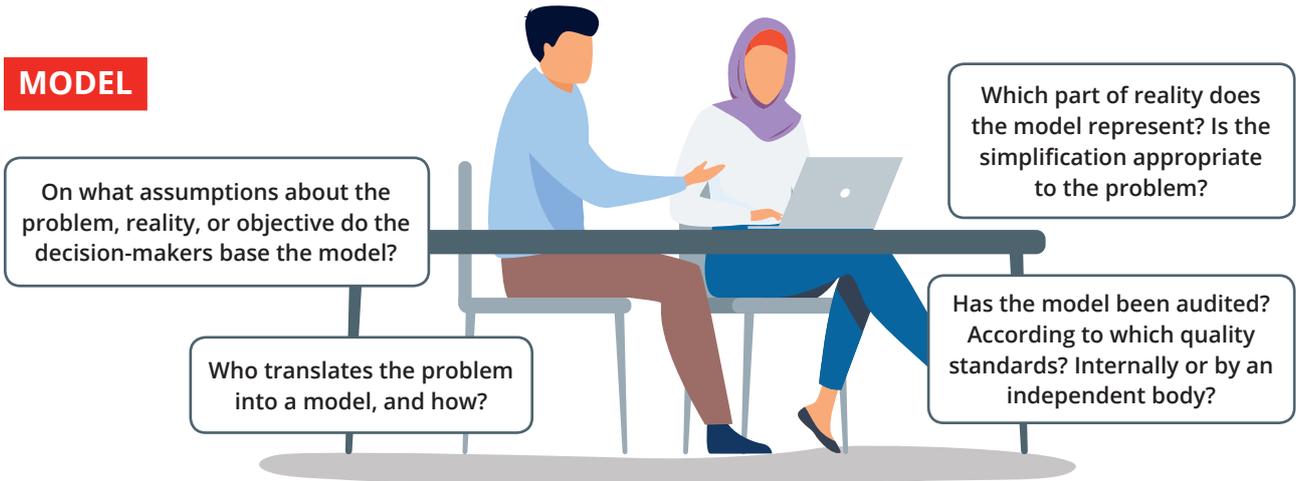
Based on which assumptions of ADM systems do people decide ADM is the best solution to an identified problem?

Who defines the objective of the application, the problem and the solution? What assumptions does this person have?

What is the decision-maker's perspective on society?



MODEL



education, language skills, etc. The decision to include work experience is based on the assumption that a person with several years of professional experience is better suited to a position than a person with less professional experience. However, this assumption could also be questioned if better criteria for suitability were included.

In scoring procedures, credit agencies assign individuals to a comparison group with similar characteristics. Thus, based on the payment behavior of the comparison group, they calculate how probable it is that a person will pay a bill. In case study a) (p. 6), the model could have caused an insufficient credit score result.²⁴ If that was the case, the assumption that gender and age influence someone's ability to pay would have entered the model.

APPLICATION

The decision to use an ADM system is based on assumptions about the possibilities and capabilities of those systems. People tend to overestimate ADM systems and rely on their recommendations. In scientific literature this tendency is called "automation bias".²⁵ This tendency increases with time pressure and a fear of negative consequences if users

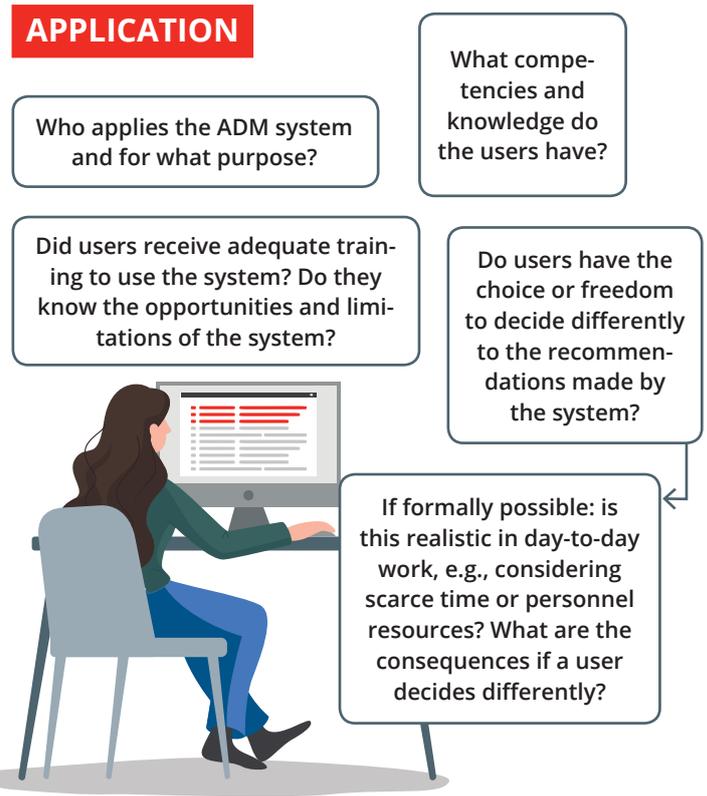
24 We do not know the actual reasons for the result, as we have no information from the credit agency about the case or the model used.

25 For further information on automation bias, see e.g., Cummings, Mary (2004). Automation Bias in Time Critical Decision Support Systems. *AIAA 1st Intelligent Systems Technical Conference*. <https://web.archive.org/web/20141101113133/http://web.mit.edu/aeroastro/labs/halab/papers/CummingsAIAAbias.pdf> (14.04.2022).

contradict the ADM system's recommendation. Therefore, users need to learn how these systems work, e.g., in trainings about the systems. As a result, users would be better able to question the recommendations appropriately.

Finally, we want to point out that even systems tested for discrimination and bias can be applied in a discriminatory manner. For example, if face recognition software worked equally well for all people, but it was used primarily to discriminate against black people. In this case, the application of the ADM system would be racist.

APPLICATION



4 Fields of Application

“Often, we do not know, where an algorithm might or might not be involved.” (Interviewee 07)

This figure describes areas where ADM systems can be applied and examples of how they are used. For the sake of clarity, we will not address how the ADM systems depicted can lead to discriminatory results. The diagram is intended to be used in consulting cases and to help develop a reasonable assumption for the use of ADM systems. The focus is on private

services and access to goods; areas beyond these are only touched upon. At present, it is difficult to say with certainty where ADM systems are used in Germany because transparency and information rights are both lacking. It is not mandatory to indicate when an ADM system is used to support a decision. In addition, organizations that use ADM systems don't need to explain the assumptions nor the data on which the systems are based. One indication that ADM systems are used is where statistical calculations have been carried out for a long time or where data are available. This is the case in the insurance sector and for credit checks. Often, ADM systems are used to sort, categorize, predict, or increase efficiency.



Payment options offered:

Payment options could change based on the information provided, such as the residential address. For example, at one address there might be the option to “purchase on account”, while at another address that option does not exist.



Customer accounts:

When creating a customer account, the customer shares information such as date of birth, gender, or citizenship. An ADM system could process this information.



Insurance Sector

- Online forms and customer accounts
- Chatbot interactions
- Group-specific insurance rates (e.g., students, senior citizens)
- Behavior-based insurance rates (e.g., telematics)
- Automated review of submitted documents
- Continuing obligation: Credit check
- Insurance concluded online, (e.g., travel cancellation, personal liability, legal protection insurance)
- Automated processing of a claim notification („digital clerk“)



Finance Sector

- Online forms and customer accounts
- Chatbot interactions
- Bank lending: Credit check
- Automated review of submitted documents
- Financial products concluded online, (e.g., current account, credit card, securities account, online loan)
- Online banks and financial technology



Online Trade

- Online forms and customer accounts
- Chatbot interactions
- Group-specific prices (e.g., students, senior citizens)
- Bonuses based on customer profiles
- Price differentiation based on time or location of purchase
- Data-based, targeted advertisements (e.g., based on search history)
- Continuing obligation for mobile or Internet contracts and subscriptions (e.g., Spotify, Netflix): credit check
- Offered payment methods (e.g., based on address)
- Installments or purchase on account: credit check
- Price differentiation based on age



Chatbot interactions: A chatbot automatically answers questions via a dialogue box. In some cases, chatbots learn through learning algorithms and adapt their answers over time.



Allocation of daycare places: Based on predefined allocation criteria, an ADM system could recommend which children should get a place in which daycare center.



Application procedures: Based on voice samples, speech recognition software is supposed to reach conclusions about the person speaking.



Pattern recognition: An ADM system could analyze images of people's skin to detect patterns related to melanomas at an early stage.



Housing Sector

- Online forms and customer accounts
- Chatbot interactions
- Tenancy: credit check
- Pre-sorting of applications for apartments (e.g., when there are a large number of apartments for rent or applications)
- Apartment ads on online platforms: Which ads are displayed to whom? (automated placement)



Health Sector

- Private healthcare and supplementary insurance (e.g., dental or hospital supplementary insurance - see insurance sector)
- Online appointment allocation (e.g., for COVID-19 vaccinations)
- Prevention and diagnostics: Pattern recognition to detect diseases at an early stage.



Education Sector

- Automated assignment of daycare places for children
- Automated selection procedure at private schools or universities



Labor Market

- Application procedure (Language analysis, Video analysis)
- Automated selection procedure of applications based on grades, CV, etc.
- Job ads on online portals (Automated placement)
- People Analytics (e.g., Performance evaluation, Personnel development, Calculate willingness to change jobs)



Group-specific insurance rates and forms: An ADM system could calculate different insurance rates automatically based on information provided, such as date of birth or employment relationship.



Supplementary dental insurance: An ADM system could automatically calculate different insurance rates based on information from online forms or the customer account.



5 Recognize

“How do I learn to recognize, whether there might be hidden digital processes behind it?” (Interviewee 10)

Below is a list of questions you can use in anti-discrimination counseling. The questions are intended to facilitate the recognition of indicators for ADM systems through asking appropriate questions. To use this list of questions you do not need prior technical knowledge.

Where did the interaction take place?

- If the interaction occurred online, this may indicate the use of ADM systems. Examples of such online interactions:

-  Search for apartments via online housing portals
-  Contracts concluded online, such as insurance contracts
-  Interactions with a chatbot, e.g., online trade
-  However: Credit checks by lenders, or video/voice recording when applying for a job, are situations in which ADM systems can be applied, even if interactions did not take place online.
- 

Was an online form filled out (contact form, chatbot, etc.)?

- For example, opening a securities account using an online account or taking out travel insurance via a website.
- Was personal information provided or personal data shared? If yes, what was shared? (Such data may also be stored in online accounts).
- Were questions asked about information that is protected by the General Equal Treatment Act (direct discrimination)? Were questions asked about information that could act as proxy variables (indirect discrimination)?

PROXY VARIABLES

Proxy variables substitute other variables for example, if those cannot be observed or measured directly. A synonym for the word “proxy” is “representative”. Proxy variables allow conclusions about certain – also protected – characteristics without recording those characteristics directly. For example, “30 years of work experience” indicates that the person must be at least in his or her mid-40s. Thus, proxy variables can have a decisive influence on the decisions of an ADM system (You can find a video explaining proxy variables here: <https://www.youtube.com/watch?v=NsIS2kIFTEY> (11.05.2022)).

“An important point: It is often assumed that discrimination does not occur if no personal data is used. This is not true.” (Interviewee 02)

- Have any documents been uploaded? If yes, what documents? Do they contain any preteed information?

How did the company or organization respond to the interaction?

- How fast did they respond? The following points indicate the involvement of ADM systems:
 - Is a decision made immediately when data is entered into an online form or is the decision communicated immediately afterwards by email?
 - Were a lot of documents uploaded, yet the response still came very quickly?
- Do certain terms or phrases appear in the response that indicate an automated decision?
 - For example, a common phrase is “This letter was generated automatically and is valid without a signature.”
 - Common terms are: “automated”, “machine”, “computerized”, “automatic”, “electronic”.

If something was purchased online, what payment options were available and which option was selected?

- With installment or payment on account options, a credit check is very likely performed.
- Is it a continuing obligation, such as rent, telephone contracts, subscriptions to Spotify or Netflix, etc.? Here, too, it is likely that a credit check is performed.
- A credit check may also be performed when participating in direct debit procedures.

Usually, explicit consent must be given for a credit check. This consent is in most cases obtained as part of the agreement to purchase. The information on this can be found in the contract documents, the General Terms and Conditions (GTC), or similar documents.

Did you give consent to automated data processing?

- In order for fully automated decisions to be made, the affected person must explicitly consent in accordance with Article 22 of the European Data Protection Act (GDPR) or Section 31 of the German Federal Data Protection Act. Information about this can be found in the contract documents, the GTC, the data protection agreement, or similar documents. Also, look for words and phrases like: “automated”, “machine”, “computerized”, “automatic”, “using algorithms”.
- Here is an example from the privacy policy of an online bookstore:
 - “On the basis of this data, our service provider calculates probability values using mathematical-statistical methods to assess the risk of non-payment and to verify your address (check for deliverability).”

“I suspect that discrimination risks through automated decision-making systems are very, very relevant, but do not reach us in consulting or I do not recognize them as a consultant.” (Interviewee 00)

6 Handling

*“What do I do if I suspect automation?
Is that an indication of discrimination?
That is of course a big question.”
(Interviewee 00)*

During a counseling case, if you suspect that an ADM system has been used, the logical question is: what are the consequences of this for the counseling process? In this chapter we describe ways of dealing with such suspicions following common interventions in anti-discrimination counseling. As in all other counseling cases, the first step is to find out what the person seeking counseling wants.²⁶ In addition, it makes sense to include a reference to an ADM system in the documentation of the counseling case. This makes it easier to retrieve such cases, bundle them together and identify any patterns.

26 Antidiskriminierungsstelle des Bundes (2012). Schritt für Schritt durch die rechtliche Antidiskriminierungsberatung – Leitfaden für Beraterinnen und Berater: S. 30-59. Antidiskriminierungsverband Deutschland (2015). Standards für eine qualifizierte Antidiskriminierungsberatung: S. 12f.

How can you integrate suspected negative impacts of ADM systems into common practices and interventions of anti-discrimination counseling?

Contact, statement, letter of complaint

Contacting a responsible person, organization, or agency in the form of a request for a statement (perhaps to reach an agreement) or a letter of complaint is an option if it is in the interest of the person seeking advice.

If you suspect an algorithm or ADM system contributed to a discriminatory or unfair decision, describe this suspicion when contacting the organization. You can use the following phrases and sample questions for this purpose:

“Based on some indications, we assume that an automated decision-making system was used in this case. By this, I mean that decision-making models were applied in an automated way by an algorithm or a computer program. Did you use such a system? If so, what software did you use? Did a fully

WHAT CAN YOU DO?

The risk of discrimination through ADM systems is a relatively new problem in anti-discrimination counseling, and cases vary greatly depending on which ADM system is used. Currently, only a few cases from Germany exist. As a result, there is little empirical knowledge you can rely on.

Therefore, if you suspect that an ADM system has been used, we recommend you contact one of the organizations listed below as a first step. On the one hand, these organizations can provide direct support on the case, and, on the other, they can refer you to the appropriate bodies (such as consumer protection or data protection agencies):

AlgorithmWatch

✉ info@algorithmwatch.org
☎ +49.30.994049.000

Federal Anti-Discrimination Agency

✉ beratung@ads.bund.de
(im Betreff: „Bitte um kollegiale Beratung zu KI-Fall“)

Anti-Discrimination Association Germany

✉ info@antidiskriminierung.org

automated decision take place or was the decision prepared in an automated manner and then made or controlled by a human? How was the automated system secured against discrimination and other unwanted influences? How was it ensured that no personal characteristics were used?"

Testing procedure²⁷

One way of checking for discrimination is by using the testing procedure. For example, if discrimination is suspected when searching for an apartment, it works as follows: "A test person and a control person each apply for an apartment. This two people have as many similarities as possible, apart from the characteristic being tested. Individuals who suspect that they have been unfairly disadvantaged in the housing market or in other areas can use single-case testing. The evidence of discrimination obtained through testing can then be used in court to substantiate a claim."²⁸ This method can also be used in other areas of the General Equal Treatment Act, such as lending.

The testing method is useful in the context of automated decisions since the decisions are always made according to the same basic pattern, i.e., systematically. However, testing procedures are often very time-consuming and demanding, e.g., for credit checks.

27 Further literature on the testing procedure: Senatsverwaltung für Arbeit, Integration und Frauen. Landesstelle für Gleichbehandlung – gegen Diskriminierung. Diskriminierung sichtbar machen (2012). Dokumentation der LADS-Fachrunde vom 8.11.2012 zu fachlichen und methodologischen Anforderungen an Testing-Verfahren. Antidiskriminierungsstelle des Bundes (2020). Fair mieten – fair wohnen. Leitfaden für Mieterinnen und Mieter und Beratungsstellen.

28 Senatsverwaltung für Arbeit, Integration und Frauen. Landesstelle für Gleichbehandlung – gegen Diskriminierung. Diskriminierung sichtbar machen (2012). Dokumentation der LADS-Fachrunde vom 8.11.2012 zu fachlichen und methodologischen Anforderungen an Testing-Verfahren, S. 10.



TESTING PROCEDURE

The woman from [case study a\)](#) [p. 6] feels discriminated against and goes to an anti-discrimination office. According to the consultant, the case is suited for testing, since the scoring procedure always follows the same logic. However, in order to be meaningful, the test would need to be designed on a large scale with a very large number of similar persons. This is too costly and resource-intensive for the counseling center at this point in time.

It is important to distinguish whether the results of the testing procedure are to be used as circumstantial evidence in legal proceedings, or in support of a statement. Unfortunately, outside of discrimination in the housing market, there is little case law to indicate whether the results of such a testing procedure would be accepted as circumstantial evidence in court. The obstacles in court are high.

In comparison, the results of a testing procedure could have great potential when a statement is requested. However, the counseling center should always decide whether it has sufficient resources available for such a procedure.

With online forms, testing is relatively easy to perform as you can change information easily. If, for example, you suspect discrimination based on gender in an apartment inquiry sent via an apartment portal, you can change the gender on a test basis. In this way, you can check whether the response or decision changes depending on gender.



NOTE

Testing can be very time-consuming and tie up a lot of resources. It makes sense to bundle cases related to automated decisions and discuss them at networking meetings, for example. It is helpful to explicitly include ADM systems in the case documentation, e.g., in the case description or as a separate category in the case database.

Would you like to exchange information with colleagues who have experience with testing procedures? The following counseling centers have conducted testing procedures in the past:

Planerladen gGmbH,

Focus on the housing market

 www.integrationsprojekt.net

 integration@planerladen.de

 +49.231.833225

Antidiskriminierungsbüro Sachsen e. V.

 <https://www.adb-sachsen.de/>

 beratung@adb-sachsen.de  +49.341.30690777

Antidiskriminierungsnetzwerk Berlin des Türkischen Bundes in Berlin-Brandenburg

 <https://www.adnb.de/>

 adnb@tbb-berlin.de  +49.30.61305328

Creating visibility: public relations and press work

Since this is a relatively new topic and few case studies from Germany are currently known, press and public relations work is important when a case occurs. Publicizing the cases supports awareness about the topic and the problem in a public discussion. Furthermore, making discrimination cases through ADM systems public is vital to help raise awareness at a political level. Gathering case studies contributes to a clearer picture of the situation regarding the use of ADM systems and the associated risks.

If the anti-discrimination counseling center you work for does not do any public relations work itself, contact AlgorithmWatch, the Federal Anti-Discrimination Agency or the Anti-Discrimination Association Germany [[contact details](#), p. 18] for support and referral.

Litigation and strategic litigation

If you have a case related to private services and access to goods that is linked to a characteristic protected under the General Equal Treatment Act, there

is also the option of legal action.²⁹ However, you need to consider that legal action entails costs and a burden for the person concerned, and resources are needed for the work done by the anti-discrimination counseling center. The resources required are comparable to other areas in which little or no case law exists.

The main difficulty is proving discrimination or gathering enough circumstantial evidence that the burden of proof is reversed.³⁰ If the person concerned decides to file a lawsuit and you are advising on the case, it is important to discuss with the legal representation to what extent attention can be drawn to the fact that the decision was (partially) automated. This could benefit the press and public relations work surrounding the case.

In addition, depending on the case and available resources, strategic litigation might be an option. Strategic litigation in anti-discrimination law „describes a method in which court proceedings are used systematically to bring about social change. It involves tactically combining legal tools and flanking measures to achieve an impact that goes beyond the individual case.“³¹ Strategic litigation can be used to address discriminating structures. Depending on the particular case of an automated decision with a

29 Note: An action under the General Equal Treatment Act requires that the claims are asserted in good time. Otherwise, the action is inadmissible. According to Section 21 of the General Equal Treatment Act, the claim must be asserted within two months in order to keep the option of a lawsuit open.

30 A policy brief in which AlgorithmWatch discussed how the General Equal Treatment Act could be adapted to better legally protect affected persons from discrimination by ADM systems: Michot, Sarah et al. (2022). Algorithmen-basierte Diskriminierung – Warum Antidiskriminierungsgesetze jetzt angepasst werden müssen. *AlgorithmWatch/Digital Autonomy Hub*. https://algorithmwatch.org/de/wp-content/uploads/2022/02/DAH_Policy_Brief_5.pdf (14.04.2022).

31 See Beigang, Steffen et al. (2021). Möglichkeiten der Rechtsdurchsetzung des Diskriminierungsschutzes bei der Begründung, Durchführung und Beendigung zivilrechtlicher Schuldverhältnisse: Bestandsaufnahme, Alternativen und Weiterentwicklung. *Nomos*: S. 159. https://www.antidiskriminierungsstelle.de/SharedDocs/downloads/DE/publikationen/Expertisen/rechtsdurchsetzung_zivilrecht.pdf?__blob=publicationFile&v=2 (03.03.2022).

discriminatory outcome, strategic litigation might be appropriate because automated decisions are systematic and often reproduce existing discriminatory structures.

ARTIFICIAL INTELLIGENCE ACT

Currently, the Artificial Intelligence Act is being negotiated at the EU level.³² The law aims to regulate the use of AI systems, especially those that involve high-risk. The proposal establishes a legal framework designed to ensure the protection of fundamental rights while promoting innovation.

Since the widespread use of automated systems is a relatively new phenomenon, many legal questions are currently pending. There is an ongoing discussion about the role that data or consumer protection could have in addition to, or in interaction with, the General Equal Treatment Act. In addition, the European Union is currently negotiating and adopting various draft laws that will have an impact on the use of ADM systems in the context of anti-discrimination law, most notably the Artificial Intelligence Act and the Digital Services Act (DSA).

DIGITAL SERVICES ACT (DSA)

EU institutions recently agreed on the final version of the DSA. It sets out new obligations for providers of digital intermediation services, especially for very large online platforms. These provisions are intended to promote a safer online environment in which people are able to exercise their fundamental rights online, including the right to non-discrimination. Specific provisions related to non-discrimination on platforms include rules that curb targeted advertising, revised reporting mechanisms, and new transparency and accountability measures. Services such as Facebook, Google, or Amazon are to assess whether they pose “systemic risks,” undergo audits, and make data available for research in the public interest.

NOTE

In order to take a closer look at legal questions with regard to protection given by the General Equal Treatment Act in cases of discrimination by algorithms, the FADA has commissioned the legal opinion „The General Equal Treatment Act and the Protection against Discrimination by Algorithmic Decision Systems“ (tender title). The expert opinion is being prepared by Prof. Dr. iur. Indra Spiecker gen. Döhmann and Prof. Dr. Emanuel V. Towfigh. It will be published at the end of 2022/beginning of 2023 on the [publication page](#) of the FADA.

32 Link to the draft law: [https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206\(14.04.2022\)](https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206(14.04.2022)).

7 Glossary

This glossary explains basic terms. It is important to us that the entire publication is understandable even without technical knowledge. With this explanation of terms, we want to contribute to a basic understanding of how ADM systems work, which is useful for anti-discrimination work.

Algorithm – is a specific form of instruction that leads to the solution of a mathematical problem in defined individual steps. Algorithms were used long before the development of computers. However, for this publication, we refer to algorithms that are used in various socio-technical systems and influence their functioning. Algorithms describe “a solution path, correctly interpretable by the computer, that computes the correct solution in finite time for each possible input defined by the mathematical problem.”³³

Artificial Intelligence (AI) – is a research field in computer science that deals with machine learning and the automation of cognitive tasks. The general goal is to enable machines to perform tasks that somehow appear ‘intelligent’. What ‘intelligence’ means in this context, and with which techniques it can be achieved, is not defined.³⁴ The term AI is commonly used for very different ADM systems e.g., speech recognition, playing chess, or translating into different languages. What is problematic about the term is that human-like intelligence is suggested and with it responsibility for decisions. However, AI systems do not bear responsibility; only humans can do that.

Automated Decision-Making Systems / (ADM) systems – Algorithms execute decision models or decision paths automatically. ADM systems can support humans in decision-making by making recommendations or preparing decisions by processing data. The execution of a decision can also be completely delegated to an ADM system, in which case no human is involved in formulating the result.

Chatbot – is a dialogue system based on software. Chatbots are increasingly used in customer service and for interaction on the Internet. Questions asked via a dialogue box on a website can be answered automatically by a chatbot. The communication between chatbots and humans happens in natural language. In some cases, chatbots even process spoken language. Chatbots based on machine learning methods can learn while they are in operation.

Data – are numerical values, which were collected by measurements, statistical surveys, observations, and the like. Data can never represent reality one-to-one. People decide which data are collected and how they are used, and thus also what can be discovered through the data.³⁵ Data reflect the structures of society and are not neutral.

Data annotation – is the process of preparing data for use by learning algorithms, e.g., by assigning data to a particular category or feature. This could be marking data that contains the correct result for a task. Annotations are often assigned by humans. For example, if an algorithm is trained to identify cat images at the end, images annotated as cat images are necessary. From these, the algorithm “learns” the correct answer.

Explainable AI – is a subfield of AI. In many artificial intelligence systems, humans cannot understand how logic or criteria decisions, recommendations, or predictions are made. Researchers looking into explainable AI attempt to make the underlying logic and individual results more comprehensible. Comprehensibility is necessary to better recognize and avoid biased and unfair results.

33 AlgorithmWatch (2016). 1. Arbeitspapier: Was ist ein Algorithmus? <https://algorithmwatch.org/de/arbeitspapier-was-ist-ein-algorithmus/> (28.02.2022). For a detailed definition and description in the context of anti-discrimination work, please see: Orwat, Carsten (2019). Diskriminierungsrisiken durch Verwendung von Algorithmen: eine Studie, erstellt mit einer Zuwendung der Antidiskriminierungsstelle des Bundes. *Nomos*. S. 3-5.

34 Döbel, Inga et al. (2018). Maschinelles Lernen – eine Analyse zu Kompetenzen, Forschung und Anwendung. *Fraunhofer*: S. 44. <https://www.bigdata-ai.fraunhofer.de/publikationen/ml-studie.html> (09.03.2022).

35 For more information and studies on this topic: Criado Perez, Caroline (2019). *Invisible women: Exposing data bias in a world designed for men*. Random House.

Financial Technology – includes new products or services in the financial sector that are based on algorithms and ADM systems.

Learning Algorithms – solve a task in two different phases: In the first phase, the algorithm tries to “learn” rules from existing data that point to the desired result. This often happens under supervision, i.e., through annotated data, the algorithm is instructed which part of a data set from the past included the desired results and which did not. The rules a learning algorithm deduces in the first phase are usually obscure and often incomprehensible to humans. In the second phase, the algorithm applies these learned rules to make decisions about unknown data of the same type.³⁶ Learning algorithms are used for machine learning.

Machine Learning (ML) – is a research field in computer science dealing with various kinds of AI methods. Without directly programming the solution path, it should be found automatically in the data. ML is based on learning algorithms that analyze patterns in existing data and use them for machine learning models. These models are then applied to new data.³⁷

(Mathematical) Model – describes a section of reality. It is an abstraction and can never represent all aspects of reality. A model should be accurate enough that it enables predictions or analyses about a certain topic or situation.

Online Platforms – are websites that enable different actors to exchange information. They serve a variety of purposes, e.g., online social networks offer communication opportunities, job platforms help people find job postings, and trading platforms facilitate online buying and selling. Algorithms can, for example, influence the order of search queries and thus who gets to see which content and how prominently.

36 AlgorithmWatch (2016). 2. Arbeitspapier: Überprüfbarkeit von Algorithmen. <https://algorithmwatch.org/de/zweites-arbeitspapier-ueberpruefbarkeit-algorithmen/> (14.04.2022).

37 Vgl. Plattform Lernende Systeme (2022). Glossar: Maschinelles Lernen. <https://www.plattform-lernende-systeme.de/glossar.html> (14.04.2022).

Proxy Variable – is a variable that is a substitute for another variable, e.g., if those cannot be observed or measured directly. Proxy variables can be used to infer certain – even protected – characteristics without directly querying or recording them. For example, “30 years of work experience” indicates that the person must be at least in his or her mid-40s. Thus, proxy variables can have a decisive influence on the decisions of an automated system.

Rule-based algorithms – humans explicitly programmed the rules and instructions for actions that the algorithm executes. This is often the case with online forms. For example, if you can only submit a form by assigning yourself to one of the binary gender categories “male” or “female”, then a human being programmed this condition in such a way. Rule-based algorithms differ from learning algorithms in that their rules and criteria were explicitly programmed.

Scoring – credit agencies calculate score values based on data. “These express how likely the credit agency believes a person is to pay their bills. In doing so, they do not evaluate the person in question individually, but assign him or her to a comparison group with similar characteristics using statistical methods. The payment behavior that this comparison group has shown in the past is then also assumed for the person in question.”³⁸

Socio-technical systems – to better understand ADM systems they should be considered as socio-technical systems, i.e., to regard their social components as well as their technical components. Such systems emerge from and are influenced by a specific historical and social context and, in turn, impact social development as a whole and on specific groups and individuals in that society.

Targeting – describes the specific dissemination of advertisements. The aim is to address people according to their interests and increase the likelihood that they will click on the ads. For example, people who want to buy a new sofa should see more ads for sofas. Data, such as previous search behavior, is used to do this.

38 Lietzau, Josefine (2019). Wie Schufa und Co. mit unseren Daten handeln. *Finanztip*. <https://www.finanztip.de/schufa/auskunftseiten/> (14.04.2022).

8 Sources and Further Information

"I would like to have a pool of experts that I can ask." (Interviewee 10)

Here we list organizations that regularly publish news about ADM systems and discrimination risks. Furthermore, we recommend relevant sources on the topic.

Relevant organizations that regularly publish reports, articles, and news about ADM

AlgorithmWatch

A nonprofit research and advocacy organization dedicated to studying and explaining ADM systems and their impact on society. Website: <https://algorithm-watch.org/>

Federal Anti-Discrimination Agency (FADA)

In addition to providing anti-discrimination counseling, FADA commissions studies, and research projects, including on ADM systems. The results are published here: <https://www.antidiskriminierungsstelle.de/DE/was-wir-machen/publikationen/publikationen-node.html>

AI + Automation Lab (Bayrischer Rundfunk)

This is an interdisciplinary team that regularly publishes articles and studies „at the intersection of journalism, computer science and product development.“ Website: <https://www.br.de/extra/ai-automation-lab/index.html>

Bertelsmann Stiftung – Ethics of Algorithms

In the Ethics of Algorithms project, the Bertelsmann Stiftung addresses the social impact of ADM. The foundation funds various projects in this area and publishes articles and studies on their project page: <https://www.bertelsmann-stiftung.de/en/our-projects/ethics-of-algorithms>

European Digital Rights (EDRI)

A network of civil society and human rights organizations from Europe that works to advance digital rights. Website: <https://edri.org/>.

Federation of German Consumer Organizations (vzbv) und Verbraucherzentrale Hesse

With a focus on consumer protection, this federation publishes easy-to-understand information on algorithms and AI applications. Websites: <https://www.vzbv.de/algorithmen-ki> and <https://www.algo-was.de/>

Gesellschaft für Freiheitsrechte e. V. (GFF)

This association works to strengthen fundamental and human rights at the German and European levels. For example, through strategic litigation. Website: <https://freiheitsrechte.org/english/>

Landesantidiskriminierungsstelle Berlin (LADS)

The tasks of LADS include public relations work on anti-discrimination laws and encouraging people to exercise their rights. This also includes the discrimination risks associated with ADM systems. You can find their publications here: <https://www.berlin.de/sen/lads/ueber-uns/materialien/>.

netzforma e. V.

This association is committed to feminist network policy and, among other things, it also deals with AI applications and algorithms. In addition to publications, they also offer workshops. Website: <https://netzforma.org/aktuelles>

Netzpolitik.org

This online medium regularly publishes articles and commentaries on the topic of digital freedom rights – in this context, the topic of anti-discrimination is regularly covered. Website: <https://netzpolitik.org/>

Superr Lab

This group explores how technologies can be used to create a more equitable digital future and it builds networks to do so. Website: <https://superr.net/news/>

Books and reports

Allen, Robin & Masters, Dee (2020). Regulating for an Equal AI: A new Role for Equality Bodies. *An Equinet Publication*. <http://ai.equineteurope.org/library/regulating-equal-ai-new-role-equality-bodies> (25.02.2022).

Benjamin, Ruha (2019). Race after technology: Abolitionist tools for the new jim code. *Polity*.

Bundesministerium für Familie, Senioren, Frauen und Jugend (2021). Dritter Gleichstellungsbericht – Digitalisierung geschlechtergerecht gestalten. <https://www.bmfsfj.de/resource/blob/184544/c0d592d2c37e7e2b5b4612379453e9f4/dritter-gleichstellungsbericht-bundestagsdrucksache-data.pdf> (03.03.2022).

Criado Perez, Caroline (2019). Invisible women: Exposing data bias in a world designed for men. *Random House*.

D'Ignazio, Catherine & Klein, Lauren F. (2020). Data Feminism. *MIT Press*.

Eubanks, Virginia (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. *St. Martin's Press*.

Lenzen, Manuela (2020). Künstliche Intelligenz: Fakten, Chancen, Risiken (Vol. 2904). *CH Beck*.

netzforma* e. V. (2021). Wenn KI dann feministisch. Impulse aus Wissenschaft und Aktivismus. https://netzforma.org/wp-content/uploads/2021/01/2020_wenn-ki-dann-feministisch_netzforma.pdf (25.02.2022).

O'Neil, Cathy (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. *Broadway Books*.

Orwat, Carsten (2019). Diskriminierungsrisiken durch Verwendung von Algorithmen: eine Studie, erstellt mit einer Zuwendung der Antidiskriminierungsstelle des Bundes. *Nomos*.

Xenidis, Raphaele & Gerards, Janneke (2021). Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law. *European Commission, Directorate-General for Justice and Consumers*. <https://op.europa.eu/en/publication-detail/-/publication/082f1dbc-821d-11eb-9ac9-01aa75ed71a1> (25.02.2022).

Zweig, Katharina (2019). Ein Algorithmus hat kein Taktgefühl: Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können. *Heyne Verlag*.

Further sources

“HR-Puzzle” is an explanatory game about machine learning (ML) in human resource management, to help people understand how ML works: <https://algorithmwatch.org/de/hrpuzzle/> (25.02.2022).

AlgorithmWatch's explainer videos provide an easy introduction to important basics: proxy variables, correlation & causality, and accountability of ADM systems: <https://algorithmwatch.org/de/auto-hr/erklervideos/> (25.02.2022).

Atlas Lab is an English-language blog for lawyers on the topic of ADM systems and legal strategies: <https://www.atlaslab.org/> (25.02.2022).

A video by Nushin Yazdani and José Rojas explains, “how machines discriminate”: “Wie Maschinen diskriminieren – Facial Recognition & the Matrix of Domination”: <https://vimeo.com/476084538> (03.03.2022).

The comic essay “We need to talk, AI” by Julia Schneider and Lena Ziyal illustrates the basics, opportunities, and limitations of AI applications: <https://weneedtotalk.ai/> (09.03.2022).

A simulator, created as part of the DataSkop project, allows you to experience the basic principles of recommendation algorithms: <https://dataskop.net/recommender-sim/?en> (09.03.2022).

The messenger service unding.de sends concerns to major Internet companies and other entities when people feel unfairly treated by, or irritated with, ADM systems. You can find examples and surveys on the topic here: <https://unding.de/news/>.

The Verbraucherzentrale des Saarlandes e. V. offers an online simulation of data-based price adjustment in online retailing: <https://wasistdeinpreis.de/index.html> (03.03.2022).

“Coded Bias” by Shalini Kantayya is a documentary about Joy Buolamwini, a researcher at MIT Lab. She works on discrimination in algorithms and, among other things, revealed in a scientific study that common face recognition software works worse on faces with darker skin tones.

AI Blindspot is a project of The Assembly Program – Berkman Klein Center For Internet & Society at Harvard University. It offers English language workshop documents on “Tools for Advancing Equity in Artificial Intelligence”: <http://aiblindspot.media.mit.edu/AL.html> (25.02.2022).

The comic books by Julia Stoyanovich and Falaah Arif Khan look at accountability and fairness in data and AI applications with a U.S. connection and explain the basics in English: <https://dataresponsibly.github.io/comics/> (25.02.2022).

The Plattform Lernende Systeme provides a glossary with explanations of terms related to AI applications: <https://www.plattform-lernende-systeme.de/glossar.html> (25.02.2022).

