Federal
Anti-Discrimination
Agency

Risks of Discrimination through the Use of Algorithms

Carsten Orwat

Risks of Discrimination through the Use of Algorithms

A study compiled with a grant from the Federal Anti-Discrimination Agency

by Dr Carsten Orwat

Institute for Technology Assessment and Systems Analysis (ITAS)

Karlsruhe Institute of Technology (KIT)

Table of Contents

List of Tables

List of abbreviations

ACLU	American Civil Liberties Union
ADS	Antidiskriminierungsstelle des Bundes (see FADA)
AGG	General Equal Treatment Act (Allgemeines Gleichbehandlungsgesetz)
AI	Artificial intelligence
AMS	Public Employment Service (Arbeitsmarktservice)
Art.	Article
BGH	Federal Court of Justice (Bundesgerichtshof)
BDSG	Federal Data Protection Act (Bundesdatenschutzgesetz)
BDSG a. F.	Federal Data Protection Act, old version
BDSG n. F.	Federal Data Protection Act, new version
BVerfG	Federal Constitutional Court (Bundesverfassungsgericht)
BVerfGE	Decision of the Federal Constitutional Court (Entscheidung des Bundesverfassungsgerichts)
CNIL	Commission Nationale de l’Informatique et des Libertés
DDD	Défenseur des Droits
DM	Data mining
GDPR	General Data Protection Regulation
DSRL	Data protection directive (Datenschutz-Richtlinie)

ibid.	ibidem
i.e.	that is to say
EDPB	European Data Protection Board
ECHR	European Convention on Human Rights
FADA	Federal Anti-Discrimination Agency (Antidiskriminierungsstelle des Bundes. ADS)
e.g.	for example
et al.	and others
etc.	et cetera
EU	European Union
GG	Basic Law for the Federal Republic of Germany (Grundgesetz)
HUD	U.S. Department of Housing and Urban Development
IT	Information Technology
ML	Machine learning
NFHA	National Fair Housing Alliance
para.	Paragraph
PoC	People of Colour
ROC AUC	Receiver Operating Characteristic, Area under the Curve
WP29	Article 29 Data Protection Working Party, renamed European Data Protection Board (EDPB)
YVTltk	Yhdenvertaisuus- ja tasa-arvolautakunta

Acknowledgement and Funding

This study was funded by a grant from the Federal Anti-Discrimination Agency (FADA), to whom we would like to express our sincerest gratitude.

The author otherwise receives funds from the institutional funding of the Helmholtz Association via the Institute for Technology Assessment and Systems Analysis. The Helmholtz Association is funded by the federal and state governments.

Ms Nathalie Schlenzka from the Federal Anti-Discrimination Agency (FADA) provided helpful information and contacts through a query to European equality bodies. I would like to thank her and her colleagues for their constructive comments. I would also like to take this opportunity to thank the employees of the European equality bodies who answered questions about cases of discrimination and their working methods and competences. The answers have been included in the study at the points cited.

I would like to thank Oliver Raabe, Moritz Renftle, Oliver Siemoneit and Reinhard Heil for numerous in-depth and extremely productive discussions. Oliver Raabe also provided invaluable assistance in reviewing legal issues. I would like to thank Arnold Sauter, Christoph Kehl and Christian Wadephul for their helpful comments. Any remaining errors can be attributed to the author.

Note

The assessments, statements and opinions expressed herein are solely those of the author and do not necessarily reflect the official opinion of the Federal Anti-Discrimination Agency.

Summary

Algorithms: The study focuses on algorithms that are used for data processing and the semi- or fully-automated implementation of decision-making rules to differentiate between individuals. Such differentiations relate to economic products, services, positions or payments as well as to state decisions and actions that affect individual freedoms or the distribution of services.

Discrimination: Algorithm-based differentiations become discriminatory if they lead to unjustified disadvantaging of persons with legally protected characteristics, in particular age, gender, ethnic origin, religion, sexual orientation or disability. The study describes cases in which algorithm- and data-based differentiations have been legally classified as discrimination or which are analysed and discussed as risks of discrimination.

Surrogate information: Algorithm- and data-based differentiations often exhibit the characteristics of so-called statistical discrimination. Typical for this kind of discrimination is the use of surrogate information, surrogate variables or proxies (e.g. age) to differentiate, because the original distinguishing characteristics (e.g. labour productivity) are difficult for the decision-makers to determine by examining individual cases. These surrogate variables can be protected characteristics, or there can be correlations with them and protected characteristics. With algorithmic methods of data mining and machine learning, complex models with a large number of variables can be used instead of one or a few surrogate variables.

Societal risks: The legitimacy of such differentiations is often justified on the grounds of efficiency in overcoming information deficits. However, they also involve societal risks such as injustice by generalisation, treatment of people as mere objects, restriction of the free development of personality, accumulation effects and growing inequality and risks to societal goals of equality or social policy. When developing and using algorithms, many discrimination risks result from the use of data that describe previous unequal treatment.

Needs for societal considerations: Although overcoming technically-based discrimination risks of algorithms and data is fundamental, the legitimacy of different forms of algorithmic differentiation requires societal considerations and decisions that take into account the aforementioned societal risks, the benefits of differentiation and, in particular, their distribution in society. This should lead to definitions of socially acceptable differentiations. Since in most cases such differentiations are based on the processing of comprehensive amounts of personal data, risks to the right to informational self-determination must be considered as well.

Data protection law: The current data protection law needs clarifications and corrections that would also serve anti-discrimination purposes. These relate to so-called informed consent, where affected persons must assess far-reaching potential consequences, including possible unequal treatment, at the time of consent. This approach no longer seems adequate in light of the actual practices of collecting, merging, transferring, and using personal data and the emerging risks of unequal treatments based on these practices.

More detailed regulation of decisions: Furthermore, a more detailed regulation of algorithmic and data-based decision-making in addition to the current regulatory focus on data processing has been suggested. Anti-discrimination law, which sets out the protected characteristics that may be used in certain decision-

making situations, can already be seen as a regulation of decisions. Improvements in regulation can range from more detailed provisions on the permitted use of certain decision-making criteria, e.g. by clarifying exemptions from the prohibited use justified on objective grounds, or the use of recognised computation methods, to the prohibition of certain algorithmic and data-based differentiations for certain types of high-risk decisions. The prohibition of automated decision-making under data protection law can be improved in several areas. Regulatory instruments should be designed according to the specific level of societal risks attributable to different types of algorithm-based differentiations.

Tasks and duties of equality bodies: In particular, the difficulty in detecting and proving algorithm-based discrimination by persons concerned suggests, according to the principle of subsidiarity, that representative bodies should take action on behalf of the persons concerned and collective redress should be used.

- Many examples illustrate that detecting and proving discrimination with algorithms is also possible without direct inspection of the algorithm or “opening” the software system. Instead, evidence of unequal treatment or discrimination can be provided by collecting and investigating publicly available data on the outcomes of differentiation decisions, which are derived from the interactions and transactions of the services and products investigated. Where the outcomes cannot be determined, however, such a procedure has its limits.

- Investigations into the outcomes of algorithm-based decision-making along with requests and information from affected individuals and the media can also serve as starting points for equality bodies tasked with advising and supporting those affected by discrimination; this also applies to the internet and computerised decision-making. If there is not enough information publicly available on the outcomes of decisions, the access options and rights of equality bodies should be extended so t hat they can fulfil their mandate to identify and reduce discrimination.

- With algorithm-enabled, customised offers and services, it can be difficult for persons concerned to detect differentiations of persons and make the required comparisons in order to provide evidence of unequal treatment. Equality bodies generally have expertise and experience with regard to groups of people, situations and treatments that are prone to discrimination, the causes of discrimination, seemingly neutral criteria and correlations with protected characteristics. Such expertise can also be the starting point for systematic empirical investigations, anti-discrimination testing and algorithm audits.

- The entities using artificial intelligence algorithms and applications in automated decision-making in particular may be required, under the legal provisions, to assess potential discrimination risks and ensure that the algorithms can be explained; they may also be obliged to document the functioning of the algorithms, the decision-making rules and their impact on persons concerned, including possible discrimination. Such documentation should be accessible to equality bodies in cases of suspected discrimination, with the right of access being regulated by law.

- Other (potential) tasks of equality bodies include advising entities that develop and use algorithms for the prevention of discrimination and (mandatory) involvement in public procurement procedures of algorithm-based systems that are particularly prone to discrimination.

1. Introduction

Algorithms and extensive data sets are increasingly involved in decisions that not only have trivial consequences for people, but also influence their way of living and personality development to a larger extent. Algorithms produce conclusions and outcomes that are used by human decision-makers as an information basis for their decisions, or the implementation of decision-making rules is completely delegated to algorithms or the computer systems containing them.

Apart from some early analyses of bias and unequal treatment through the use of computer systems (Friedman & Nissenbaum 1996; Bruce & Adam 1989), the risks of discrimination in connection with information and communication technologies have only been comprehensively addressed in this decade. Particularly in the course of the big data development, researchers and policy-makers have pointed out the associated risks of discrimination (e.g. Crawford 2013; Dwork & Mulligan 2013; The White House 2014; US CEA 2015; FTC 2016; Schneider & Ulbricht 2018). Researchers and journalists have uncovered numerous cases of unequal treatment and discrimination resulting from the application of algorithms, which will be discussed below.

This study takes a problem-oriented view on the consequences of the use of algorithms for the differentiation of individuals. The study focuses on algorithms that are used to differentiate individuals with regard to differentiated information, products, services, payments, positions, etc.[1] Such differentiations can lead to unjustified unequal treatment or discrimination. It is therefore a societal task to define what unequal treatment is considered unjustified and to regulate unjustified unequal treatment. This task also concerns the forms of differentiation that are put into practice with the use of algorithms.

After a brief introduction to algorithms, relevant developments in data processing and algorithm-based differentiation (Chapter 2), relevant types of discrimination are presented (Chapter 3). Discrimination caused by algorithmic and data-based differentiation is mainly a reflection of statistical discrimination, which is also evident from the examples described in Chapter 4 The causes of risks of discrimination should not only be seen in the way algorithms and data sets are generated, selected and used, but also in the way that differentiations are used themselves (Chapter 5). This is followed by considerations of the required and potential actions whose primary aim is to prevent discrimination, and calls for societal consideration processes that extend beyond this context (Chapter 6).

[1] Given this problem-oriented view, the numerous non-discriminatory applications of algorithms are not taken into account. The study also does not deal with the effects of systems that, through the algorithm-based control of information, have an impact on information perception ("filter bubbles"), freedom of opinion, opinion formation or voting behaviour in democratic processes. Although examples from the field of government action are also given, the conclusions and the required or potential actions derived thereof apply mainly to the private sector.

2. Terms and Basic Developments

2.1 Algorithms

In this study, the term “algorithm” is used from an information technology perspective.[2] According to this, algorithms are basic, formalised and precisely defined computation rules or rules for a sequence of computation steps that are set up to execute a given task. For a computable task, such as sorting lists, there are often many different algorithms.

Algorithms have to be implemented or programmed in one of the many programming languages (such as Python, Java, JavaScript, C++, etc.) in order to be executed by a computer. They are subsequently available as programme parts, which are combined with data structures to form software or software systems. Algorithms in software then carry out the task of generating an output – usually in other data formats – from an input, usually in the form of data. When we talk about algorithms below, we always mean implementations of algorithms as possible components of software.

Algorithms are implemented in software, combined and organised to fulfil specific purposes. These purposes are primarily determined by humans, in particular software developers and contractors, who also determine the interpretations, values, prioritisations and exclusions contained in algorithms.[3] Their purposes and applications may also have unintended consequences[4] both for those directly affected and for indirectly affected third parties. This makes it more difficult to implement societal and fundamental values such as the protection of human dignity, the preservation of the free development of personality and informational self-determination, the avoidance of discrimination or the safeguarding of the rule of law.

Algorithms only display their societal consequences once they are deployed in software applications using specific data sets in economic, social, administrative and legal practices. Consequently, there is a greater emphasis in this study on the applications of algorithms in software systems for specific purposes than on algorithms per se.

Algorithms are currently the subject of much attention, especially in the social sciences and humanities, public life and politics, which can be explained by the fact that (a) IT systems with algorithms are used almost ubiquitously in all areas of life, not only in production processes, office applications and economic transactions, but also in social interactions and communication, (b) their widespread use for the automated handling of large amounts of data, accelerated interactions and transactions appears indispensable and

[2] For a discussion on the term algorithm and its delimitation, see e. g. Hill (2016), Cormen et al. (2010: 5-15), Mittelstadt et al. (2016) or Yeung (2017).

[3] Cf., e. g., Schinzel (2017), Zweig, Fischer & Lischka (2018) or Kitchin (2017).

[4] Here according to Brey (2000), (2009) and Kitchin (2017).

inevitable to some extent, due to economic network effects,[5] (c) they are increasingly used in decisions that have consequences on the life opportunities and personality development of people and (d) they are sometimes expected to have a certain ability to act and make decisions, or a shift of responsibility to algorithms is even implied. For this study, algorithms will be divided into several **types**:[6] (1) algorithms whose rules are developed entirely by human logic and whose rules are implemented as "direct programming", virtually "by hand" by the developers, and (2) data mining or machine learning algorithms whose rules are based on correlations generated by data analysis.

The latter are also referred to as "learning algorithms" and are usually assigned to the field of artificial intelligence (see Sections 2.2.2 and 3.4 for details).[7] Also for this type, "learning" does not take place without people, since entities developing and using machine-learning methods have to make many design decisions. In many cases, human decisions are also the source of risks of discrimination (see Chapters 4 and 5 for more details).

Algorithms are used today not only to automate data processing, including data analysis and inference, but also to automatically apply and enforce decision-making rules. For illustrative purposes, algorithms can therefore be further subdivided into (1) those for automatic data processing and analysis and (2) those that execute decision-making rules automatically[8] (Ernst 2017; Kleinberg et al. 2019). Many of the software systems of interest here contain both, but for later considerations it is useful to distinguish between them in their representation. After all, not every application of algorithms means a fully automated decision. Algorithms are also often only used for data analysis, whose outcomes are used as recommendations or support for human decisions.

[5] In the case of economic network effects, the benefit for a single user increases with the number of other users, as the possibilities for communication or exchange between the users increase. This can lead to concentration or monopolisation tendencies or the dominance of one system. For individual users who want to participate in the communication or exchange, the options can be drastically reduced and the use of one system can become almost inevitable. Network effects occur in particular in telecommunication networks, on "matchmaking" platforms (e.g. job market platforms, e-commerce platforms such as eBay, dating platforms) or "audience-making" platforms (e.g. social networks, search engines), in transaction systems (e.g. electronic payment systems such as PayPal) and on software platforms (especially operating systems). For an overview and discussion see e.g. Dewenter & Lüth (2018).

[6] Here according to Lehr & Ohm (2017), Selbst & Barocas (2018).

[7] Unfortunately, in the discussion about algorithms, it is often the case that the term "algorithms" only refers to the machine-learning algorithms. However, this meaning is not pursued in this study, given that risks of discrimination can also arise from the use of a variety of algorithms.

[8] For more details, see Section 2.3.3.

2.2 Developments in data processing

2.2.1 Increasing the amount of data relating to an identifiable person

In recent decades, the amount of personal data and data relating to an identifiable person that is generated as a product or by-product of computerisation (now increasingly referred to as "digitisation"), not only by and between organisations but also in public and private spheres of life, has grown significantly. The collection of usage, location and movement data from mobile devices, the recording of the diverse uses of the internet, such as communication in online social media, search engine queries, website visits and evaluation of browser histories, the use of online commerce and other internet services (e.g. streaming services) and electronic financial transactions and payment systems are all significant sources of personal data (Christl & Spiekermann 2016; Christl 2017; Constantiou & Kallinikos 2015; Weichert 2013; Pasquale 2015). The collection and commercial analysis of personal data are often obtained in return for the "free" use of internet services such as search engines or online social media ("personal data as counterperformance") (EDPS 2017). Moreover, due to the progressive spread of so-called "smart" devices ("smart homes", "smart cars", "wearables", "fitness trackers", "personal assistants", virtual assistants or speech assistants, etc.) or the implementation of the "Internet of Things", extensively equipped with sensors and networked objects, very large volumes of personal data are collected whenever products and services are used.

This increase in the amount of personal data has led to several **consequences** in terms of differentiations and risks of discrimination that are relevant to this study. The growth in volume of available data enables many differentiations of persons through the use of algorithms. In particular, the consolidation of data – either within the company or with the help of data trading or data brokerage – enables differentiation on the basis of extensive personal profiles. New forms of data analysis, and much of machine learning in particular, only work meaningfully when based on large volumes of data.

2.2.2 Expansion of algorithm-based analysis methods

The following section outlines some of the developments that are particularly relevant to this study and which have taken place over several decades and are still ongoing. It should be noted that some of the terms used may relate to the same developments with a large degree of overlap, or the terms may describe developments at different levels of representation, and therefore cannot be clearly distinguished from one another.

The purpose of the **data mining methods** is to identify findings or statistical correlations in large data sets (Custers 2013; Calders & Custers 2013; Linoff & Berry 2011: 2). Automated procedures that reveal patterns or regularities in the data sets are characteristic here and also – unlike in classical statistical analyses – that there are no hypotheses to be tested about possible correlations between variables as the basis for the procedure. The outcomes, which are generated as a set of calculated relationships, are also called models. Outcomes or models can be used to create classes or categories to which people can be automatically assigned. In many cases, the formation of categories is based solely on the automated generation of correlations (Barocas & Selbst 2016: 677).[9]

[9] See details in Section 3.4 and Section 5.1.

The developments that have taken place under the vague umbrella term **"big data"** are mainly aimed at the merging and processing of large and different data sets, i.e. which may be heterogeneous in their formats (e.g. numbers, images, text, video and audio formats) and which have been collected in different contexts. Algorithms serve here primarily for data analysis. If personal data is used, algorithms from big data analytics help to create and maintain comprehensive personality profiles that are often used to predict behaviour (e.g. expected buying behaviour). In a narrow, technical sense, big data techniques or big data analytics are primarily used for the automated collecting, processing, management and analysis of large volumes of data (Chen, Mao & Liu 2014).

In a broad sense, the term "big data" also encompasses the organisational arrangements, practices and business models in which the connecting processing of large, sometimes heterogeneous volumes of data plays the central role and is primarily aimed at forecasts and reactions in real-time (e.g. Zuboff 2015; see also Kolany-Raiser et al. 2018; Hoeren & Kolany-Raiser 2018).

The transition between data mining and **machine learning**, which is usually considered a subset of **AI**,[10] is fluid. Machine learning is a vague umbrella term for very different concepts and methods, which may even include conventional statistical analysis methods.[11] The term machine learning refers to procedures for automatically finding correlations – also known as relationships, regularities or patterns – between variables in a data set. In doing so, one tries to reproduce the human process of learning by using machine methods (but not without the involvement of humans) to identify the relevant patterns or characteristics for the object to be analysed from a large number of examples in the form of learning or training data sets. The patterns or characteristics are generated as a model. In most cases, machine-learning methods are used to generate predictions or estimates of outcomes (according to Lehr & Ohm 2017: 671). Together with the further increase in the performance and cost reduction of computers and the widespread use of cloud data centres, the growth in volume of personal data available in sufficient quality and quantity has contributed to the fact that machine-learning systems are now being applied to an increasing number of situations involving people. There are many such applications aimed at pattern recognition in databases (e.g. patterns of fraudulent behaviour in financial data), computer vision or text or audio processing of speech, including natural language processing. In the field of computer vision, facial recognition systems are particularly noteworthy; these are not only designed to recognise people, but also to recognise states (e.g. emotional states, see below) and behavioural patterns of people.

Machine learning overlaps with other methods of statistics and data processing like data mining. The basic algorithms, which are combined to new systems with AI, have in some instances been known for decades (e.g. regression methods). In other instances, new learning algorithms have been added. In the latter case, the so-called "artificial neural networks", in particular "deep learning",[12] have attracted attention in the last few years.

[10] A generally accepted definition of artificial intelligence (AI) has yet to be established. Many authors describe AI as a technical means to reproduce human intelligence.

[11] For example, see Domingos (2012), The Royal Society (2017), Jordan & Mitchell (2015), Alpaydin (2016), Leis et al. (2018), Mullainathan & Spiess (2017), Strauss (2018), WIPO (2019).

[12] Deep learning methods belong to the procedures of "artificial neural networks" and use several "node layers" of computational stages ("neurons"), which are connected to each other in a weighted manner in order to identify patterns or correlations in a data set (e.g. a digital image of a dog). Each layer is dedicated to learning with a different level of abstraction. During the training process, the weighting of the links is adjusted. In the example of the dog image, the bottom layer learns simple details such as the values of a pixel, the next layer up tries to learn edges and upper layers learn to interpret the combination of edges as a dog's nose, for example. Cf. Alpaydin (2016), Beck et al. (2019).

Data mining, big data analytics and machine learning can be used for profiling. The term **profiling** covers algorithm-based techniques and practices of processing large volumes of data to create and link, update and use records on natural persons in order to produce a comprehensive picture of a person or group of persons, which is mainly used for categorisation, assessment, forecasting and decision-making. Profiles are often created in order to form categories based on known characteristics or persons and correlations, in which the category affiliation of unknown persons is deduced by identifying common characteristics. Best known examples of the use of profiling are law enforcement, border controls, commercial and governmental web tracking, marketing and insurance (Hildebrandt & Gutwirth 2008; van Otterlo 2013; Helberger 2016; FRA 2018; Hänold 2018).[13]

Scoring is understood as the assignment of numerical values to persons, usually with the assignment of persons on a scale and by calculating probability values for a certain future behaviour (ULD & GP Forschungsgruppe 2014; Dixon & Gellman 2014; Weichert 2018; in detail SVRV 2018). Applications of scoring, which are currently attracting particular attention with regard to possible unfair treatment or discrimination, are credit scoring in the granting of loans, scores of labour market opportunities in the employment service or risk scores in the judiciary system.

The above-mentioned data processing procedures are increasingly used for personalised data **forecasts**. Unlike ex-post evaluations, which examine whether or not a target value or target state of a differentiation objective has been achieved, forecasts are designed to classify people ex ante into classes formed according to a differentiation objective on the basis of calculated probabilities, or to assign individual values to express how likely it is that certain states will be achieved in the future. In this study, the focus is mainly on algorithms for forecasting.

In particular, developments in machine learning as a sub-area of AI have enabled analyses of the automated **identification of personality traits** – for example, the health status or sexual orientation of a person – to be carried out in an increasing breadth, with (presumed) better accuracy, in (near) real time and on the basis of previously unusual data material. For example, while the recording of the personality trait “trustworthiness” when determining credit scores was mainly based on payment history and other financial information for many years, credit scores are now also generated using data on communication and relationships in “social” online networks (Wei et al. 2016).[14] In addition, there are many descriptions and experiments by researchers and developers on the identification of personality traits, including:

- The recognition of emotional states using keyboard strokes (Epp, Lippold & Mandryk 2011);
- The derivation of sensitive information (including health status) from telephone metadata (Mayer, Mutchler & Mitchell 2016);
- The determination of naivety or sophistication and its use in loan offerings (Ru & Schoar 2016);

[13] Article 4 para. 4 GDPR provides a legal definition of profiling, which refers to profiling as “any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person’s performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements [...]”.

[14] One of the international leaders in credit scoring, Lenddo, claims to process data from online social networks, telecommunications data, browser data, mobile data, data from e-commerce and financial transactions, data from the analysis of application completion and psychometric data. Cf. https://lenddo.com/ (last retrieved on 17 April 2019). See also the patent filed by the company Facebook, which describes a method of calculating the creditworthiness of people from the credit scores of “associates” in the network (Meyer 2015).

- The recognition of emotions and development of psychodemographic “profiles” based on data from the online network Twitter (Volkova & Bachrach 2015);
- The detection of criminal inclination (Wu & Zhang 2016) and genetic diseases with automated facial recognition (Gurovich et al. 2019);
- The determination of sexual orientation on the basis of Facebook contact lists (Jernigan & Mistree 2009);
- The determination of “racial” or ethnic origin on the basis of personal pictures (Fu, He & Hou 2014);
- The identification of psychological characteristics (extroversion, introversion, openness to innovation) from “digital footprints”, such as “likes” or posts on the online network Twitter (Matz et al. 2017);
- The determination of various personality traits, such as sexual orientation, ethnicity, religious and political attitudes, age, gender or intelligence from “likes” on Facebook (Kosinski, Stillwell & Graepel 2013) or;
- The recognition of sexual orientation, especially homosexuality, from images of persons (Wang & Kosinski 2018).[15]

The extent to which such methods of analysis are already being used in practice is in most cases unclear, as there is no systematic survey of this. However, what these examples make clear once again is that there is no such thing as “unimportant” data (belangloses Datum) – as the Federal Constitutional Court (BVerfG 1983) enshrined in 1983 – and that seemingly “harmless” communication and behaviour can potentially provide the basis and criteria for unequal treatment and discrimination.

[15] See also Matz & Netzer (2017) for overviews on psychological analyses based on big data techniques, or Yue et al. (2018) on sentiment analysis/sentiment detection.

2.3 Algorithmic and data-based differentiations

2.3.1 Types of differentiation

Algorithms are used within applications of statistical data analysis, data mining, big data analytics and machine learning for differentiating individuals, either to enable new forms of differentiation or to rationalise or refine existing differentiations.[16] Differentiation of persons means the division of a population of persons by assigning them to classes, categories, groupings or (market) segments, or by identifying "outsiders". At the same time, different objects of differentiation are provided or carried out for the differentiated groups of persons or individuals. An extreme form of differentiation is individualisation, i.e. differentiation targeted towards a single individual. Fundamentally, the individuals or groups of individuals need to be identified by using the wide range of options available today for processing personal data and drawing conclusions from it (Gandy Jr. 2010: 30). Differentiations are aimed at providing information, goods, services, payments, positions, granting of freedoms etc. to specific groups or individuals (see Table 1).

Table 1: Objects of algorithmic and data-based differentiations

Objects	Examples
Information	Website content, advertising, search results, (partner) contacts
Products and services	Goods, including information or media products, real estate, insurance, loans, education, medical treatment, infrastructure services
Payments	Prices, premiums, tariffs, interest, wages, wage replacement benefits
Development opportunities and positions	Training, working positions, working conditions, offices, other positions
Freedoms	(Non)restrictions (detention, control, penalties)

Source: own compilation

There are many different **reasons** behind differentiations and the resulting unequal treatment of persons: (1) in order to manage different risks of persons and groups of persons, risk classes or risk measures (e.g. risk scores) are determined (e.g. risk of recidivism if a crime was committed, risk of credit default, risk of job change, risk of unsuitable staffing); (2) in order to determine the different value of the clientèle or strategic value of persons in economic relations and to use them to generate profit, classes and measures of economic potential (e.g. labour productivity, demand behaviour, sales potential or willingness to pay) are formed; (3) similarly, persons may be differentiated on social grounds, for example, on the basis of need, distribution of opportunities or solidarity, as is the case with special awards for students or pensioners and support programmes for certain groups of persons.

[16] Cf. also Mittelstadt et al. (2016). For business applications in the context of marketing or business relationship management see, e.g., Vercellis (2011).

Algorithm-based differentiations can also serve to differentiate **behaviour controls**.[17] A distinction must be made between "hard" and "soft" behaviour controls. "Hard" behaviour controls are aimed at the technical exclusion of rule deviations and effective rule enforcement, e.g. through programmed access and usage rules for media products with digital rights management or programmed contract components for smart contracts or other block chain applications. "Soft" behavioural controls take place through financial incentives, information provided, recommendations, other "nudges", etc.[18] or "dark patterns",[19] which can be offered and administered with algorithms. These include, for example, incentives to reduce risky behaviour through personalised insurance rates, the selection of recipients of certain information or the design of the user's options, such as personalised advertisements on websites or product recommendations in online trading. A characteristic of algorithmic systems is that they execute behavioural controls automatically on a large scale, i.e. not only for a single user, but for the entire number of relevant users (Yeung 2017, 2018: 19, 29).

Economic differentiations through the design of products, services and payments relating to groups or individuals have a long tradition in market economies. Nevertheless, the societal effects and the acceptability of such differentiations are the subject of controversy and repeated discussions. Algorithmic and data-based differentiations bring these controversies about the advantages and disadvantages of differentiations to light. One of the **advantages** is that differentiated information, products or services can better meet the different preferences of demand. This can not only increase the demand, customer satisfaction and the identification of members of the clientèle with the offers and goods, but also reduce the costs of advertising by avoiding "unused" information or wastage in advertising. In principle, lower prices can also be conferred to certain groups. However, economic differentiations are also subject to criticism regarding a number of **disadvantages** that they bring – in particular, that they take unreasonable advantage of the willingness to pay and unilaterally reduce the consumer surplus.[20] Differentiations, especially in the form of individualisations, can also limit the choices of the individual and thus their autonomy (Barocas &

[17] For a discussion based on various keywords, see "Lex informatica" Reidenberg (1998), "Code is Law" Lessig (1999), (2006), "Regulation by Software" Grimmelmann (2005), "Regulation by Design" Yeung (2008), "Software as Governance" Shah & Kesan (2010), "Regulating Code" Brown & Marsden (2013), "Governing Algorithm" Barocas, Hood & Ziewitz (2013), "Governance by Algorithm" Just & Latzer (2016), "Algorithmic Regulation" Medina (2015), Yeung (2017), Hildebrandt (2018), "Governing through Technology" Kallinikos (2011), "Verhaltenssteuerung durch Algorithmen" (Behavioural Control through Algorithms) Hoffmann-Riem (2017) or "Software als Institution" (Software as an Institution) Orwat et al. (2010), Orwat & Bless (2016). An overview of relevant research strands can be found in von Grafenstein et al. (2018), for example.

[18] "Nudging" refers to entrepreneurial and political measures and instruments for influencing behaviour, which are mostly based on findings from behavioural research. They attempt to take the behavioural characteristics of people as a starting point and to prescribe their "decision-making and selection architecture". This usually occurs without restricting their freedom of choice. The latter distinguishes them from commandments or prohibitions. Sometimes they also covertly target unconscious behavioural traits or exploit certain human characteristics, such as the tendency to avoid extra effort or to follow social norms and expectations. Examples include the default settings of computer systems or online services, specific ways of presenting or arranging goods (e.g. healthy food before unhealthy), pleas that appeal to social norms, certain ways of representing information on websites, or political or business measures designed as games ("gamification"). The boundaries between the longer known financial incentives or information policy measures are not always clear and may include them. Cf. Sunstein (2014), Smeddinck & Bornemann (2018), von Grafenstein et al. (2018).

[19] "Dark patterns" are practices of designing the elements of computer systems, online services, e-commerce platforms or websites at the interface with the users, which attempt to steer users towards unintentional behaviour and decisions that are harmful to them. In terms of content, there is a strong overlap with "nudging", whereby "dark pattern" refers primarily to the manipulative practices that lead to harm among the users. Examples include tracking-intensive default settings, restricting choices or making it more difficult to choose privacy-friendly settings, forcing registration, hiding cost information or providing information that creates a sense of urgency about a decision. Cf. Forbrukerrådet (2018), Mathur et al. (2019).

[20] The so-called consumer surplus is the difference between the price that a consumer would be willing to pay for a product or service, and the market price that has actually formed for the product or service in the market. The greater the difference, the greater the financial benefit for the person on the demand side. One of the objectives of price differentiation is to impose higher prices on those customers who are willing to pay a high price for the product or service.

Nissenbaum 2014: 54). The question of whether and how risks of discrimination can arise from differentiation will be considered in the course of the study.

Overall, technical, methodological and organisational developments have led to the fact that algorithmic and data-based differentiations can be carried out at lower costs, in a finer degree of detail and above all along new features, such as presumed determined characters and personality traits, compared to conventional forms of differentiation (Agrawal, Gans & Goldfarb 2016, 2018). For example, developments in digital technologies, including the internet, have reduced the cost of identifying and tracking individuals' behaviour, characteristics and conditions, and have enabled better verification of identities (Goldfarb & Tucker 2017). Technically, the various possibilities of more or less unnoticed tracking[21] (e.g. visits and behaviour from websites, in social networks and online trading platforms, when using apps), registrations and user accounts (and/or accounts or logins) serve this purpose. Identification of persons not only means the recognition of an individual; to an increasing extent, it also means the identification of characteristics and conditions such as age, emotional states, social status or sexual orientation. For applications with continuous data streams (e.g. posts on online social networks) and their analysis, the providers can also implement differentiations with constant adjustments and experiments within the business models (Varian 2014).

Differentiation or personalisation related to groups or individuals can be better implemented by computer-based means, and better online than offline, not only because that the necessary database of identified persons or groups and their behaviour or states is available or can be obtained, but also because the separate group or person can be addressed in a more optimal manner from a technical standpoint. This is because information and communication technologies and automation have also reduced the costs of adaptation, such as menu costs or the costs of presenting information in a way that is adapted to people (e.g. Varian, Farrell & Shapiro 2004: pp. 12ff.). Personalisation can then be achieved, for example, on websites or in apps, in that the persons concerned only perceive "their" offer or the decision relating to them and do not have the opportunity to make direct comparisons. Comparability can be improved, for instance, with the help of comparison portals or by exchanging information with other users.

2.3.2 Scope of application

In the meantime, a wide range of international applications can be found in which systems with algorithmic and data-based differentiations are used, which have consequences on the way people live and their opportunities for development.[22] Chapter 4 provides examples of cases of unequal treatment and risks of discrimination.

When it comes to **working life**, algorithms and data-based differentiation occurs within the context of selecting job applicants and determining different salary levels and working conditions for employees. The systems used for this purpose are called "talent analytics", "people analytics", "workplace analytics" or "human resources analytics". The personal data collected includes data from job seekers' application documents, from work processes including communication, and from the work results of the employees. In addition to determining the applicant's suitability for the hiring organisation, the purpose of the systems is to check compliance with work regulations and organisational guidelines, to determine productivity for sanctions or rewards or promotions, and to determine the likelihood of absenteeism or leaving the

[21] See, for example, Klebert et al. (2012).

[22] Lischka & Klingel (2017), Spielkamp (2019), Matzat et al. (2019) also provide overviews.

company.[23] An example of "talent analytics" is the digital recording of job interviews, which are evaluated using machine-learning methods of "social sensing" (Chamorro-Premuzic et al. 2017). Artificial intelligence algorithms are used for pattern recognition in the machine review of digitally available application documents, for voice and word choice analysis in (electronic) job interviews or for the recognition of certain facial expressions (e.g. when lying) in video job interviews. Another area is online platforms such as "social" online networks or online job market services, on which differentiated job advertisements can be placed, but which are also used to manage and evaluate the provided services by the service users or the platform company.[24]

Within **trade**, attempts are being made with algorithmic and data-based business practices to achieve price differentiation in various forms, in addition to targeted product recommendations and advertising (e.g. Lecuyer et al. 2015).[25] Only a few examples of price differentiation with individualised prices for single individuals using personal data[26] can be found in some online shops, more common is the granting of individual discounts, premiums or coupons in customer loyalty programmes (e.g. Payback) (US CEA 2015; Schwaiger & Hufnagel 2018). Price differentiation can take a further form by creating different versions of a product or service ("versioning") and different market segments ("market segmentation"), on which products and services are offered at different prices depending on different quantities sold, at different times of the offer or in different (quality or feature) variants. Usually, the demanders assign themselves to certain market segments (self-selection), or can choose between different versions.[27] Here, algorithmic procedures play a role above all in data analysis for the formation of market and customer segments, which can also be performed with anonymised data. In a third form of price differentiation, providers charge different prices for different groups (e.g. senior citizens' discounts).[28] Furthermore, algorithms and computer systems are used in customer management to calculate the so-called "customer lifetime value", the economic value of a person as a customer over the entire (potential) life cycle of the customer relationship, in order to determine individual or group-related offers and advertising, to prevent customer churn or to selectively terminate relationships ("demarketing") (Blömeke & Clement 2009; Vercellis 2011).

In the **banking industry**, new forms of credit scoring are applied, in which the data basis for credit score creation is expanded and new analysis methods are applied. While risk scores have long been used in lending, the current discussion is mainly focused on the extent to which the current regulation adequately addresses the risks of the new procedures.[29] In the **insurance industry** differentiated insurance tariffs are offered which are based on new methods of collecting and analysing personal data. These include telematics tariffs for motor vehicle insurance, which include the recording and analysis of individual driving behaviour.[30]

[23] See Rosenblat, Kneese & Boyd (2014), Burdon & Harpur (2014), Marler & Boudreau (2017), Chamorro-Premuzic et al. (2016), 2017), Dzida (2017), Weichert (2018): 59-61), Angrave et al. (2016), Kornwachs (2018), von Grafenstein et al. (2018: 25-26).

[24] See overview e.g. in Bogen & Rieke (2018).

[25] For a discussion of data-based price differentiation see US CEA (2015), Miller (2014), Ezrachi & Stucke (2016), Steppe (2017), Acquisti, Taylor & Wagman (2016), Zuiderveen & Poort (2017), Christl & Spiekermann (2016: pp. 41ff.), Schwaiger & Hufnagel (2018), Zander-Hayat, Reisch & Steffen (2016), Tillmann & Vogt (2018a), (2018b). For a discussion of price differentiation by gender and from the perspective of anti-discrimination law, see an der Heiden & Wersig (2017).

[26] Also known as first-degree price differentiation. It is generally criticised that personal data is required for its implementation and that the privacy of the data subjects is curtailed. Cf. Varian, Farrell & Shapiro (2004: 14).

[27] Also known as second-degree price differentiation.

[28] Also known as third-degree price differentiation.

[29] See Citron & Pasquale (2014), Weichert (2014), ULD & GP research group (2014), Hurley & Adebayo (2016), Ferretti (2017), Wei et al. (2016), Christl (2017), Eschholz (2017), Dorfleitner & Hornuf (2018).

[30] See, for example, SVRV (2018), Hänold (2019).

In **health care**, behavioural tariffs of (private) health insurance companies, in which personal data in the form of transaction data or vital parameters are recorded via wearables or smartphones via apps. In the discussion, not only legal concerns are raised, but also ethical ones, such as negative effects on the principle of solidarity and redistributive effects.[31] Applications using algorithmic procedures, including machine learning, in particular for the analysis of image material, are found in medical diagnostic procedures. Systems are also used to assign patients to specific treatments and programmes.

In the **public domain**, systems are used for border control and predictive policing,[32] to support court decisions, for the surveillance of public spaces, to identify potential criminals or terrorists and to manage social benefits, schools and universities or study places.[33] According to Matzat et al. (2019: 28), various decision-making support systems are currently being tested or are in the pipeline in the job centres of the German Employment Agency (Arbeitsagentur).

From the above-mentioned algorithmic and data-based differentiations in the various areas of life, it does not necessarily follow that discrimination takes place. However, Chapter 4 gives examples of cases of unequal treatment and discrimination in individual areas of life.

2.3.3 Automated decision-making

The use of the term "automated decision-making" has become common in scientific discussion and (legal) practice. The term addresses both the use of algorithms for decision-making support of human decision-makers and the automated execution of decisions, although these are not always clearly differentiated from each other. For both types, the terms "automated decision-making systems" (ADM Systems) or "automated decision systems" are also used (e.g. Zweig, Fischer & Lischka 2018; Zweig 2019).

For illustration purposes, the decision-making process can be divided abstractly into several steps, ranging from the recording of the outcomes of the data analyses, the evaluation of the situation and the alternatives including the reconciliation of predefined conditions, the selection between alternatives, to the triggering of an action.[34] In a fully automated decision, all steps of the decision-making rules are executed by software. Humans are the ones to set the decision-making rules or, in the case of machine-learning methods, algorithms generate parts of the decision-making rules based on the analysis of data. Examples of fully automated decision-making systems are automated (online) bank lending, negative selection in systems for managing applications in the personnel sector, recommendation systems in electronic commerce, automated price adjustments, application and processing procedures in the insurance sector, spam filters in email programmes or (expected) automated administrative acts in public authorities, such as fully automated tax assessment notices (Busch 2018; Weichert 2018; Straker & Niehoff 2018; Hänold 2019).

How data analysis and decision-making processes relate in practice is very different and can, for illustration purposes, be divided into several types: (a) automated data processing and the decision-making process are separated and the outcomes of the data processing are virtually "manually" transferred to automated

[31] See Weichert (2018), German Ethics Council (2017), ten Have (2013), Christl & Spiekermann (2016: pp. 35ff.), Arentz & Rehm (2016), Bitter & Uphues (2017), Swedloff (2014), Selke et al. (2018).

[32] On predictive policing see, for example, Merz (2016), Robinson & Koepke (2016), Selbst (2017), Richardson, Schultz & Crawford (2019).

[33] Overview e.g. in Spielkamp (2019).

[34] See also Parasuraman & Riley (1997: 232) (with further references), which indicate that automations are to be understood rather as specific manifestations within a spectrum between the extremes of manual handling on the one hand and complete automation on the other, where the machine controls all aspects of the function. Similarly, Cummings (2004a), Vercellis (2011: 25-28). For the decision-making process, see similarly Kornwachs (2018: 174-179).

decision-making processes or programmed there as decision-making rules or (b) the data processing is integrated into the decision-making process. With data mining and machine learning methods, for example, the outcomes in the form of optimised models can be directly embedded in decision systems as programme components as rules of differentiation (Barocas & Selbst 2016: 677; Lehr & Ohm 2017; Kleinberg et al. 2019). The distinction between these two types is important for the identification of discrimination, since in the case of the latter, the outcomes are often less comprehensible.[35] For further considerations, an additional distinction is to be made between: (a) static systems, which perform data analyses once or at intervals separated by time and adapt the decision-making rules, and (b) dynamic systems, which constantly adapt and optimise the decision rules or models by continuously analysing data streams (e.g. Yeung 2017).

A precise distinction between decision support by automated data processing systems and fully automated decision execution is not only important from an ethical perspective[36] nor only necessary in terms of attributing responsibility, but also from a legal point of view. This is because, in principle, Article 22 para. 1 GDPR prohibits individual decisions based solely on automated processing, i.e. those that are made without human intervention (see Section 6.2.3, also on exemptions and permitted forms of application). In addition, there may be a tendency in practice that even in decision support systems, human decision-makers – particularly for reasons of efficiency, assumed neutrality or higher objectivity of computer conclusions, or because they find it difficult to justify any deviation from computer recommendations to superiors – to adopt the computer recommendations directly. As a result, decision support systems also tend to almost acquire the character of systems of fully automated decision execution.[37] The advantages of semi- and fully automated decisions are seen in efficiency gains, avoidance of errors and prejudices in human decisions and enabling offers or reactions in "real time". In addition to risks of discrimination, risks of automated decisions include potential manipulation, shifting of responsibility, lack of traceability and contestability by data subjects (Mittelstadt et al. 2016; Busch 2018; Weichert 2018; Ernst 2017: 1027-1029; Zarsky 2016).

When **comparing human and automated decisions**, it should be noted that many human decisions can be influenced by prejudices, stereotypes or other biases. Here, the initial expectation is that automated decisions made via computer systems are more "neutral" and "objective", since decision-making rules can be executed without human emotions or subjective preferences, and algorithms can process much more information in a decision-making situation and thus make "better" decisions. However, it will be shown below that the expectations of greater rationality and neutrality are not necessarily fulfilled and that partially or fully automated decision-making systems may also give rise to new risks of discrimination.

Another important distinction is the number of decisions that are made. Where human decision-makers – such as administrators or company employees – have some scope for decision-making within decision-making rules, the discrimination that occurs there may be limited to one or a few persons, depending on the number of discriminating employees.

In the case of automated differentiation decisions that have a potential to discriminate, all decisions taken by the system have the risk of discrimination. Risks of discrimination can thus become a mass phenomenon and easily lead to cumulative disadvantages.[38]

[35] See also Section 3.4.

[36] However, Wiegerling, Nerurkar & Wadephul (2018) emphasise from an ethical perspective that such decision-making systems are not really a matter of the system "deciding", since such systems do not have an understanding of responsibility for consequences and do not pursue their own intentions.

[37] A similar phenomenon is "automation bias", which leads to people trusting the answers provided by computers more than their own assessments, e.g. Cummings (2004a).

[38] See Gandy Jr (2010).

3. Discrimination

3.1 Terms and understanding

What actions are considered discriminatory is viewed differently in different societies, eras and regions. The demarcation is the outcome of societal conflicts, negotiations and agreements. These societal enshrinements take place above all in human and fundamental rights as well as in the laws and institutions that substantiate and enforce human and fundamental rights.

This study follows the common understanding of discrimination in the EU and the Federal Republic of Germany and understands discrimination as disadvantageous, unjustified unequal treatment of persons in connection with a protected characteristic.[39] The unequal treatment is based on the categorisation and attribution of characteristics to persons. The categorisation and formation of characteristics can, for example, be based on stereotyping, prejudices or rational calculations, be hidden in rules and practices or be unintentional. Various legal catalogues define categories and characteristics as legally protected characteristics – synonymously called discrimination grounds – according to which persons must not be disadvantaged in unjustified ways. The most important are summarised in Table 2. Unjustified primarily means that there is no objective reason or objective justification for the unequal treatment. In other words: unequal treatment in itself may also be acceptable from a societal point of view if a recognised objective reason[40] exists for this (see below).

[39] The term “discrimination” is used here, as in general usage in German-speaking countries as well as in European and German federal law with its negative connotation as societally undesirable unequal treatment or disadvantaged treatment of persons, see e.g. Berghahn et al. (2016: 25). On the other hand, the term “discrimination” is often used in English literature, in particular in scientific papers, to describe any form of “distinction” or “differentiation” of persons, which also may be societally beneficial and acceptable.

[40] Cf. Berghahn et al. (2014: pp. 57ff.), Schrader & Schubert (2018: AGG Section 3 points 68ff., Sections 8, 9, 10 and 20).

Table 2: Legally protected characteristics[41]

Protected characteristic	Article 3 GG	Sections 1 and others AGG**	Recital 71 GDPR	Article 9 GDPR
"Race" or ethnic origin	yes	yes	yes	yes
Ancestry, home country, origin	yes			
Gender	yes	yes		
Language	yes			
Political opinion or viewpoint and other opinion	yes		yes	yes
Religion and belief	yes	yes	yes	yes
Disability	yes	yes		
Age		yes		
Trade union affiliation	yes*		yes	yes
Genetic characteristics or dispositions and health status	yes		yes	yes
Biometric characteristics				yes
Sex life, sexual identity or orientation		yes	yes	yes

Source: own compilation. * According to Article 9 para. 3 GG ** There is a graded use of the characteristics, e.g. "political world view" does not apply in the civil law section (see e.g. Wersig 2017)

The objects of differentiation mentioned in the previous Section 2.3 (e.g. goods, real estate, positions, etc.) are potentially eligible as objects of discrimination and risks of discrimination, but not all of them are regulated by law. Differentiation of products and services always also means that one person or group may be deprived of something or access may be made more difficult, while access may be made easier for another person or group. Price differentiation means that individuals may be hindered in their access to resources, goods and services that serve the economic, social and cultural development of personality or the development of skills. Disadvantages are then expressed in concrete economic losses, e.g. in the form of loans or prices for products and services, or denial of access to opportunities for personal development or development opportunities, such as employment, housing or educational opportunities. Even the differentiation of information can be problematic from an anti-discrimination perspective if the information relates to goods or positions (e.g. jobs) that serve personal development, social integration or political participation (e.g. information and information technologies such as internet access). Last but not least, differentiated information or non-information can limit the options available.

[41] Other catalogues of protected characteristics can be found in the EU Charter of Fundamental Rights and the European Convention on Human Rights (ECHR), which also cover "property" and "birth", as well as the open "other status" clause in the ECHR.

3.2 Types of discrimination

Different types of discrimination are distinguished according to the purpose of research, discourse, discussion, and anti-discrimination decisions and measures. The most common distinction is between direct and indirect discrimination.[42] According to Section 3 para. 1 of the General Equal Treatment Act (Allgemeines Gleichbehandlungsgesetz, German abbreviation AGG), **direct discrimination** occurs "[...] where one person is treated less favourably than another is, has been or would be treated in a comparable situation on any of the grounds referred to under Section 1" (translated by the Federal Anti-Discrimination Agency 2009). Section 1 AGG itemises the reasons or characteristics including "race"[43] or ethnic origin, gender, religion or belief, disability, age or sexual orientation (see also Table 2). Pursuant to Section 3 para. 2 AGG, **indirect discrimination** "shall be taken to occur where an apparently neutral provision, criterion or practice would put persons at a particular disadvantage compared with other persons on any of the grounds referred to under Section 1, unless that provision, criterion or practice is objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary" (translated by the Federal Anti-Discrimination Agency 2009).

A further distinction between **taste-based** and **statistical discrimination** is geared towards the different motivations of the decision-makers who make a differentiation (Lorenz 1993). In the case of taste-based discrimination, unequal treatment is based on the personal, prejudiced dislikes or preferences of the decision-makers against or for a certain group of people or on dislikes or preferences for certain products (Becker 1957/1971; for criticism see Arrow 1998). Taste-based discrimination can be based on affective likes and dislikes. It can also be based on other rationales, such as welfare state redistribution targets, for example maximum age limits for professors that are not based on statistical evidence but are rather a redistribution measure in favour of younger professors or candidates (Britz 2008: 23).

Risks of discrimination that can occur through the use of algorithms and data sets in differentiating individuals often have the character of statistical discrimination (Calders & Žliobaitė 2013: 53; Barocas & Self 2016: 677, 688-692; Goodman 2016; Williams, Brooks & Shmargad 2018).

[42] Direct discrimination is also called "disparate treatment", while indirect discrimination is also referred to as "systematic discrimination", "disparate impact" or "unintended discrimination".

[43] In the following, the term "race" is placed in inverted commas or replaced by the term "ethnic origin" if it is used in the cited original texts, such as current legal texts or English language scientific literature. This follows the recommendation of UNESCO (1951) as well as biologists who point out the lack of a scientific basis for the term. See also e.g. Wersig (2017: 42).

3.3 Statistical discrimination

The concept of statistical discrimination is understood to mean the unjustified unequal treatment of persons on the basis of surrogate information.[44] The decision-makers have incomplete information about the main characteristic of persons on whom a differentiation decision is to be made. There is a cost and (time) effort involved in the detailed examination of the person's characteristics in order to obtain information about the main characteristic of differentiation. Costs or effort are judged by the decision-makers to be so high that they resort to surrogate information that is comparatively cheaper or can be obtained with less effort.[45] The surrogate information may also include variables belonging to a group that are the protected characteristics (e.g. the age variable) or variables that have a correlation with legally protected characteristics (e.g. years of professional experience). It is therefore possible to distinguish between two types of statistical discrimination: (1) if there is unjustified unequal treatment, which uses one or more legally protected characteristics as surrogate information, this can be referred to as direct static discrimination. An example would be the use of the ethnic origin characteristic when a supposed statistical link to labour productivity is suspected and people of a certain ethnic origin are excluded from jobs. (2) Where there are correlations between apparently neutral variables used and protected characteristics, there is indirect statistical discrimination. An example here is the use of the characteristic "part-time employment", where there is a correlation with the protected characteristic gender, as women more often work part-time. For both types, the surrogate information is also called "proxies".

If variables of a group membership are used as surrogate information (e.g. age of employees in the form of an age limit), a statistical relationship is often assumed between these variables and the differentiation objective (e.g. allocation of employees who are no longer capable of performing to retirement) and the main characteristic of differentiation (performance as an employee). This relationship is then generally assumed for further decisions about other or all individual group members, i.e. it is generalised and the surrogate information is used for **generalisation**. The assumptions about the "statistical" relationship can also be based on (supposed) empirical knowledge (Britz 2008: 8) or on statistical surveys and evidence. In the further course of the study, this "statistical" relationship and its changes with the use of algorithms will be examined in more detail.

According to Scherr (2016), the form of statistical discrimination also exists when decision-makers in markets (e.g. labour or housing markets) claim to have no prejudices or intentions of discrimination. However, due to a (supposedly) uncertain information basis about the characteristics, abilities and potential of individual applicants, "[...] assumptions about probable differences between social groups to which individuals are assigned are instead used as additional information to simplify the decision-making process" (Scherr 2016: 5; translated from the German original) (e.g. gender and skin colour instead of qualifications). This is often done when the time required to look at the individual case in detail is limited: "As a result, more or less plausible assumptions about the probable characteristics of categorically

[44] On statistical discrimination, see, for example, Britz (2008), Scherr (2016), Hellman (1998), Schauer (2003), Lippert-Rasmussen (2007), Gandy Jr. (2009), Fang & Moro (2011), Schauer (2018).

[45] For example, HR decision-makers in law firms have the differentiation objective of avoiding the risk of unsuitable recruitment. However, the main characteristic of being a good lawyer is difficult to determine because (a) it is not necessarily clear what characteristics and qualities make good lawyers, (b) some clearly relevant characteristics, such as judgment, could only be determined by extensive testing, and (c) even for relevant and testable characteristics, such as written expression, the determination requires costly assessment procedures. Therefore, it is more efficient for the HR decision-makers to use a substitute figure, such as the fact that applicants must be among the 10 per cent best in a year at a prestigious university. Example from Hellman (1998).

differentiated groups are already a gateway for discrimination [...]" (Scherr 2016: 5; translated from the German original).

Example and special case

A recent case of statistical discrimination is currently being investigated in Belgium. There, the energy supplier EDF Luminus refuses to supply electricity to persons living within a certain postcode area. For the energy supplier, this postal code area represents an area with many people with poor payment habits. Even solvent potential buyers are excluded from supply without taking into account their individual solvency.[46] This case represents a special form of statistical discrimination, the so-called "redlining", which is based on the surrogate variable "place of residence" and has been given its name by encircling areas with red lines (e.g. Barocas & Selbst 2016: 689).

Some authors point out that statistical discrimination is a type of discrimination based on "rational" decisions made by the decision-makers. Therefore, this form is often attributed to rational discrimination (Gandy Jr. 2009, 2010; Hellman 2008). In contrast to taste-based discrimination, the decision-makers do not have an intrinsic aversion to a particular group as such, but discrimination is based on "rational" calculations in order to deal with an information deficit as efficiently as possible. The concepts and theories of statistical discrimination were first developed in economics using the example of the labour market (Phelps 1972; Arrow 1973). This type of discrimination has later been studied in particular for the housing market, the credit and insurance industry and various forms of age discrimination (e.g. Arrow 1998; Hinz & Ausprung 2017; Britz 2008).

However, not all differentiations based on the formation of categories by statistical methods and analyses and the use of surrogate information are discrimination in the legal sense. There are challenges in assessing their legitimacy. The law provides standards by which it can be judged whether a form of statistical differentiation is considered unjustified. If, for example, it has the characteristic of indirect discrimination, the AGG requires that the objective justification and proportionality be examined.[47] (1) The use of the allegedly neutral characteristic (also procedure, rule, regulation) may be objectively justified by a legitimate objective, for example, for reasons of labour market and social policy or for reasons related to the enterprise or production. In this context, the sole reference to reasons of costs is not admissible. The objective justification must always be assessed on a case-by-case basis. (2) Whether the means of achieving the objective is proportionate – i.e. necessary and appropriate – must also be examined. It must be examined whether a milder, equally suitable means is not available and whether the means is proportionate to the desired objective (here according to Wersig 2017: 26f.) (For details, see Section 6.1.3.2).

[46] Information provided by employees of the Belgian equality body Unia – Interfederal Centre for Equal Opportunities, by email, 14 November 2018.

[47] Furthermore, the AGG regulates the justification reasons for unequal treatment in other areas: for direct discrimination in employment relationships in Sections 5, 8, 9 and 10 AGG, for other civil law relationships in Sections 5, 19 and 20 AGG and for direct discrimination in Section 3 para. 2 AGG; according to Wersig (2017: 29-30).

3.4 Changes in statistical discrimination

In data mining, big data analysis and machine-learning methods, the phenomenon of statistical discrimination is changed by replacing one or a few surrogate variables with entire models containing a large number of variables and their weighted relations to one another. Such models are generated by analysing large amounts of data and can be used as decision-making rules for differentiation or unequal treatment in software.

In general, machine-learning methods enable that mathematical models,[48] which represent the linear or non-linear relationships between variables, are optimised on data sets, which are characterised by a large, previously unknown set of relevant variables. A distinction is often made between the learning or training phase on the one hand and the application or productive phase on the other (Géron 2018). As an outcome of the training phase, recognised correlations or patterns are stored as models and, in the application phase, incorporated into decision-making rules and transferred to new decision situations.

In simplified terms, machine learning consists of training a model with the following elements (1) collecting and compiling a data set, (2) specifying a concrete outcome to be predicted in the data set, (3) deciding which possible influence variables are formed and provided to the training algorithm to be considered in the final model, (4) constructing a procedure to find the best influence variable that uses all other variables to predict the desired outcome (the outcome is the differentiation model or differentiation algorithm) which can be used to make predictions about the outcome, e.g. the rating of a person, and finally (5) the validation of the procedure with a retained part of the data set ("hold out set" or "test data set") that was not used for training (Kleinberg et al. 2019: pp. 17f.).

In the training phase, a mathematical model with learning algorithms is optimised in an iterative process by gradually adjusting the parameters with the aid of feedback until the model is best adapted (or "fit") to the data set. Typically, one or more initial models are used, on which different learning algorithms are tried out until one of the learning algorithms produces the best performance of the model in terms of the most accurate prediction or estimation of the outcome (Géron 2018: 30). In the test phase, the model is applied to the test data set and checked for so-called "overfitting" or "underfitting". Usually the problems of "overfitting" or "underfitting" arise when the generated model "fits" too much only to the training data set, does not generalise well, and does not "fit" well to the original data set from which the training data was taken.[49] After the training and test phase, the generated model can be used in the so-called productive phase to actually make predictions or classifications based on new data.

Compared to conventional programming, the use of machine-learning methods can save time and money, or enable the processing of complex data processing tasks in the first place. With traditional programming, for example, to detect and filter out email spam, you would have to programme rules for individual terms, patterns or typical email components that are known to be common in spam email. This would require a complex list, which would also have to be reprogrammed at great expense if spammers changed terms or components. When using machine learning for spam filters, one takes emails that users have previously

[48] In this context, the terms model and algorithm or learning algorithm are sometimes used synonymously, e.g. Lehr & Ohm (2017). For the sake of clarity, however, we will only talk about models in the following.

[49] Here, "generalise" means that the model generalises from a given number of training examples to data never used before. "Overfitting" occurs when the model is too complex in view of the training data and does not generalise well when applied. One speaks of "underfitting" if the model is not complex enough to be able to reproduce complex relationships in reality. Cf. Géron (2018: 17, 26-29).

marked as spam and lets the learning algorithm detect the relevant words or components (Géron 2018: 4-6). The example shows that machine-learning methods are particularly suitable for tasks that are too complex or too costly for direct programming “by hand”.

Typical characteristics of machine learning are that machine-learning methods are able to process more dimensions of variables than conventional statistical methods, they are able to sometimes achieve higher accuracy in prediction or categorisation, and there are usually many models that can be tested and from which the most appropriate ones can be selected. In particular, with more variables available for differentiation decisions, there is some hope that the use of protected characteristics as criteria for differentiation may become unattractive and thereby reduce the risk of direct discrimination (US CEA 2015: 16).

However, it became apparent early on that these characteristics can be regarded as advantages, but also have a number of disadvantages. For example, higher accuracy is associated with a loss of simplicity (and thus comprehensibility) (Breiman 2001). Or the problem of “overfitting” is caused, so that the machine-generated categories no longer correspond to those that the user has actually intended (Hand 2006). Other issues that may lead to risks of discrimination are discussed in Chapter 5.

4. Cases of Unequal Treatment, Discrimination and Evidence

In the following cases, unequal treatment of persons has resulted from the use of algorithms within differentiation applications, which are discussed as potential risks of discrimination or have been determined by the courts to be discriminating. In addition, some examples illustrate the general possibilities of proving discrimination without proving actual discrimination in the legal sense. The cases were compiled within the framework of a literature review, which was carried out between February 2018 and July 2019 and repeatedly updated. With a multitude of algorithms for differentiation and a very large, almost unmanageable number of applications in various systems, it is not possible to systematically record all applications. It should therefore be noted that the cases presented cannot correspond to a complete systematisation. The examples are assigned and discussed in Chapter 5 and 6. Chapter 5 outlines possible causes of risks of discrimination and societal consequences and Chapter 6 needs for action and options.

4.1 Working life

Case 1: Personnel software at Amazon

According to a media report, Amazon has been using a software system under development since 2014 to search and evaluate the CVs of potential employees found on the web. The machine-learning process was trained on CVs to search for word patterns that would indicate successful employees. During the development period, it was noticed that the system was not gender neutral. The system downgraded terms with "women's" and names of two exclusively female (high) schools. CVs from the last 10 years, mainly from men, were used as training data. This reflected the male majority of employees in the technology sector. Even with adjustments to the system, it was not possible to ensure that the system would not have developed other ways of discriminating against applicants. The development team was dissolved in 2017. According to the media report, staff members had considered the recommendations of the system but had not fully relied on the ranking (Dastin 2018).

Case 2: Online platforms TaskRabbit and Fiverr for freelancers

In a scientific study, Hannák et al. (2016) examined TaskRabbit (an online platform for freelancers) for unequal treatment. The online marketplace mediates smaller work services, such as household chores or the completion of errands. For the analysis, 3,707 profiles of those offering smaller work services from 30 cities in the US were collected over a period of five years. They were evaluated with regression analyses according to their ratings, the algorithm-based ranking of the search results of the online platform and according to the ratings of the clientèle in relation to the characteristics gender and ethnic origin. It was found that (1) women – in particular white women – received 10 per cent fewer ratings than men with comparable qualifications, (2) black people received significantly lower rating scores than other individuals offering services with similar characteristics and (3) the algorithm for ranking search results correlated significantly with ethnicity and gender, with the lower ranking grouping varying from city to city. The researchers recommended that online marketplaces should proactively identify and reduce bias (Hannák et al. 2016).

In a similar scientific study on the online marketplaces TaskRabbit and Fiverr with 13,500 profiles of people offering their work services, unequal treatment was also found in the reviews of the people offering their work with regard to the perceived characteristics of gender and ethnic origin ("race") (Hannák et al. 2017). In contrast to TaskRabbit, the online marketplace Fiverr mediates smaller "virtual" work services, such as the design of digital documents, help with programming or video editing. The personal data automatically collected via web crawling was assessed by human evaluators and assigned to gender and ethnic origin, as this concrete information is otherwise not used on the platforms. The evaluators were commissioned through the Amazon Mechanical Turk service. As an outcome, it has been demonstrated that black people on Fiverr have received fewer reviews and lower ratings. Also, the language used in the reviews, which were evaluated through linguistic analysis, differed according to gender and ethnic origin at the Fiverr service. In addition, they found an algorithmic bias in the ranking of search results on the TaskRabbit service, which resulted in negative correlations between the search result rank on the one hand and gender and ethnic origin on the other. However, the cause of the latter outcome could not be determined. Instead, the researchers assume that the algorithm for the search results is based on the reviews and ratings of the users who have used the services. As these were biased, so were the rankings of the search results (ibid., pp. 1915, 1927). The risks of discrimination were transferred and reinforced.

Case 3: Gender-biased unequal treatment on online platforms for job seekers

In a scientific study, Chen (2018) investigated gender-biased unequal treatment in specialised online platforms or search engines for the field of work in the USA. On the one hand, the online platforms allow job seekers to upload their CVs and short profiles to the site, and on the other give those looking for staff the opportunity to view the digital CVs that have been automatically sorted and ranked. The study looked at the algorithms used to rank the search engine results of Indeed, Monster and CareerBuilder. None of the search engines allowed results to be filtered or sorted by demographic characteristics (e.g. gender, ethnicity), but they did allow the use of surrogate variables, such as years of work experience, as an indicator of age. To generate the survey data, searches for applicants were conducted for 20 cities and 35 job titles using an automatic web browser, resulting in data on 355,000 applicants. The researchers were able to deduce the gender from the first names.

As a result, gender differences were found in the search results for all three online platforms. With regard to individual fairness, which was determined from the ranking according to gender with otherwise identical characteristics, a slight disadvantage of women was shown (although with only low effect sizes). With regard to group fairness, which would exist if the ranking algorithm assigned an equal distribution of ranks for women and men, men were better off in 12 of 35 occupational groups. As the websites did not collect any information on gender, the researchers did not consider that there was any direct discrimination, but other hidden characteristics (unemployment and institution of higher education) might have been taken into account.

The researchers were unable to clearly interpret the outcomes. In the case of individual fairness, according to one of their hypotheses, the outcome could also have been obtained if the algorithm had adjusted the rank according to how many jobseekers clicked on the respective profile. According to their assessment, the outcomes on group fairness can be explained by the structural inequality that already exists in some of the occupational groups considered (e.g. software developers). As a result, the cause of unequal treatment cannot be seen in the algorithms; instead, the researchers assess the search engines as successful when it comes to fairness in the sense of equal treatment of equal applicants. However, they do not see this for a fairness interpretation that means a distribution of employees according to the distribution in the overall population and would mean an active recruitment of underrepresented employees.

Case 4: Discriminatory job ads on Facebook

According to the Danish Institute for Human Rights (Institute for Menneskerettigheder), legal action is currently being taken against companies that have used or are using Facebook's differentiation feature to place selective job advertisements for men only on the online platform. The action is not directed against Facebook itself, but against the companies placing the ads. The unequal treatment was uncovered by a journalist who also researched the evidence and provided it to The Danish Institute for Human Rights. Further details, e.g. on algorithms for basic profiling and for enabling selective advertising, are not known.

On the basis of The Danish Act on the Equal Treatment Board and the Discrimination Act, the Danish Institute for Human Rights is empowered to bring potential discrimination cases to the tribunal without a specific complainant.[50]

Case 5: Age-based discrimination in job advertisements on Facebook

An investigation by the journalists' association ProPublica and the New York Times (Angwin, Scheiber & Tobin 2017) revealed age discrimination on the online platform Facebook. Companies such as Amazon, Verizon, UPS, Goldman Sachs and Facebook itself used the opportunity to post job ads on Facebook only for certain age groups. This was made possible by the approximately 5,000 options for the personalisation of advertisements, known as "microtargeting". The setting options are based on the detailed, algorithm-based analysis of data on Facebook users. As the older Facebook users did not see the ads, questions arose about the unequal treatment of people over the age of 40, which is prohibited under the Age Discrimination in Employment Act. The prohibition also refers to "assistance" in or "support" for age discrimination. Ultimately, the reporting led to a lawsuit filed by the union "Communications Workers of America" and others in the San Francisco District Court.

The union "Communications Workers of America" and others then filed a class action lawsuit against the companies T-Mobile, Amazon, Cox Communications and Cox Media in 2017 (United States District Court for the Northern District of California 2018). The procedural documentation indicates that the plaintiffs accused the designated companies and many other companies of having posted age-discriminatory job advertisements on Facebook. Together with other court cases, an agreement with Facebook was reached as an outcome of the proceedings (see Case 7).

Case 6: Age-based discrimination in job advertisements on online recruitment sites

According to the public prosecutor's office, the state attorney of the US state of Illinois, Lisa Madigan, brought formal action against the online employment websites Beyond.com, CareerBuilder, Indeed Inc., Ladders Inc., Monster Worldwide Inc. and Vault for suspected age discrimination. She sent letters to the companies warning that older users could face disadvantages when seeking employment due to the requirement that users comply with certain age requirements on the websites. For example, companies allow education and work experience information in the website menus only from a certain year limit, or the information must be provided at intervals from a certain year that do not fit older applicants with previous education and longer work experience, so that they can only create incomplete application profiles (Illinois Attorney General 2017). Algorithms are used here to control which information data subjects can provide, how personal data is analysed and classified, and they enable the automated and targeted addressing of specific groups of people.

[50] Information provided by staff at The Danish Institute for Human Rights, by email to the author, March 2019.

Case 7: Gender discrimination through job ads on Facebook

In 2018, the American Civil Liberties Union (ACLU), law firm Outten & Golden LLP and the Communications Workers of America (CWA) union, together with the Equal Employment Opportunity Commission (EEOC), filed a lawsuit against Facebook and 10 employing companies for unlawful discrimination based on gender, as job ads on Facebook were only posted to a male target audience, thereby excluding all women and non-male users from receiving the ads.[51]

In addition to the three protected characteristics (location, age, gender) that advertisers had to select, Facebook provided numerous other categories for detailed targeting that differentiated explicitly or implicitly by gender, such as "[...] Single Dads, Single Moms, Soccer Mom, Working Moms, Working Mother, Bad Moms, Strong Single Moms, Proud Single Mother, The Single Moms Club."[52]

Among the problematic practices cited in the indictment was the so-called "lookalike audience" service. Here, employers or employment agencies could transfer lists of their existing employees to Facebook. Facebook compared these with the data records on Facebook users and provided employers or employment agencies with lists of demographically similar Facebook users to whom targeted job advertisements could be sent. Facebook used features such as location, age, gender and interests in the processing. In the opinion of the complainants, this constitutes direct discrimination.[53]

In March 2019, in a total of five court cases,[54] an agreement between the company Facebook and the plaintiff organisations was reached, in which Facebook pledged, among other things, to set up a separate area on its platform for advertisements on staff positions, housing and loans, in which it would no longer be possible to address people according to age and gender and settings correlating with protected characteristics. The targeted advertising approach based on postcodes or within a region below a 15-mile radius will be abolished. The categories used in the "lookalike audience" service will be limited to "country, region, profession and field of study". Advertisers will also have to confirm compliance with anti-discrimination rights. The company also intends to set up a system of automated and human verification for correct identification and classification of advertisements.[55]

Case 8: Gender-biased unequal treatment in occupational classification

When it comes to online personnel searches and automated procedures of recruitment, the websites and online biographies of job seekers and professionals are becoming increasingly important. Likewise, whether and how these websites and biographies are found – and thus give job seekers access to employment positions – is also gaining in importance. Automated decision-making systems must be able to precisely record the jobs, skills, interests, etc. To this end, machine-learning methods should improve the assignment of persons to occupational group classifications on the basis of the descriptions (in particular the words used and their combinations) on the websites.

[51] See information on the ACLU website, https://www.aclu.org/cases/facebook-eeoc-complaints (last retrieved on 28 August 2019), and in the indictment "Charge of discrimination", available on the ACLU website https://www.aclu.org/legal-document/facebook-eeoc-complaint-charge-discrimination (last retrieved on 28 August 2019).

[52] See the indictment in footnote 51.

[53] See the indictment in footnote 51.

[54] According to Gillum & Tobin (2019).

[55] Sherwin & Bhandari (2019) and information from the "Exhibit A – Programmatic Relief" agreement document, available on the ACLU website: https://www.aclu.org/legal-document/exhibit-describing-programmatic-relief-facebook-settlement (last retrieved on 28 August 2019).

In a scientific study, DeArteaga et al. (2019) uncovered gender-biased unequal treatment in occupational classification based on existing gender inequalities in employment. The search engine Common Crawl was used to collect 397,340 online biographies. They tested three methods of machine learning for semantic representation with the outcome that scrubbing explicit gender indicators, such as first names or pronouns, was not sufficient in removing the gender imbalance, and that even in the absence of gender indicators the recognition rate (true positive rate) correlated with the existing gender imbalances in the occupational groups. Therefore, job classifications could further increase gender imbalances (DeArteaga et al. 2019: 2).

4.2 Real estate market

Case 9: Discrimination in housing ads on Facebook

Research by the journalists' association ProPublica showed that the company Facebook allowed to discriminate on its social networks when advertising flat rentals. To this end, ProPublica had placed ads itself and used the settings for the targeted ads to ensure that they were not targeted at African Americans, mothers with high school children, people who needed a wheelchair ramp, people of the Jewish faith, people emigrating from Argentina and Spanish-speaking people. These groups of people are protected under the "Fair Housing Act", the US anti-discrimination law for the housing market. Facebook approved all the ads, although their own company policies should have prevented this. As a result of the research, the U.S. Department of Housing and Urban Development (HUD), which also monitors the prohibition of discrimination in housing and renting, took action (Angwin, Tobin & Varner 2017).

In March 2018, the National Fair Housing Alliance (NFHA) filed a lawsuit against Facebook. In March 2019, the court proceedings were concluded with an agreement. After this, NFHA now offers the company a Fair Housing training programme, Facebook's advertising policy is regularly monitored and Facebook supports programmes to expand Fair Housing. Furthermore, Facebook has promised to set up a separate advertising portal for housing, employment and credit advertising with limited opportunities for targeted advertising (NFHA 2019).[56]

The HUD (US HUD 2019b) filed an additional lawsuit in March 2019.[57] The indictment focuses on the role of Facebook in the selection of those Facebook users who are displayed an advertisement or not. This selection decision would be based as much as possible on conclusions and predictions about the likelihood of users responding to the advertisement. According to the HUD, the conclusions and predictions are based on an analysis of the data that the company has on the individual, as well as data on other users that Facebook considers similar, and data on the "friends" and other people associated with the individual through Facebook ("associates"). The analysis is carried out using machine learning or other prediction techniques. Facebook uses gender and proxies for other legally protected characteristics, such as the websites visited, what apps a user has, where a user goes during the day and what purchases a user makes. This information would also be used to determine the prices that advertisers would have to pay for the targeted placement of the ad. Facebook determines the selection of those who see the ad, not the advertisers. In addition, advertisers who want to address a broad audience cannot achieve this, because the Facebook system makes selective decisions based solely on the characteristics of people who are most likely to respond to the ad (US HUD 2019a: 5).

[56] See for the agreement on the advertisements for jobs, housing and loans, which concerned several complaints at once (Case 7).

[57] As of March 2019.

Case 10: Unequal treatment on Airbnb based on ethnic origin

In a scientific study, Edelman and Luca (2014) point to unequal treatment based on ethnic origin on the commercial online marketplace for short-term rentals, Airbnb.com. In order to build reputation and trust, Airbnb enables landlords to place self-descriptions in the form of personal profiles, as well as allowing tenants to post ratings about landlords and landlords about tenants online. In a 2014 study analysing data on rentals in New York City, which establishes a correlation between the photos of the landlords and the rental prices, they show that non-black landlords can achieve twelve per cent higher rental prices for comparable offers as compared to black landlords. However, they cannot draw any clear conclusion from the outcomes as to whether this is a form of taste-based or statistical discrimination. They interpret the outcomes as unintended consequences of the mechanisms for building reputation and trust.

In another study using data on transactions via the platform Airbnb.com in the cities of Baltimore, Dallas, Los Angeles, St. Louis and Washington D. C., Edelman, Luca & Svirsky (2017) showed that the landlords accepted rent seekers with names that sounded "white" in 50 per cent of the rental requests, while they accepted seekers whose names sounded African American in only 42 per cent of the rental requests. Discrimination, i.e. the rejection of guests, results in "costs" for those offering properties on the site, in the form of lost profits due to rooms remaining empty. According to the authors, however, the study could not explain the mechanisms leading to discrimination, and this study also does not provide clear evidence on whether discrimination is taste-based or statistical (Edelman, Luca & Svirsky 2017: 17).

Case 11: Ethnicity-based unequal treatment on Airbnb

Based on the study by Edelman and Luca (see Case 10), Gilheany et al. (2015) show that Asian landlords on the rental service website Airbnb fetch 20 per cent lower prices than white landlords for similar rentals. To this end, they examined a total of 101 Airbnb landlords in the cities of Oakland and Berkeley who could be identified as white and "Asian". However, the authors cannot clearly identify the causes of the differences. This example also shows that the identifiability of the provider leads to the risk of discriminatory behaviour by other beneficiaries. Questions arise as to whether platform operators could not better prevent this risk using their possibilities of managing and controlling what information is communicated via the platforms.

4.3 Trade

Case 12: Proof of price differentiation in electronic commerce

In a scientific study of e-commerce companies in the general trade and travel sector (Hannák et al. 2014), researchers used the accounts and cookies of 300 fictitious users to prove price differentiation, i.e. prices adjusted to some users, and price control, i.e. controlled arrangement of search results for particularly expensive or less expensive products in high rankings. They set up user accounts to find out the impact of different characteristics of the users, such as the type of web browser, operating system, the existence of a user account or the products purchased and viewed in the past. For the investigation, they used control accounts to separate actual personalisation from interference or noise. To measure this, they developed a metric for information retrieval. For the study, work services were purchased via the crowdsourcing platform Amazon Mechanical Turk. They found personalisations at nine of the 16 online retailers investigated. Two companies use price differentiation to grant reduced prices to "members". Two companies use A/B tests (two variants of a website are shown for different groups of people), which lead a subgroup of users to more expensive hotels, two companies use personalised search results for mobile devices and one company personalises the search results based on past clicks and purchases (Hannák et al. 2014: 306). Although the example does not show discrimination in the legal sense, since no protected group of persons seems to be disadvantaged, it does illustrate that algorithm-based price differentiation can be proven "from the outside" without direct inspection of the algorithms.

Case 13: Unequal treatment in logistics at Amazon

According to a media report, Amazon's "Same Day Delivery" offer used an algorithm to calculate the areas in which the company was the first to introduce the then new form of delivery. Although the algorithm did not consider ethnic origin as an input, it excluded neighbourhoods with predominantly black inhabitants. According to Amazon, features that were considered in the algorithm included proximity to the nearest distribution centre and the number of people with a prime membership in an area. Presumably a correlation to the characteristic of ethnic origin was created. Journalists researched the case for six cities. They tested the availability of the service according to postal codes and compared the data and maps obtained with data from official population statistics. Following the reporting and protests, the company promised to extend the service to previously disadvantaged areas (Ingold & Soper 2016).

4.4 Advertising and search engines

Case 14: Stereotypes in search engine results

Noble (2018) used numerous examples to show how search engines contributed to the increase in racism by referencing disparaging stereotypes in the ranking of the output of links, in the images displayed or in word completions in the auto-complete search box (e.g. predominantly pornographic images when searching for the keyword "black girls"). She mainly used the search engine Google.

An early and much-cited example is described in a media report. According to this, a Google photo app automatically assigned a derogatory designation with the tag "gorillas" to photos of black people (Kasperkevic 2015).

Case 15: Gender differences in search results for images for occupations

In several scientific studies, Kay et al. (2015) examined the results of searches for images of occupational groups that were output by Google's search engine and whether the search results reinforced stereotypes in representation and perception. Among other things, the investigations showed an under-representation of women in the search results for those occupational groups stereotypically dominated by men compared to the gender ratio of the official employment statistics for these occupational groups. This reinforces stereotypes. It was also shown that the quality of the presentation (particularly in terms of the professionalism shown) assessed by respondents was also higher for those occupational groups that corresponded to gender stereotypes. The researchers also showed that the perception of gender relations in search results has an impact on perceptions of actual gender relations in occupations. The authors discuss their findings with regard to possible amplification effects on inequalities in occupations, including the fact that this could also influence or limit the pursuit of careers in the occupations. However, they expect improvements in the automated marking of images (ibid., p. 3826).

Case 16: Unequal treatment in targeted advertising on Gmail

Through case studies, Lecuyer et al. (2015) show that the Google service Gmail also uses sensitive or legally protected characteristics for targeted advertising (e.g. health, religious affiliation and interests, sexual orientation or tight financial situation). The authors used the open source system "Sunlight" they developed for the investigation, which served to detect and statistically prove personalisation on the web, i.e. in the form of personalised advertising, recommendations or personalised content.

Case 17: Gender-biased unequal treatment in advertising on Google services

Another scientific study was devoted to the ad settings of Google's services (Datta, Tschantz & Datta 2015). For the study, the authors used the "AdFisher" system, which served to investigate the interrelationships between user behaviour, Google's advertising and user settings. The system collects large amounts of personalisation results, such as personalised advertisements, using computer-simulated agents. They demonstrated that advertising for higher paid jobs had been shown comparatively more to men than to women when the gender was changed to women in the "Ad setting" settings, although the web browsing behaviour was identical. Since the investigation "from the outside" looked at a complex "ecosystem" of online advertising, the exact cause of the unequal treatment could not be uncovered.[58]

Case 18: Ethnicity-based unequal treatment in the placement of advertising

In an empirical study, Sweeney (2013) proves that advertisements for commercial products of the documentation of arrests, previous convictions, criminal offences, etc. ("arrest records") in Google's search engine results on web pages using the ad service are displayed differently, depending on whether the names searched for indicate a particular ethnic origin. For example, employers use this kind of documentation when evaluating jobseekers. In the survey, the ads for the products were more often found when searching for names that sounded "non-white" than for names that sounded "white". The cause was seen in the algorithmic process that Google used in the Google AdSense service to display targeted ads for certain search queries. Behind this was an automated real-time auction mechanism that controls the prices and placement of advertisements. The mechanism included, among others, the (previous) rates of clicking on advertisements. The author has not (yet) been able to identify any clear causes for unequal treatment (ibid., p. 52), but the algorithm for the placement of advertisements is used to analyse the behaviour of users on the search engine website, i.e. the recorded click behaviour as well as past and current trends, and accordingly reflects societal unequal treatment.[59]

Case 19: Gender-biased unequal treatment in advertising on Facebook

Lambrecht and Tucker (2019) demonstrated gender differences in online advertising for STEM (Science, Technology, Engineering and Maths) careers in an empirical case study. They showed that the test ad, which was placed in 191 countries on the online "social" network Facebook, was shown 20 per cent more often to men than to women, although women were more likely to respond to the ad ("click through rates"). The authors interpret this to mean that the group "women" is more expensive for advertisers, since the algorithmic mechanism calculates prices according to the probability of users viewing the advertisement, i.e. the "click through rate". The authors assume that advertisers lose more auctions for the target group "women" than for the target group "men". A comparable unequal display of ads between women and men has also been demonstrated for the advertising providers Google Display Network, Instagram and Twitter. With these results, the authors illustrate that economic mechanisms – in this case the advertising auction mechanism of online platforms – in particular can lead to discriminatory outcomes of algorithm-based differentiations. The study of Ali et al. (2019) confirmed the results of the study (see Case 20).

Case 20: Gender-biased unequal treatment in advertising algorithms

Ali et al. (2019) examined selective or personalised advertising on the Facebook platform for possible discrimination. For this purpose, they used the "Custom Audience" customisation service that Facebook made available to advertisers[60] and determined the target group with a list prepared according to telephone numbers (and randomly selected users for control purposes). They tested the image recognition algorithm

[58] For further analysis of the case, which leads to similar outcomes, and legal classification under US law, see Datta et al. (2018).

[59] Guidance on explanation is given in the examinations of the Cases 19 and 20 below.

[60] The authors group the various possibilities of targeted advertising on the Facebook platform, (1) by group-based selection according to demographic characteristics, such as age, gender, location or profile information of users, (2) by individual persons who can be specified to Facebook, such as in the form of lists of names, addresses, telephone numbers, birthdays, or in the form of persons who are recognised by web tracking tools ("custom audience"), or (3) certain groups of persons can be selectively addressed by referring to users who are similar to those who have already been selected ("lookalike audience"), cf. Ali et al. (2019: 4).

by placing various advertisements with stereotypical images for women and men, with some of the advertisements showing "female" or "male" stereotypical features that could only be recognised by machine image recognition (e.g. construction machinery or military for men, bridal bouquets or perfume for women), while for human viewers these images were not recognisable. Nevertheless, the majority of the prepared images reached a "female" or "male" target group. The researchers therefore concluded that Facebook had analysed the image data and automatically assigned and delivered the ads according to stereotypically unequal gender. From their findings, they concluded that the company's internal algorithm determined which advertisements were displayed to which users and that the algorithm could therefore be discriminatory. The algorithm acted independently of the settings that advertisers could define in the selection of target groups. However, the authors admit the limitation that their outcomes cannot be generalised.

The outcomes of the study point to possible legal consequences, as they show that not only those who place the advertisements are responsible, but also the platform company Facebook, which develops and uses the algorithms of the ad regulations. The outcomes can also be linked to the lawsuit filed by the U.S. Department of Housing and Urban Development (HUD) on housing claims (see Case 9) (Matsakis 2019).

4.5 Banking industry

Case 21: Ethnicity-based discrimination and the emergence of FinTechs

By comparing traditional lending with algorithm-based lending, Bartlett et al. (2018) show that ethnicity-based discrimination has continued to exist in the US mortgage market (for African Americans and Latin Americans), even with algorithm-based lending decisions. However, algorithms and the emergence of FinTechs have changed the nature of discrimination from discrimination based on human prejudice or aversion to illegitimate applications of statistical discrimination using big data variables. Big data variables replace variables for risk assessment that are not or only poorly determinable with surrogate variables (e.g. the authors mention high school graduation for income increases as one of many variables). Algorithm-based lending has increased competition, facilitated comparability between providers and also facilitated switching between them. In addition, FinTech companies do not discriminate by refusing loans, as conventional lenders do, but by setting higher prices or interest rates. In the latter case, however, the degree of discrimination (measured by additional interest rate premiums) was as high as in conventional lending.

Case 22: Multiple discrimination in online lending by Svea Ekonomi

According to the judgement document (YVTltk 2018), the National Non-Discrimination and Equality Tribunal of Finland (Yhdenvertaisuus- ja tasa-arvolautakunta) sentenced a credit institution in Finland for using an inappropriate statistical method, using protected characteristics and failing to perform an individual assessment of solvency (on the judgement, see YVTltk 2018).

The credit institution Svea Ekonomi AB had refused to extend the loan of a male applicant, who had applied for it on a website in connection with an online purchase. The person concerned reported the case to the anti-discrimination ombudsman, who brought the credit institution before the tribunal. The tribunal decided that the loan-granting procedure must no longer be used and imposed a fine of 100,000 euros. The tribunal based its decision on the fact that this was a case of direct multiple discrimination, as the legally protected characteristics of gender, mother tongue, age and place of residence had been used, and that the applicant had not been individually assessed with regard to his credit behaviour and creditworthiness; instead, formal and abstract credit data based on the credit behaviour of others had been used (YVTltk 2018). In this case, the characteristics of statistical discrimination are clearly evident, i.e. eschewing individual assessments and instead using surrogate variables which, in the case in question, were the protected characteristics of gender, mother tongue, age and place of residence, and constitute direct or indirect statistical discrimination.

A schematic credit decision was available, based on internal data of the credit company, the credit data file and score data. The score represented factors such as gender, language, age and place of residence and was based on statistical correlations calculated using data on other people. These indicated that men had repayment problems more often, which is why they received a lower score than women. Similarly, Finnish-speaking residents received a lower score compared to Swedish-speaking residents. The plaintiff's mother tongue is Finnish. If he had been a woman or had Swedish been his mother tongue, the score would have been enough for the loan. The ombudsman pointed out that the credit applicant had not defaulted (YVTltk 2018: 7). Since the man also lived in a region that was assigned a value for unknown areas by the system, this was also seen as a disadvantage (YVTltk 2018: 6).

In the lawsuit, the ombudsman distinguished precisely that the credit scoring system cannot be used to obtain precise information on the actual situation of an individual applicant, as the system can only give a statistical assessment of how likely it is on average that applicants will match the profile of applicants with a poor credit score. He also pointed out that the applicant had not been treated as an individual, but instead as a "representative of statistical profiling", based mainly on variables of the protected characteristics that the lender applies to all persons who fit the profile, such as men who live in a certain neighbourhood or have a certain mother tongue and are of a certain age (YVTltk 2018: 4f.). As a result of this method, people with a stable income and with evidence of the ability to repay the loan would be refused a loan (YVTltk 2018: 5).

The defendant credit institution stated in its response that under anti-discrimination law, unequal treatment does not constitute discrimination if the treatment is based on legislation and has an otherwise acceptable objective and the measures to attain that objective are proportionate (YVTltk 2018: 8). Moreover, the credit procedure had complied with the Credit Data Act and the supervisory authority did not have any objections to the procedure. The credit decision procedure is a part of other operators' online sales systems; this type of online financing is purchase-bound and is a fast and automated process. The individual credit assessment of credit-seekers using personal information and documents, such as salary or tax certificates, is not suitable for this type of financing process (YVTltk 2018: 10f.).

The tribunal nevertheless ruled that the procedure was not proportionate and therefore not acceptable under the applicable discrimination laws. The tribunal also ruled that the credit institution could not refute the presumption of discrimination (YVTltk 2018: 2). The judgement also states that assessments of solvency are increasingly based on assumptions generated with data collected from other people. These assumptions cannot be used to provide the credit applicant with acceptable reasons for refusing the loan, in particular if the credit applicant is not given an opportunity to clarify their actual ability to pay and the factors that affect it (YVTltk 2018: 17).

4.6 Medicine

Case 23: Biased data records in a diagnostic system

Biased training data was found in a health care machine-learning system designed to predict the risk of mortality for patients with pneumonia. The system is intended to support decisions on whether patients can be treated as outpatients or inpatients. The system, based on a neural network, identified a higher chance of survival for patients with asthma, an assumption that in fact contradicted medical experience. The bias in the data set resulted from the fact that patients with pneumonia and a (long-term) asthma disorder were taken directly to intensive care units. As a result, better outcomes were achieved for patients with pneumonia and asthma than for patients who "only" had pneumonia. Although the system made correct predictions, the problem arose because this information about the context was not included in the decision support system (Caruana et al. 2015). The case also illustrates the problems that arise when contextual factors that are relevant but not represented in data sets are ignored in decisions (Cabitza, Rasoini & Gensini 2017: E1).

Case 24: Ethnicity-based discrimination in patient allocation

Obermeyer and Mullainathan (2019)[61] describe finding evidence of ethnicity-based discrimination or "racial" bias in a widely-used commercial system for allocating patients in need of intensive medical care to a care management programme. The assignment to the "care management" programme is associated with a higher allocation of resources. White patients were more likely to be assigned to the programme than black patients in a comparable state of health. The allocation was made using an algorithmically generated risk score. The calculation included data on the total medical expenditure in a given year and fine-grained data on the use of health services in the previous year. The score therefore did not reflect the expected state of health, but instead predicted the cost of treatments. From the authors' point of view, these cost predictions are also accurate and unbiased.

However, the problem is that, although treatment costs can be used as proxies for health status, this is insufficient surrogate information. This is because factors other than health status alone determine the cost level, such as ethnic origin. As such, black patients would cost less to treat, depending on their health status. An algorithm that correctly predicts the costs for individual ethnic groups inevitably provides biased predictions about health conditions. They attribute the problem to the determination of the objective function and the selection of labels, which unilaterally target cost optimisation and generate externalities with regard to health (Obermeyer & Mullainathan 2019).

[61] At the time of this study, only an abstract was available.

4.7 Transport

Case 25: Quasi-segregation in navigation services

The navigation service “Ghetto Tracker”, which was renamed “Good Part of Town” in the US, was suspected of pointing out “unsafe” areas with predominantly non-white inhabitants and was shut down again after protests. However, according to Silver, various other navigation services also have similar functionalities that warn of “unsafe” neighbourhoods (Silver 2013). Route planning is one of the most prominent examples of the use of modern algorithms, and the processed data sets with their discriminatory ratings and stereotypes are likely to be the main cause of risks of discrimination.

Case 26: Possibility of unequal treatment at the ride sharing company Uber

Using the case study of Uber, Rosenblat et al. (2017) illustrate that the system of rating drivers by customers/users of the ride sharing service can be a source of discrimination based on ethnicity or “race”. The case study is based on qualitative field research with ethnographic studies and interviews with the vehicle drivers. The rating system is a fundamental element of the company’s quality assurance system for the individualised, scattered workforce. User ratings are a key factor in the company’s automated assessments and HR decisions, used to determine whether to “activate” the drivers or “deactivate” them with a temporary suspension or complete termination of the relationship, for example. They also have a direct impact on the drivers’ earnings and their chances of getting higher-paid work.

The authors assume that ratings given by users and customers are systematically characterised by unequal treatment according to ethnic origin and gender and refer to findings from similar areas. However, they lacked access to company data on user ratings and the composition of the group of drivers that would be required in order to provide accurate evidence. Because the user ratings have such a central place in the business model of the company and biased user ratings are incorporated into the company decisions, indirect discrimination is likely. The authors assume that such risks of discrimination are relevant to all platforms where a user rating system regulates a distributed workforce.

4.8 State social benefits and supervision

Case 27: Continued unequal treatment in forecasting systems of state supervision

Altenburger and Ho (2018) first show that ratings of Asian restaurants on the online platform Yelp were biased by the fact that visitors and customers rated Asian restaurants disproportionately worse than other restaurants. For the New York and King County regions in the US, they compared the evaluation on the platform and complaints to emergency services by customers and data on inspections by health authorities. The authors then demonstrate that such unequal treatment can be perpetuated through biased ratings from the private to the public sector, as the customer ratings are also used for public tasks of food and health inspections. According to the authors, this is a general development trend in state regulation, with state controls being replaced by big data analyses based on ratings data from platforms. Bias continue to occur when algorithm-based forecasting systems (“predictive analytics”) are used for state supervision, which are developed and operated on the basis of biased customer ratings. To demonstrate this, they estimated the consequences of using a machine-learning method for predictions based on biased ratings. If, according to their opinion, state food and health inspections of restaurants were to be replaced by forecasting systems based on the scores from biased customer opinions, they believe that biased algorithms will become regulatory instruments in the future.

Case 28: Unequal treatment in a system for the prevention of child abuse

The Allegheny Family Screening Tool (AFST) is a system for predicting and identifying preventive interventions in potential cases of child neglect and abuse, which is used in Allegheny County, Pennsylvania (in operation since 2016, revised in 2018). For each telephone call, a risk assessment system assigns a score from 1 to 20 for each case, which is displayed to the person processing the call. A score of 20 indicates the highest risk. The calls or tip-offs usually come from the community, i.e. from neighbours or teachers.

Surrogate information or proxies are used to determine the risk scores (Eubanks 2017: 143-144). This is data on cases from the past, including, (a) new reports from the "community", i.e. if, within two years, a new report was received on the same child for whom no further investigation was carried out on the first call ("screened out"), or (b) if removal from the family occurred ("child placement") after a further investigation call, which resulted in the child having to be removed from the family and assigned to a care institution. The algorithmic conclusions thus reflect former "social" ratings of the families by the community, the authorities and the courts. The model could also only provide predictions about future clues and future removals from families, but not about the future actual abuse of children. In addition, important variables in the modelling, such as geographical isolation, are missing. The system assigns disproportionately high risk scores to families who have already made frequent use of social benefits and produces outcomes with ethnicity-based unequal treatment (Eubanks 2017: 143-144; Courtland 2018). A quality measure for classifications was determined for the prediction accuracy, the value for ROC AUC,[62] which was given as 76 per cent (Eubanks 2017: 145). As there were problems with previous versions of the system, it was revised and a group of researchers was commissioned to review it (Chouldechova et al. 2018). They see a risk of negative amplification effects in the use of predictive analytics, since more data of some regional communities, in particular poorer or specific ethnic groups, are stored in state institutions, e.g. because they are included in social assistance systems, and these communities are therefore identified as high-risk and checked more frequently (Chouldechova et al. 2018: 2).

Case 29: Risk of direct and indirect discrimination in a system for classifying unemployed persons

A study by the Polish non-governmental organisation Fundacja Panoptykon highlighted the risks of discrimination through a system of employment services (Niklas, Sztandar-Sztanderska & Szymielewicz 2015). In 2014, the Ministry of Labour in Poland introduced a system of job placement that classified jobseekers into three categories ("Profiles I-III") based on the creation of profiles and scores. The objective is to determine the "distance" from the labour market and the willingness to enter or re-enter the labour market (Niklas, Sztandar-Sztanderska & Szymielewicz 2015: 11). The classification of each jobseeker into one of the three categories determines the type of labour market programme to be applied (e.g. job placement, vocational training, apprenticeship). In the process, 24 characteristics are processed that are intended to characterise unemployed people. These characteristics are derived from the registration of jobseekers in the employment centres and the computer-assisted interviews with the caseworkers. The recorded and processed characteristics include age, gender and degree of disability (Niklas, Sztandar-Sztanderska & Szymielewicz 2015: 5, 11) as well as information on periods of childcare and care of persons requiring assistance (ibid., p. 21). Information on the need and willingness to accept a job is also recorded,

[62] The so-called ROC (Receiver Operating Characteristic) curve is configured in a diagram in which the true positive rate is entered on the ordinate and the false positive rate on the abscissa. When comparing the quality of classifiers, the area under the curve (AUC) is considered. A perfect classifier has a value of 1 and would result in a point in the upper left corner of the diagram, and a completely random one would have a value of 0.5 or would represent the diagonal in the diagram (Géron 2018: 92-94). The objective is to obtain the highest possible ROC AUC value. The ROC AUC value illustrates the compromise between the hits or true positive classifications and the "cost" or false positive classifications.

e.g. personal commitment to find employment, willingness to meet labour market needs, flexibility or past or current willingness to cooperate with the relevant labour market authorities (ibid., p. 11).

In the employment centres, the caseworkers handled the systems differently. For example, the computer system was seen as the final decision-maker, profiling was seen as part of a comprehensive investigation and attempts were made to adapt the profiles to the unemployed person's expectations (Niklas 2018). According to the authors, one of the problems of the system is that legal provisions, e.g. to know the types and scope of data to be processed and to adapt the data and profile, have not been translated correctly in the functions and rules of the system and its operation. This makes it more difficult to adjust and correct entries, for example (Niklas, Sztandar-Sztanderska & Szymielewicz 2015: 17; Niklas 2018). According to the authors, the application can lead to both direct and indirect discrimination. Since the categories are also created using the protected characteristics of age, gender and disability, there would be direct discrimination. Indirect discrimination would result from the application of the characteristics "periods of childcare and other care", which statistically affect women more often (Niklas, Sztandar-Sztanderska & Szymielewicz 2015: 21). In real terms, jobseekers assigned to the least favourable category "Profile III" would then have a lower chance of receiving a support measure. This assessment of opportunities is based on statistical analyses and experience (Niklas, Sztandar-Sztanderska & Szymielewicz 2015: 25-27).

Furthermore, jobseekers do not have sufficient access to the data and analyses that make up their assessment and assignment to one of the three categories. This makes it difficult for them to assert their right to compensation in the event of discrimination. Indeed, even if the burden of proof is shifted to those accused of discrimination, those affected by potential discrimination must prove that there is a likelihood of discrimination (Niklas, Sztandar-Sztanderska & Szymielewicz 2015: 21). Furthermore, according to the authors, the system violates the right to procedural fairness, as there are no clear procedures for appealing, expressing the opinion of the persons affected and requesting re-verification of the assigned profiles (Niklas, Sztandar-Sztanderska & Szymielewicz 2015: 22f.). According to Niklas, however, there are plans to abolish the system (Niklas 2019).

Case 30: Risk of discrimination in the classification of unemployed persons

According to media reports, a system is currently being tested[63] in Austrian job centres (Arbeitsmarktservice, AMS) that divides jobseekers into three categories indicating the likelihood of integration into the labour market (high, medium, low). On this basis, recommendations are to be made on the allocation of funds for qualification measures, such as training or additional courses (Fanta 2018). The likelihood of integration is calculated as the probability of accepting work. According to the company Synthesis Forschung GmbH, which developed the system, the computation of probabilities is based on a logistic regression model (Holl, Kernbeiß & Wagner-Pinter 2018). According to the AMS Management Board, the hit rate is 85 per cent, which means that about 50,000 people a year are misclassified. This system does not have a fully automated decision-making system, as the system only makes recommendations (Wimmer 2018b, 2018a).

One criticism was that the AMS administrators tend to adopt the recommendations of the system, because additional evaluations are required if the human evaluation of an administrator deviates from the computer recommendation and the person affected is to be downgraded (Wimmer 2018b). On the other hand, there is criticism that certain groups of persons could be disadvantaged, because the algorithm not only takes into account the protected characteristics of gender, age and nationality, but also implements weightings which, for example, assign women a lower number of points than men. Other controversial

[63] As of January 2019.

characteristics are care responsibilities, such as time spent on childcare, or health impairments (Szigetvari 2018).

However, AMS countered the criticism that classification in a category of labour market opportunities does not mean that the groups concerned are worse off in terms of qualification measures. In this context, strategies and programmes for the promotion of disadvantaged groups are receiving particular consideration and are taken into account in decisions on qualification measures (Wimmer 2018a; Salzburger Nachrichten 2019). It is also likely that the extent to which the criticism will have an effect – that the creation of the algorithm is based on the statistical analysis of past data, which could lead to a perpetuation of inequalities on the labour market – will only be assessed through the accumulated decisions on support funding with regard to the implementation of support programmes, which are allegedly aimed against the further disadvantaging of groups that are already disadvantaged.[64]

4.9 Education

Case 31: Risks of discrimination in a system for allocating places in universities

According to the French equality body and the documentation of the public procedure, the French "Parcoursup" system of university place allocation requires prospective students to make 10 to 20 requests for university places and enter personal details in an online system. The system at national level matches the requests to the respective intake capacities of (public and private) higher education institutions. At local level, each higher education institution has its own system to manage requests and select students. First of all, the lack of transparency in the decision-making procedure has been criticised. Although the government made the source code of the matching algorithm public at national level, the sorting algorithms were not public at the local level of the respective higher education institutions. The personal data that must be provided when registering include income and place of residence, which led to concerns among the plaintiffs of discrimination against less wealthy prospective students or those from suburbs.[65]

An association of elected representatives, student representatives, teachers and lawyers presented the case to the French equality body Défenseur des Droits (DDD), which conducted a legal investigation. The investigation led to two decisions by DDD in the form of recommendations to the Ministry of Education: Decision No 2018-323 refers to the special consideration of students with disabilities.[66]

The second Decision no. 2019-021 of January 2019 refers to "local algorithms" and recommends that all information concerning processing, including that relating to algorithms, and the evaluation of application documents by the local commissions of higher education institutions should be published in advance. The aim of this measure is to ensure transparency of the procedure and allow applicants to make their decisions in full knowledge of the facts. The decision also recalls that the use of the applicants' school as a characteristic in the selection process, in which certain applicants are favoured or others are excluded on the basis of the

[64] The Ombudsman Board therefore requested a review and parliamentary discussion of the AMS system. Cf. The Standard (2019). However, Fröhlich & Spiecker called Döhmann (2018), who regard the use of a lower weighting for women as discriminatory, are critical of this, as is Allhutter (2019), who assumes self-reinforcing processes, in particular for people who have several characteristics that are attributed lower labour market opportunities.

[65] Information provided by staff at Défenseur des Droits, by email, November 2018 and March 2019.

[66] Information provided by employees at Défenseur des Droits, by email, March 2019. Interpretive supplement in square brackets by the author.

geographical location of the school, may be considered a discriminatory practice if it leads to the exclusion of applicants on that basis.[67]

During the investigation, DDD did not have access to the algorithms of the "local level", but did receive general information about their use. In principle, DDD has the legal means to collect all information that seems necessary for the facts that are shown to them. Restrictions on information due to the protection of secrets, which only apply to secrecy for national security, state security or foreign policy, were not relevant in this case. Systems at "local level" in individual higher education institutions have been introduced on a voluntary basis, without the introduction of an [overarching] system, and are designed to prioritise applications before the actual selection by the selection committees. The system can therefore be regarded as a decision-support service. DDD could not identify any system with fully automated data processing during their investigation. The equality body DDD cooperated with the French data protection authority CNIL (Commission Nationale de l'Informatique et des Libertés).

From the legal situation for such admission decisions at higher education institutions, they concluded that examination commissions must be set up to answer applicants and that such decisions therefore cannot be fully automated. The "Parcoursup" system at "national level" is one of the algorithms that must be authorised by ministerial decision after the CNIL data protection authority pronounces and publishes its decision. After reviewing compliance with the legal provisions on transparency, the right to human intervention, the types of data collected and the categories of persons with authorised access to data, the CNIL issued an opinion in favour of the introduction of the "Parcoursup" system. However, the systems at local level were not sent to the CNIL for review and comment.[68]

4.10 Police

Case 32: Reinforcing unequal treatment in a system of predictive policing

In a simulation study, Lum and Isaac examined the "PredPol" system for predictive policing, which predicts which areas will have a higher probability of crime (spatially based predictions). They assumed that bias in the (training) data sets lead to biased predictions of crimes and correspondingly to more operations in areas that are already overrepresented in police statistics. Due to the likelihood of increased deployments in the same areas, additional offences were observed, confirming the previous assumptions about the distribution of offences. The aim of the simulation study was to reveal the strength of this bias. Lum and Isaac underscore that police records are not a complete survey of all criminal acts, nor do they represent a random sample. In order to determine the extent of the bias in the police records, they correlated the police records with a complete set of crime statistics taken from the National Survey on Drug Use and Health. This data was simulated on the regional population composition of the city of Oakland, allowing estimates of drug-related crime at a regionally high-resolution level for individual areas of the city, which in their opinion provided a more accurate picture of drug abuse than police arrest data (Lum & Isaac 2016).

According to the results of the simulation, drug-related offences should have been scattered more widely across the city, but the actual arrests by police for drug-related offences were concentrated in a narrowly defined area. They used a publicly available version of the algorithm for the investigation of the PredPol predictive policing system. According to the company, only three data points are used for prediction: the

[67] See Décision 2019-021 of 18 January 2019 on the operation of the national platform for pre-enrolment in the first year of higher education (Parcoursup), available at https://juridique.defenseurdesdroits.fr/index.php?lvl=notice_ display&id=27285 (last retrieved on 26 March 2019).

[68] Information provided by employees at Défenseur des Droits, by email, March 2019. Interpretative supplements in square brackets by the author.

previous type of crime, the crime scene and the time of the crime. The algorithm was applied to the police statistics for a given year and showed the already "over-policed" areas as areas with high prediction values. In comparison with the simulated distribution of crime, where there was a distribution of areas with a higher proportion of white people, the predictions provided by the PredPol system affected fewer areas, particularly those in which a higher proportion of black people lived. In addition, they simulated a situation in which there were incentives for further police operations in "predicted" areas, with the outcome of confirming the above-mentioned effect (ibid.).

The study was criticised because the PredPol system was not developed to combat drug-related crime and the use of drug data was therefore incorrect, and because the system was not used in Oakland (Ferguson 2017). The authors responded that the aim was to prove the possibility of the amplification effect, whereby the problematic issue is that the crime data used is generated by the police (virtually as a secondary activity) and that this is not a representation of all crimes, that the police are not notified of all crimes and that the police do not document all crimes to which they react. "Police-recorded crime data is a combination of policing strategy, police community relations and criminality." (Isaac & Lum 2018: without page reference). Furthermore, the system was abolished after testing due to a lack of evidence of improvements in crime prevention (ibid.).

Case 33: Ethnicity-based discrimination in a predictive policing system

Brantingham et al. (2018) report on an empirical study (randomised controlled study) to uncover possible discrimination based on ethnic origin (or minorities), which was part of comprehensive accompanying research on the introduction of a predictive policing system in Los Angeles. To this end, the effects of police force deployment plans with crime forecasts for respective areas of deployment, which were created by the algorithmic system, were compared with those of a human planner. Police forces had been told that the areas of operation are the areas with the highest crime rates for their shift. In the comprehensive accompanying research, the reduction of the crime rate could be proven for both algorithmic and human prognoses, whereby the decrease in the crime rate was higher for algorithmic prognoses. The results of the study on risks of discrimination revealed that the use of the algorithmic system did not lead to differences in arrests of persons belonging to minorities. In contrast, the arrest rate increased in the areas where the algorithmic system was used (across all population groups).

Case 34: Data sets on discriminatory police practices in predictive policing systems

In the context of systems of predictive policing, Richardson et al. expands the notion of "dirty data", which now includes "dirty" records based on corrupt, discriminatory or unlawful police practices, in addition to missing or incorrect data or non-standard representation of data. They may also have been created and changed through deliberate manipulation. It is especially problematic when the development and application of predictive policing systems are based on such "dirty" police data sets. These authors also emphasise that police data is not objective data, nor do they reflect actual criminal behaviour or patterns. Rather, they merely reflect the practices, policies and programmes, bias and political or financial needs of a particular department (Richardson, Schultz & Crawford 2019: 8).

Their study is based on the documents and outcomes of legal investigations commissioned by the government, or on court settlements under the supervision of the Federal Court of Justice, consensus agreements or other agreements based on legal investigations. They looked at 13 regional administrative units or "jurisdictions",[69] which were using or had previously used predictive policing systems in trials during the periods under investigation. The distinctive feature here is that while the administrative units were under legal investigation or under judicial settlement procedures, it was discovered that their police authorities were engaged in corruption, ethnicity-based unequal treatment or other illegal practices. In nine of these, the authors collected evidence proving that the systems were used during periods of illegal police practices. Three administrative units (Chicago, New Orleans and Maricopa) were presented in detailed case studies leading to the conclusions on the problem of using dirty data (Richardson, Schultz & Crawford 2019).

The authors attribute these situations to a lack of supervision and verification measures in the collection, analysis and use of police data, which was neither carried out by an authority nor by the producers of such systems (ibid., p. 20). Furthermore, no evidence was found in the cases that manufacturers or providers of the systems independently verified the police data used (ibid., p. 7). Providers of these systems, which themselves point to biased police data records, would not sufficiently take into account the structural and systematic errors in this data. The authors added that detecting and correcting such errors would be too great a challenge (if not an insurmountable one), and this raises doubts as to whether a distinction can be made between problematic and less problematic categories of data. Even if a differentiation were possible, this would only be possible for one respective administrative unit and would hardly allow for comparative or aggregated conclusions (ibid., pp. 8f.). Moreover, without an "empowered and independent authority" (ibid., p. 24), it is to be expected that potentially unlawful and discriminatory police practices and data based on them could go untreated and uncorrected, particularly since there would be few political and institutional incentives for self-evaluation and reform (ibid., pp. 24f.).

4.11 Judicial and penal system

Case 35: Ethnicity-based discrimination in criminal recidivism systems

One of the most frequently cited cases of algorithmic systems for decision support is the COMPAS system (Correctional Offender Management Profiling for Alternative Sanctions) of the company Northpointe (renamed to equivant). It is used for risk prediction to assist judges in many US states to determine the risk of recidivism in case of early release. The COMPAS system uses a variety of personal characteristics and a proprietary algorithm that determines the probability of re-arrest for each person charged. Based on publicly available documentation and crime statistics, research by the journalist association ProPublica (Angwin et al. 2016) found that the predictions for black defendants systematically overestimated the risk. Of those who were not re-arrested, 45 per cent of the black defendants had been identified as high risk. In comparison, only 23 per cent of the white defendants who were not re-arrested had been put at high risk. With regard to the accuracy of the predictions they concluded that the probability of black people being falsely labelled as high risk is twice as high as for white defendants. In a reaction of the company, the statistical approach of ProPublica was criticised and its own computation methods were presented. They showed that people with similar risk scores, whether black or white, had the same probability of being arrested again (Dieterich, Mendoza & Brennan 2016). However, it was found that both parties used different approaches and fairness concepts, which could not be applied and fulfilled simultaneously (Chouldechova 2017; Eckhouse et al. 2019).

[69] These are Boston, Chicago, Ferguson, Miami, Maricopa County, Milwaukee, New Orleans, New York, Newark, Philadelphia, Seattle and Suffolk County.

In 2016, the Wisconsin Supreme Court issued a court decision that found a "due process violation" in the use of the COMPAS system for risk assessment.[70] The decisive factor in the dismissive court decision was the undermining of the right to a court ruling based on precise information (Freeman 2016; Citron 2016). Although the court confirmed that the evaluation of individuals on the basis of group data and generalising statistical evaluations is fundamentally legally problematic, the Supreme Court relativised this problem by stating that the risk scores of the COMPAS system are only one of several information bases used by the judges. The Supreme Court's decision is criticised for not sufficiently taking into account the (possibly dominant) importance of computer recommendations for human decisions or the risk of automation bias (Citron 2016; Freeman 2016: 96). It was also previously stated that such systems were not continuously checked for accuracy (Klingele 2015) and had not been examined for hidden bias (Starr 2014), which was also critically discussed in the Supreme Court decision (according to Citron 2016).

Case 36: Comparison of risk analysis system with machine-learning methods

Tolan et al. (2019) compare the system SAVRY ("structured assessment of violence risk in youth"), a risk analysis system for predicting the probability of recidivism in juvenile justice, using machine-learning methods with regard to group fairness for the protected characteristics of gender and nationality. The SAVRY system calculates an overall score by entering assessments by evaluative experts on risk and prevention factors of the respective persons concerned. The assignment to risk classes in the final decision was also carried out by the experts and was therefore not algorithmic. The experts had been previously informed that the recidivism rates for male and female offenders were usually different.

The SAVRY system was compared with methods of supervised machine learning. The comparison was made on the basis of numerous input data on demographic information and the criminal history of the accused. A data set on juvenile delinquency in Catalonia was also used. As a result, it was shown that the SAVRY system is considered "fair" in terms of various measures of fairness,[71] while machine-learning methods tended to discriminate against male and foreign defendants and those of certain nationalities.

In addition to data analyses, they also used tools for the interpretation of ML procedures. The researchers emphasise the tension between predictive accuracy, where ML methods perform better, and fairness measures, where they perform worse. One explanation is that the basic distribution of recidivists in the various population groups ("base rates") affects the predictions of ML procedures. For example, the ML methods would adopt the empirical correlations between population characteristics and relapse rates.

In addition, they discuss possible technical countermeasures,[72] which would actually lead to further deterioration. (a) Scrubbing protected characteristics would be useless because many other characteristics have correlations with them. (b) The use of different thresholds (of fairness measures) for various legally protected characteristics may lead to incorrect classifications, with the result that unidentified recidivists may endanger public safety or young people may be wrongly imprisoned. (c) Adaptation of the model or classification algorithm, e.g. by inserting some kind of correction variable without understanding the underlying mechanism, may in turn lead to other instances of discrimination, stigmatisation or injustice.

[70] See judgement of the Supreme Court of Wisconsin "State v. Loomis", (881 N.W.2d 749, 763-64 (Wis. 2016)).

[71] The fairness measures of demographic parity and error rate balance were examined. See also Section 6.1.1 on fairness measures. In predicting recidivism probabilities, equal demographic parity means that any person with a protected characteristic has the same probability of being classified as a recidivist as a person in a reference group. An equal error ratio means that every person with a protected characteristic has the same probability of being falsely classified as a recidivist as a person in a reference group. The same false negative rate and the same false positive rate were examined. Cf. Tolan et al. (2019).

[72] See also Section 6.1.1.

4.12 General cases of artificial intelligence

Case 37: Unequal accuracy in facial recognition systems

The researchers Klare et al. (2012) demonstrate in an experimental study that six investigated facial recognition systems, which are used by law enforcement agencies in the US, systematically recognised images of persons with the markers “women”, “ethnic origin” and persons between the ages of 18 and 30 with less accuracy. The systems included three commercial systems that were revealed to be less accurate in all tests. Based on the overall test results, the authors conclude that all groups of people subject to recognition by the systems at a later stage should be sufficiently represented when selecting the training data.

Regarding these outcomes, Garvie emphasises that the error rates in accuracy may in practice lead to more frequent controls of innocent persons from the less accurately recognised groups. In this context, the authors point out that facial recognition systems have hardly been tested for possible unequal treatment based on ethnic origin (Klare et al. 2012: 53-56).

Case 38: Adoption of gender-based stereotypes in machine text analysis

Bolukbasi et al. showed in a process of text analysis using machine learning of natural language (“natural language processing”) that gender-based stereotypes concealed in the texts are repeated in the outcomes. The text analysis method investigated, known as “word embedding”, converts text data into (number) vectors, which are made available for further machine processing. The method is trained with regard to the simultaneous occurrence of words in certain text corpora and the search for certain patterns of coherence between words. The geometry between the vectors displays the semantic connection between the words, which reveals the stereotypes in the form of calculated word associations (Bolukbasi et al. 2016).

In the study, the publicly available method “Word2Vec” was examined using the Google News text corpus, which contains three million English words. The assumption was that it would contain little gender bias as it was mainly written by professional journalists. However, stereotype-like outcomes of the machine-learning process were demonstrated in two ways. (a) For assigned job titles, it was shown that the word “she” was accompanied by words such as “homemaker”, “nurse”, “receptionist”, “librarian”, “socialite”, and “social worker”, etc., whereas the word “he” was associated with words such as “maestro”, “skipper”, “protege”, “philosopher”, “captain”, etc. Crowdworkers[73] were tasked with assessing whether the computationally assigned terms were female, male or neutral stereotypes. (b) When analogous pairs were created by machine in the form “man is to king as woman is to queen”, the method produced words that were considered to be gender stereotypes in 29 of the 150 analogy words, while 72 of the 150 words were judged to be gender-consistent. This assessment of the computational assignments was also carried out by commissioned crowdworkers. They also present a procedure to reduce gender bias. To this end, they changed the gender link for certain words; for example, this produced a female and male connotation for the term “nurse” (Bolukbasi et al. 2016).

Machine text analysis is used in a wide range of applications, such as the automated analysis of documents, resumes or written communication in social networks, and automated ranking in search engine results, product recommendations or machine translations. If the “embedding” algorithms thus generated, which

[73] “Crowdworkers” are people who provide and perform mostly minor computer work services via online platforms on the internet, such website or software testing, copywriting, photo categorising, the programming of software components or design work. In most cases, they are not permanent employees at a company. The activity of the crowdworkers is also assigned to an area called “crowdsourcing”, in reference to outsourcing.

have adopted stereotypical word relationships, are used in such applications, problematic outcomes can occur in that traditional gender roles are perpetuated.

Case 39: Adoption of cultural stereotypes in machine text analysis

Researchers (Caliskan, Bryson & Narayanan 2017) demonstrated that common machine-learning methods (in this example, word embedding) can "learn" everyday cultural stereotypes from text files. The machine learning methods were trained on a text corpus with normal human language from the internet, the "common crawl" corpus. The researchers showed that when machine-learning methods are applied to texts, the outcomes are just as stereotypical as was previously demonstrated by other researchers for human behaviour using association tests. This was demonstrated by the fact that machine-determined associations between words were similar to those of humans. For example, they revealed that the ML methods associated female names more frequently with words for "family" than for "career", as compared to male names. (Similar results were obtained for words such as "mathematics" or "science", which were associated with male terms, and "arts", which was associated with female terms.) They also confirmed previous research results that European-American names were more frequently associated with terms for "pleasant" than African-American names.

Based on these results, they conclude that AI systems that learn and reproduce the characteristics of language adopt cultural conceptions from the past, some of them stereotypical. This is particularly problematic if these systems are used in current applications, such as online text translation, or if decisions are left to them, such as the reviewing of CVs, where systems with inherited cultural stereotypes would produce outcomes that are biased (Caliskan, Bryson & Narayanan 2017).

Case 40: Adoption of gender and ethnic stereotypes in machine text analysis

Similar to the previous example, Garg et al. (2018) demonstrated that the machine-learning procedure "word embedding" is also suitable for quantitatively recording widespread gender-related stereotypes and attitudes towards ethnic minorities and their changes over time in large quantities of text and over historical time sequences. They demonstrated this with texts from the 20th and 21st centuries in the US, including the Google Books/Corpus of Historical American English and the New York Times Annotated Corpus. For example, in 1910, the word "women" was associated with such words as "charming", "placid", "delicate", "passionate", "sweet", "dreamy", "indulgent", "playful", "mellow" or "sentimental". For 1990, these were "maternal", "morbid", "artificial", "physical", "caring", "emotional", "protective", "attractive", "soft" and "tidy". The authors point out that stereotypes learned automatically become problematic when they are used in "sensitive" products and services, such as search engine ranking, product recommendations or automated translations (Garg et al. 2018: E3635).

Case 41: Unequal accuracy by gender and dialect in automatic subtitles of the video service YouTube

In a scientific study, Tatman showed that the YouTube platform's service for generating automatic subtitles on uploaded videos ("automatic caption") has varying levels of accuracy, with significantly lower accuracy in recognising female speech and for videos featuring people who speak in a Scottish dialect. The service is based on a machine-learning process. As one of the possible reasons, the author suspects insufficient training data (Tatman 2017: 57).

Case 42: Ethnicity-based unequal accuracy in speech recognition systems

Blodgett et al. examined four common speech recognition systems or speech recognition cloud services (langid.py, IBM Watson, Microsoft Azure and Twitter's internal identification service) using the dialect of

African-American English used in the Twitter social network.[74] They used a publicly available Twitter corpus with 59.2 million tweets. As a result, they found differences in the accuracy of speech recognition, with greater accuracy for text messages assigned to white authors than text messages assigned to African-American authors. The differences are particularly high for short text messages. Since data from social networks is often used for sentiment and opinion analysis on products or political figures, the fact that the opinions of African Americans are less well captured than those of white participants constitutes a social problem, according to the authors (Blodgett & O'Connor 2017: 3).

Systems of speech recognition are used in personal assistance systems (e.g. Siri, Alexa, Amazon Echo), chatbots or automated telephone systems. In the case of insufficient training data sets, in which certain population groups are not sufficiently represented, their dialects or accents cannot be sufficiently learned; as a result, the members of these population groups are less easily recognised or understood in applications.

Case 43: Unequal detection rates by gender in machine opinion and sentiment analysis

A study of machine-learning methods for the recognition of opinions and sentiments ("sentiment analysis") based on texts revealed inequalities between the genders, with the methods being better suited to determining the sentiments of women than those of men. The data basis was compiled from hotel and restaurant ratings on the travel platform TripAdvisor.com for the United Kingdom. The gender differences revealed appear to indicate that women's opinions are slightly overrepresented in opinion and sentiment analyses, as a higher proportion of male sentiments are not recorded (Thelwall 2018). However, the author admits that the outcomes for this specific data set cannot be generalised. That being said, he showed that biases were not present in the data set, but only came about as the outcomes of seemingly objective machine-learning processes. Therefore, when reviewing the algorithms, he argues for testing not only the input in the form of the data and algorithms, but also the output of the system, including the different groupings, e.g. by gender, ethnic origin, etc. (Thelwall 2018). Automated opinion and sentiment analyses are used in marketing to evaluate products or services and can record the sentiments of a large number of users, e.g. social networks, in almost real time.

Case 44: Unequal detection in 219 sentiment and opinion analysis systems

Kiritchenko and Mohammad, together with cooperating institutions of the evaluation study SemEval, examined over 219 systems of natural language processing for sentiment analysis, some of which achieved different results depending on gender or ethnic origin. The data basis was a uniform corpus of texts, the "equity evaluation corpus" (EEC), which was compiled from 8,640 English sentences. A network of cooperating institutions carried out the tests on the systems. The study was based on a systematic quantification of inequalities by gender and ethnic origin. The outcomes were presented in the form of systematically higher or lower sentiment intensity scores (Kiritchenko & Mohammad 2018).

Case 45: Unequal recognition according to skin colour in pedestrian detection systems

Wilson et al. tested the inconsistent prediction quality of object detection using machine-learning techniques for detecting pedestrians with different skin colours. They investigated whether the unequal prediction rates were due to the time of day (or light conditions), the degree to which the individuals were concealed, or weightings in the objective function of the machine-learning process. Their results showed that standard models of object detection trained on standard data sets have a higher accuracy for images

[74] A previous publication by Blodgett, Green & O'Connor (2016) describes the data production. To do this, they had to assign the tweets to different ethnic population groups by comparing the location of the target region of the Twitter message with the demographic data of the official population statistics. The study of the three speech recognition systems langid.py and two internal Twitter systems with poorer recognition of African-American dialect was also presented.

containing persons with “light” skin (skin types lower on the Fitzpatrick scale) than for “dark” skin. This inequality even increased when researchers removed covered persons (as a possible source of the differences in accuracy) from the images (Wilson et al. 2019).

Case 46: Unequal recognition according to skin colour in commercial facial recognition systems

Buolamwini and Gebru conducted experiments with commercial face recognition systems from Microsoft, IBM and Face++ revealing that facial recognition worked with varying degrees of efficiency for different genders and people with different skin colours (classified according to the Fitzpatrick scale for skin types). The systems generally detected male faces (with an error rate of 8.1 per cent) better than female faces (with an error rate of 20.6 per cent). Likewise, they detected “light-skinned” faces (with an error rate of 11.8 per cent) better than “dark-skinned” faces (with an error rate of 19.2 per cent). For the tests, they compiled their own comparative data set with the personal images of 1,270 African and European parliamentarians, the “Pilot Parliaments Benchmark” (PPB). In their opinion, existing comparative data sets for determining the accuracy of the systems were characterised by an over-representation of male and “light-skinned” person types and an under-representation of female and dark-skinned person types (Buolamwini and Gebru 2018).

The publication of the paper by Buolamwini and Gebru has led to reactions from companies. For example, IBM adapted the data set for optimising its software to the Pilot Parliaments Benchmark (Puri 2018). Other manufacturers also improved their facial recognition accuracy. These reactions were analysed and discussed in another study by Raji and Buolamwini, which examined the systems of IBM, Microsoft, Face++, Amazon and Kairos (Raji & Buolamwini 2019).

Case 47: Unequal recognition by skin colour in a commercial facial recognition system

During a test of the facial recognition system “rekognition” of the company Amazon, the American Civil Liberties Union (ACLU) discovered that the system incorrectly recognised 28 members of the US Congress, i.e. falsely recognised them as wanted persons, and that the rate of false recognition in PoC (People of Colour) was disproportionately high. In testing the system, which is publicly available via an online service, the images were compared with a database of 25,000 publicly available photos of people who have been arrested. According to the civil rights organisation, Oregon police forces are using the system to match pictures from “body cams” to a database of mug shots. At the time of reporting, a protest campaign was underway against the use of the monitoring system (Snow 2018).

5. Causes of Risks of Discrimination

By way of illustration, **two types** (albeit interrelated) of discrimination risks are distinguished below:

(1) risks of discrimination resulting from the use of algorithms due to their special technical properties (see Section 5.1) and;

(2) risks that arise through the use of algorithmic and data-based differentiations and decision-making systems themselves and occur as societal risks (see Section 5.2).[75]

In this context, Powles and Nissenbaum point out that the current focus on technical solutions for issues caused by algorithms and AI is also associated with societal dangers: (1) Societal unequal treatment is a social issue and attempts to solve it using the technical logic of automation are always inadequate. (2) Even if technical problems are successfully solved, this does not say anything about the legitimacy of the intended use. For example, if the problem that persons with a certain skin colour are less easily recognised by a facial recognition system were to be solved, the facial recognition or identification system could still be used for surveillance activities with disproportionate restrictions of fundamental rights of personality protection. (3) Finally, a unilateral focus on technical solutions can divert attention and (research) resources from solving the actual societal issues of unequal treatment (Powels & Nissenbaum 2018).

5.1 Risks in the use of algorithms, models and data sets

Risks of discrimination can result from the special technical properties of the procedures of algorithm-based data analysis and automated decision procedures, whereby the following focuses primarily on data mining and machine learning (according to Calders & Žliobaitė 2013; Barocas & Selbst 2016; Kim 2016; Lehr & Ohm 2017; Zweig, Fischer & Lischka 2018; Schweighofer u. a. 2018; Zuiderveen Borgesius 2018; Favaretto, De Clercq & Elger 2019; Tolan 2018; FRA 2019).

5.1.1 Risks in the development of algorithms and models

In data mining and machine learning, the purpose of analysis is usually determined, i.e. what to predict or estimate and how to measure it. From verbal descriptions of the purpose of the analysis, a calculable **objective function** needs to be determined, which is to be optimised on the basis of historical or other data. In this context, an objective function is a mathematical expression of the objective of the entire process, usually in the form of target variables (or outcome variables) that are minimised or maximised (Lehr & Ohm 2017: 671). If the processes are used in organisations such as companies or authorities, the objectives, provisions and requirements of the organisation must be transformed into calculable functions and values. Specifications such as the “creditworthiness” of applicants, the “efficiency” of employees or the “value” of

[75] Other authors make similar classifications, cf. Citron & Pasquale (2014), Zarsky (2016), Crawford et al. (2016: 67), Britz (2008: 120-136), Gandy Jr. (2010), Eckhouse et al. (2019).

customers must be analysed or determined and converted into calculable values or formalised. System developers and analysts are therefore tasked with determining what constitutes “good creditworthiness” or “good employees”, for example (Barocas & Selbst 2016: 678f.).

To determine what constitutes a “good” employee, for example, there are usually multiple and varied target variables available. The example of personnel selection can be used to illustrate that risks of discrimination can already arise when selecting the target variables (Barocas & Selbst 2016: 680). For example, when decision-making systems for personnel selection are based on target variables formed by evaluating seniority, certain groups that usually have a higher employment turnover rate (such as women) are systematically disadvantaged, even though they can perform equally well or better. The problem is also illustrated in Case 24 of a computer system for assigning patients to treatment programmes, in which an objective function that is unilaterally aimed at costs and fails to sufficiently reflect health aspects led to ethnicity-based discrimination. In other words, it can be stated that the very selection and determination of the target variables can become a risk of discrimination that is virtually “programmed” into algorithms and computer systems.

Risks of discrimination may also arise from the selection and determination of the labels of the categories (“labels” or “class labels”). This is important for the later phase of application (to new records), because the automated assignment of persons to categories or groups is carried out with these labels. Normally, when selecting and determining the categories or labels, all possible values of the target variable are divided into mutually exclusive categories. Since the selection of the objective function with target variables and the labelling of the categories already influence the outcomes of the machine-learning process, they form the basis for risks of discrimination in the other steps of the data mining or machine-learning processes (Barocas & Selbst 2016: 678, 680). The decision-making rules formed on this basis tend to reflect existing prejudices and bias through the subjective decisions and systematically lead to disadvantages (Barocas & Selbst 2016: 682): (a) Objective labelling is less likely to cause such bias because there is agreement between different stakeholders on whether the characteristic is met or not and individual interpretations are not necessary (Calders & Žliobaitė 2013: 48). For example, everyone can clearly identify and understand the characteristics “loan repaid or not” or “tested for alcohol or not”. In particular, however, if (b) the selection of labels contains subjective interpretations, risks of discrimination may arise. For these characteristics, not everyone can clearly and unanimously understand or share how to define the characteristic or make it measurable, such as the characteristic ‘a good fit’ of applicants for a position (Calders & Žliobaitė 2013: 48).

Furthermore, there may be risks in the **selection of influencing variables** when training models. Data mining or machine-learning methods generate models that can also contain a large number of influencing variables (referred to here as “features”, also called predictors). A feature selection is necessary when it is technically impossible to include all influencing variables. However, the selection decisions can produce models that no longer have a sufficient level of detail to detect critical differentiation points. This can lead to systematically higher error rates in the detection of certain groups of persons who may also have protected characteristics. Although the characteristics used may be statistically sufficient, they may simply be inappropriate for the group of persons not sufficiently recognised. In other words, they are not universally valid.[76] At this point, the peculiarities and problems of statistical discrimination become especially clear. In favour of efficiency objectives, surrogate variables are used for differentiation which are too coarse to do sufficient justice to the characteristics of the data subjects. This results in an injustice by generalisation (for further details, see Section 5.2.1).

[76] Cf. Barocas & Selbst (2016: 688) with reference to Schauer (2003).

5.1.2 Risks in the compilation of data sets and characteristics

In statistics and computer science, it has been known for some time that biased data sets lead to discriminatory models. They also occur in data mining and machine-learning methods (Custers 2013; Barocas & Selbst 2016: 680-690; Lehr & Ohm 2017; FRA 2019). Under-representation or over-representation of groups of persons and the reproduction of previous unequal treatment in the used data sets and correlations of surrogate variables or proxies to protected characteristics can lead to risks of discrimination in the outcomes of the differentiation systems.

Risks of discrimination arise from intentional or unintentional **under-representation or over-representation of groups of people** or the complete scrubbing in evaluated data sets or training data – in other words, when data on certain groups of people is not available in a ratio that would be required for correct representation (Calders & Žliobaitė 2013: 47-49; Barocas & Self 2016: 684-686). Possible causes for this are:

(1) The data relates to situations in which **unequal treatment and unequal distribution** of groups of persons existed or currently exist. For example, if a model were to be built by analysing data on existing or historical employment relationships, perhaps in order to establish a correlation between being “a good fit” for a job and certain characteristics from job applications or personnel records, and if women were under-represented in these employment relationships – for example, because they were denied access to these employment relationships – then this imbalance is also to be expected in the composition of the characteristics of the model (Calders & Žliobaitė 2013: 50). This issue can be found in Case 1 of Amazon’s personnel search system, in which the models were trained with data sets that were biased to the detriment of the proportion of women. Or, if historical data sets are used for modelling, the data may correctly reflect previous inequalities. However, if the conditions have changed in the meantime, the models do not reflect the current conditions correctly. For example, if a model based on historical data on income ratios between women and men were to be used, for example, to target advertising towards economically promising female customers, the historical inequalities in income ratios would be correctly reflected. However, if occupational activities and income relations have changed in the meantime, the generated selection and relations of characteristics would not correctly reflect the current conditions in the model (Calders & Žliobaitė 2013: 50f.).

Many of the examples in Chapter 4 featuring analyses or uses of machine learning or data mining methods reveal risks of discrimination due to biased (training) data sets that reflect (previous) unequal treatment, stereotypes and discrimination.[77]

The example of the image search engine results that reinforce stereotypes (Case 15) also shows that qualitative bias in the data records can occur if the image data are labelled as stereotype-laden. The special study on the data bases of systems of predictive policing in the US (Case 34) makes it clear that the mapping of irregular police practices in the data sets could call into question the legitimacy of the systems as a whole.

[77] See cases 1, 8, 34, 35, 37, 38, 39, 40, 41, 45 and 46.

(2) In addition, data may originate from the use of services or products that certain groups of people use less or not at all (Lerman 2013). This may be the case, for example, for evaluations of the use of certain IT services (e.g. online services or broadband use), in particular data sets originating from online social networks (Hargittai 2015). (3) Similarly, for cost considerations data collection may be deliberately limited or data sets are used that are available at low cost but are unsuitable. (4) Also for reasons of data protection or privacy, certain surveys cannot be carried out (Calders & Žliobaitė 2013: 52f.). (5) Over-representation of a population group may occur when activities involving data collection are disproportionately concentrated on a specific group of people and thus a disproportionate number of survey subjects are covered (Calders & Žliobaitė 2013: 51). The latter is the case with police activities, for example, which, as an outcome of data analysis, take place in areas that are already heavily controlled and can thus lead to amplification effects (Case 32).

The simple removing of protected characteristics with the intention of avoiding discrimination can itself create risks of discrimination (Calders & Žliobaitė 2013; Žliobaitė & Custers 2016). Models that have apparently "neutral" features or **proxies** instead of protected characteristics can also lead to risks of indirect discrimination if there is a correlation between the proxies and the protected characteristics (Barocas & Selbst 2016: 720-722). For example, in a model designed to determine creditworthiness, there may be a link between place of residence and ethnicity.

Even if the legally protected characteristic "ethnicity" were to be removed, it might be possible to infer ethnic origin from the place of residence if a majority of persons of the same ethnic origin inhabit that place of residence. In these situations, the use of "neutral" factors would also penalise groups that are actually protected. In general, the problem arises when the characteristics are not independent of each other, as it is then impossible to determine which characteristic contributes to the model and to what extent (Calders & Žliobaitė 2013: 47). The problem is also illustrated in Case 8, where the scrubbing of gender indicators did not improve inequalities in occupational classifications and where, in the authors' opinion, the use of machine-learning methods can further increase gender disparities.

Overall, more variable types or dimensions can be processed with data mining and machine-learning methods than with "classical" statistical methods. This also increases the risk of (unnoticed) correlations to protected characteristics. Furthermore, sensitive personal data may be necessary precisely in order to identify and compensate for discrimination, if necessary in an aggregated state by means of statistical studies (e.g. FRA 2018: pp. 9f. on the characteristic 'ethnic origin').

Similar problems may arise when models or systems contained in them are applied in other areas of life or contexts and the original population used for the modelling does not match the population of the new scope.[78] This raises fundamental questions about the transferability and (commercial) tradeability of such systems. For example, Schweighofer points out that the COMPASS system (Case 35) was originally intended to assist with probation decisions but is also used for criminal sentencing in some states (Schweighofer et al. 2018: 39).

[78] This can lead to poorer recognition of certain groups of people in AI-based facial or speech recognition systems, for example. See cases 37, 42 or 46.

5.1.3 Risks with online platforms

Online platforms differ from 'simple' websites in that they bring together different actors and offer them an opportunity for social and economic exchange by providing the means of communication (e.g. in online social networks) or the trading function or 'matchmaking' (e.g. in online platforms for employment services). Many examples[79] revealed unequal treatment and discrimination in online platforms. This was caused in part by allowing users to rate and select other users. As Edelman and Luca aptly put it, "Full of salient pictures and social profiles, these platforms make it easy to discriminate [...]" (Edelman & Luca 2014: 10).[80] Algorithms can perform the analysis of personal data and the categorisation or ranking of individuals or administer the information that is adjustable or visible and usable for the participants of the interactions and transactions.

In addition, algorithms have other discrimination-related functions in connection with online platforms. In Case 2 it is suspected that, on one of the employment agency's online platforms, the ranking of search results is also based on the reviews and ratings of employers provided by applicants. As the ratings were already gender and ethnically biased, the bias in the ranking of the search results will continue. Being found on online platforms can have considerable economic consequences on access to jobs and income opportunities. Case 27 illustrates that biased restaurant ratings can continue to exist in unequal treatment in the algorithm-based risk prediction of state control. The continued algorithmic bias and discrimination can thus lead to societal accumulation and amplification risks (see Section 5.2.2).

In addition, certain online platforms, in particular search engines or "social" online networks, are companies whose aim is to generate attention and that rely on individual or group advertising as their main source of income. Algorithms also enable the pricing and market mechanisms (e.g. auction mechanisms) for advertising and customer selection. Case 17, Case 18, Case 19 and Case 20 show that these algorithmic marketing mechanisms can be the cause of unequal treatment and discrimination. Algorithms can lead to discrimination even if users have not actually made discriminatory settings, as in Case 20, for example, where the selection of target groups for advertising is algorithm-based.

5.1.4 Intentional discrimination and obfuscation in and by computer systems

All of the sources of risk mentioned could also be used by developing and applying entities to disguise intentional discrimination, i.e. the data sets could be knowingly selected with bias or the very ones known to reflect former unequal treatments could be used. Similarly, models with characteristics that do not correctly detect individual groups prone to discrimination could be deliberately chosen (Barocas & Selbst 2016: 692; Dwork & Mulligan 2013; Kim 2016). Since data mining and machine learning allow legally protected characteristics from data sets that do not contain them to be derived, decision-makers can potentially discriminate even with data sets without protected characteristics. If they were called to account, they could claim to have used only "non-protected" characteristics. In areas where it is already

[79] See cases 2, 3, 4, 5, 6, 7, 9, 10, 17, 18, 19, 20 and 26.

[80] For example, in a scientific study, researchers reveal unequal treatment in "peer-to-peer lending", i.e. the granting of loans via online platforms. They use data on transactions in the online marketplace Prosper.com as an example. Their findings include the fact that credit-seekers who have an African-American appearance in their profile pictures are less likely to have received financing, with a probability of 25 to 30 per cent. However, since it was also possible to prove preferential treatment for African-American credit applicants in terms of net interest rates, the authors were unable to draw any clear conclusions about the existence of preference-based or statistical discrimination. However, in this example, algorithms do not appear to be the cause of discrimination (Pope & Sydnor 2011).

difficult to prove indirect discrimination, such as employment, such practices could make it even more difficult to prove (Barocas & Selbst 2016: 692f.).

5.1.5 Insufficient incentives for revision or abolition

Kim uses the field of employment to illustrate that a competitive approach, market-economy interests and the efficiency-mindedness of the users cannot provide sufficient incentives to question the application of analysis and decision-making systems and, if necessary, to change or abolish them, even if they lead to discrimination. She explains this by citing the following reasons: (a) Systems with a risk of discrimination can still be "accurate" enough so that their application is not questioned; (b) feedback effects can stabilise the "accuracy" of the system and; (c) systems can even be efficient because they discriminate (Kim 2016: 892897).

For example, the accuracy of one criterion may mean that the verification of the suitability of the entire system is neglected. Methods for machine-learning analysis also only have a limited amount of data with limited characteristics. Therefore, the resulting models also represent only a limited range of criteria potentially relevant for decisions. In particular, the more "successful" (in the sense of accuracy with one criterion) an analysis and decision-making system is, the more the focus can be on this criterion. The entities applying the system then no longer have an incentive to question its conclusions and mechanisms, even if it systematically causes discrimination. Above all, they will not question whether completely different decision-relevant criteria are relevant and need to be considered (Kim 2016: 894f.).

5.2 Societal risks of algorithmic differentiation

Even if risks of discrimination through algorithms and data were to be avoided to the greatest possible extent, societal risks of discrimination can result from the use of algorithmic differentiation methods and automated decision-making systems themselves. They cannot be eliminated through technical solutions, but require societal solutions through appropriate forms of political handling and, if necessary, regulation. Such risks are also called unintended consequences, negative consequences, social costs or externalities (Gandy Jr. 2010: 36-39) because the applying entities do not take them sufficiently into account when deciding on the use of differentiation methods and they are incurred by affected persons or third parties. However, the societal risks can be drastically increased and intensified if additional systematic errors in algorithms and data sets are present.

5.2.1 Group membership and injustice by generalisation

As illustrated, algorithmic and data-based differentiation decisions often involve the phenomenon of statistical discrimination, since differentiations take place along surrogate information that either consists of the protected characteristics or has correlations to them (see Section 3.3). Typically, the surrogate information is often generated by analysing data on groups of people. This is largely the case with the procedures of data mining and machine learning. This raises the question of whether it is fair for individuals to be assessed using data about other people, i.e. on the basis of groups to which the individuals concerned need not necessarily belong themselves (e.g. Eckhouse et al. 2019: 198f.).

The fundamental problem of statistical analyses with the use of aggregating parameters or measured variables is that they are merely statements about characteristics of a certain population or an aggregation of persons. They should actually refer to only one aspect of the population, but the parameters are often used as if each member of the grouping had the characteristic. This gives the parameters the features of stereotypes (Gandy Jr. 2010: 34).

In the case of prediction algorithms, such as the computation of risk scores in particular, the prediction outcome is not the probable future behaviour or condition of the persons concerned, but usually an extrapolation of previous ratings of other persons by other persons. This becomes particularly evident in Case 22 on risk scores in the Finnish case of online lending and Case 35 on the COMPAS system with risk scores on the probability of recidivism.81 In the court decisions it was pointed out that the ratings of individuals based solely on statistical analyses of data, which in turn reflect ratings of other individuals or are group data, are legally problematic. This is particularly the case when other information about the individuals is not used in their assessment. The case studies on rating data for the Uber transport service (Case 26), risk assessments with restaurant ratings (Case 27) or risk assessments with quasi neighbourhood ratings in child protection (Case 28) also illustrate the problem.

This causes risks of misinterpretation and wrong conclusions[82] on the basis of (supposed) group membership (or group membership by region) (Schauer 2018; Lippert-Rasmussen 2007; Kamp & Weichert 2005: 51; SRP 2018: 48; Zweig, Fischer & Lischka 2018: 25). They are particularly problematic when this leads to stigmatisation due to misattribution of negative personal characteristics, like in cases of statistically determined "unreliability" in credit decisions (Britz 2008: 124) or the "propensity to commit crimes" in sentences of imprisonment (Eckhouse et al. 2019).

Such statistical and algorithm-based decisions then do not do justice to the individual case. Adverse effects result from the fact that an assumption is made about one or more persons who possess a certain characteristic, which may apply to the majority of those who possess the characteristic, but does not necessarily apply to the specific person or persons in the individual case (according to Britz 2008: 120f.). From a constitutional point of view, with regard to the principles of equality of the Basic Law (Article 3 GG)[83] in statistical discrimination, **injustice by generalisation** occurs when "atypical" persons are excluded from certain occupational activities, for example, according to a surrogate characteristic (e.g. age), but would actually still be able to perform these activities (Britz 2008: 211). In the same way it is possible to speak of an incompatibility with doing justice on an individual case basis, since, "In cases of statistical discrimination, a person is judged and treated on the basis of stereotypical perceptions of a person because of a certain (proxy) characteristic, without any appreciation of his/her actual characteristics." (Britz 2008: 12, translated from the German original)

It is therefore a question of unequal treatment when specific features of the individual case are disregarded. Since this is done – as is characteristic of statistical discrimination – to overcome information deficits in a cost-effective manner, and since risks of injustice are thus almost "accepted", it leads to situations in which the values of efficiency and equity, which are actually incommensurable, are weighed against each other (Gandy Jr. 2010: 36f.). These can hardly be solved by technical or organisational improvements, but require **societal considerations and decisions** to be made in political and law-making processes.[84] The universally

[81] This also applies to risk scores in the social sector, as illustrated by Case 28.

[82] Here, the risks described in Section 5.1 are directly related.

[83] "(1) All persons shall be equal before the law. (2) Men and women shall have equal rights. The state shall promote the actual implementation of equal rights for women and men and take steps to eliminate disadvantages that now exist. (3) No person shall be favoured or disfavoured because of sex, parentage, race, language, homeland and origin, faith or religious or political opinions. No person shall be disfavoured because of disability." Article 3 GG (translation by Deutscher Bundestag 2019).

[84] See also Section 6.4.2.

binding balancing should actually be provided by law, in particular by the AGG. Due to the relatively unspecific, general clause-like exemptions of the AGG (Britz 2008: 72) and whenever new forms and applications of statistical discrimination emerge, the question of legitimacy must always be posed anew for each individual situation. To this end, algorithmic procedures would have to undergo the usual proportionality test for their legitimate purpose, their suitability, necessity and appropriateness (Britz 2008: 151-179).

In doing so, the particular characteristics of the respective differentiation situations have to be taken into account, in particular whether differentiations are based on group data or on recordings of individual behaviour, but also the degree of accuracy of the algorithmic outcomes or the error rates and other technical risks as well as knowledge about causal relationships[85] or the extent or severity of the disadvantage resulting from the wrong conclusions in differentiation decisions (e.g. refusal of a loan, non-employment, higher insurance rates).

The degree of disadvantage, for example, is determined by the goods that the disadvantaged person is deprived of and the severity of the restriction of the person's opportunities for development (Britz 2008: 125). For example, decisions on imprisonment and deprivation of liberty and decisions on the distribution of opportunities for personal development (e.g. on educational or occupational access) must be assessed differently from decisions on the selection of targeted advertising for consumer goods. Nevertheless, for financial reasons, algorithm-based systems are also used to support decisions with serious consequences for the development of personality in court proceedings (Eckhouse et al. 2019).

5.2.2 Accumulation and amplification effects

The risks and negative effects of economically rational differentiation can accumulate and intensify. Overall, there may be cumulative disadvantages in the sense of limiting life chances and personal development, income inequality, the degree of political involvement and the enforcement of justice in the legal system (Gandy Jr. 2010: 37). These are not fundamentally new risks first caused by algorithms, but the risks of discrimination caused by the use of algorithms can also contribute to accumulation and amplification effects. Risks and effects occur in particular when a characteristic prone to discrimination is used as surrogate information for differentiation decisions, and there is a mutually reinforcing effect of stigmatisation, impairment of self-representation,[86] discrimination through false assignments to categories and the resulting difficulty in accessing goods that serve to develop personality.

A further amplification effect results if existing unequal treatments are perceived within the population and those affected are disincentivised, e.g. to obtain further qualifications. If future discrimination is anticipated, investment in skills may no longer appear to be worthwhile (Britz 2008: 126f.). This is illustrated by Case 15, in which the risk is discussed that stereotype-enhancing image search results for occupational groups can impair the career aspirations of the groups concerned. In general, such amplification effects are likely to come mainly from systems that exhibit imbalances in the representation of groups of people or representation risks ("repressive harms") (e.g. Tolan 2018: 17).

Other types of algorithm-based accumulation and amplification effects result when algorithms based on already biased or discriminatory ratings of humans by humans are formed (and continuously adjusted)

[85] See Schauer (2018: 46).

[86] See Section 5.2.5.

and used for other functions, as shown in Case 28 and Case 32 for the formation of risk scores based on human ratings or assessment practices. This is also illustrated by Case 27, in which biased ratings of customers of various ethnic restaurants are used as the basis for algorithms of forecasting systems in state supervision.

5.2.3 Differentiations against socio-political ideas

Even if there would be efficiency gains from the use of differentiation with surrogate information, it may be desirable from a socio-political point of view that differentiation is abandoned in order to achieve objectives of equality and equal treatment, such as equality in court, equal access to infrastructure, equal educational and career opportunities and equal treatment with respect and dignity. Likewise, possible differentiation can be prevented if groups that were previously discriminated against or particularly prone to discrimination are to be protected and promoted (Schauer 2018: 50).

Specifically, an economically rational differentiation or disadvantage of certain characteristic carriers can be rejected if (a) there is a desire for compensation for past discrimination injustice or structural disadvantage of certain characteristic carriers, precisely in order to break through accumulation and amplification effects. Furthermore, differentiation should not be applied if (b) differentiation would make it more difficult for members of a structurally disadvantaged group to access goods, resources and positions that they would need precisely to overcome their disadvantaged group status (e.g. access to employment or credit). Furthermore, an economically rational differentiation can be rejected in order to (c) avoid expanding stereotyping, as if differentiation is associated with the attribution of a negative characteristic, this can lead to expanding stereotyping, in particular if a grouping is affected that is confronted with negative stereotypes anyway. Finally, (d) other rationales, such as health or social policy objectives, may speak against economically rational differentiation, e.g. in health or car insurance tariffs (Britz 2008: 127-130).

Societal risks could arise in the future if societal considerations about differentiation made possible or improved by algorithms and appearing economically sensible are unilaterally in favour of efficiency objectives and at the expense of equality or socio-political objectives. The danger can increase if cost reductions through automation in algorithmic and data-based differentiation make differentiation in many new application areas possible and displace other forms of governing and management.

5.2.4 Treatment as a mere means and psychological distancing

The use of algorithmic differentiation increases the risk that people are no longer considered as individuals or in recognition of their constitutionally granted human dignity and their unique individual subject quality,[87] and rather only **as a mere object or means**. It is true that according to the prohibition of instrumentalisation or the object formula, it is possible to treat other people also as means for one's own ends, but according to this moral principle it is forbidden to use them exclusively as mere means or to degrade them as mere means for one's own ends. Treating people as a mere means entails treating them in a way that they **do not agree** to. This can be the case, for example, with a false promise if the persons concerned do not know what is actually intended for them. Similarly, they may not consent to treatment

[87] Cf. Wiegerling (2016) and Hänold (2018: 130f.). See also Härtel (2019: 60) demands that the basic principle of the protection of human dignity should permeate all regulations on digital transformation. She refers to the categorical imperative of Immanuel Kant, "Act in such a way that you treat humanity, whether in your own person or in the person of any other, never merely as a means to an end, but always at the same time as an end." (Kant (1786/1977: 60, translation by Atwell J.E. (1986) The principle of humanity. In: Ends and Principles in Kant's Moral Thought. Nijhoff International Philosophy Series, vol 22. Springer, Dordrecht).

because they do not have reason or if it would seem irrational to do so. Respecting the dignity of other persons means then to treat them in a way that gives them the possibility to reasonably agree or disagree with what is done to them (Schaber 2012: 40-42).

Entities applying economic applications want to record and analyse the behaviour and conditions of persons concerned primarily in order to obtain or increase the monetary value from customer relations and not to find out the actual rationales for the behaviour (Yeung 2018: 30). Algorithmic data analyses of data mining, big data analysis or machine learning typically generate correlations and not causal relationships. As a result, the decision-makers lack the basis for sufficiently explaining the rationales for their decision to the persons concerned, e.g. if they are sorted out. This means, the persons concerned do not have the opportunity to consent or refuse the treatment they are undergoing. This risk became particularly clear in Case 22, where the use of the system did not sufficiently explain the refusal of the credit to the person concerned, which was included in the justification of the discrimination.

In addition, there is a risk that algorithm-based decision-making procedures could lead to a **psychological distancing** of the controllers from the decisions and the persons concerned. The risk posed by the interposition of computers as "moral buffers" and an apparent shift of moral responsibility for decisions to computers has so far been discussed mainly for autonomous weapon systems (e.g. Cummings 2004b; Brundage et al. 2018: 17), but it can be seen to have negative consequences for all semi- and fully automated decision-making procedures with negative consequences for the persons concerned.

Data protection law has developed the prohibition of automated individual decision-making to reduce such risks. This is intended to prevent disadvantageous differentiation decisions from being made solely by automated processing, and to ensure that "[...] no one may become the mere object of an assessment of personal data based solely on algorithms" (Scholz 2019: GDPR Art. 22 point 3, translated from German original; also, Martini 2018: GDPR Art. 22 point 1). The prohibition of automated decisions is discussed in more detail in Section 6.2.3 and its loopholes are pointed out, which make it questionable whether the actual protection objective is still being achieved.

5.2.5 Endangering the free development of personality and the right to self-expression

In addition to the equality rights of the Basic Law for the Federal Republic of Germany (see above), the phenomenon of statistical or algorithmic and data-based discrimination also affects the constitutionally granted rights of personality, in particular the **free development of personality** granted by Article 2 (1) of the Basic Law.[88] The problem results from the fact that evaluators form a certain image of the persons concerned by one or more characteristics. Persons concerned are confronted with externally-produced constructions of their identity and how other perceive a person, i.e. with external images (Britz 2008: pp. 179f.; Fröhlich & Spiecker aka Döhmann 2018).

"Statistical discrimination deprives the person concerned of the opportunity to present themselves to their counterpart and thus to influence the way they are perceived. Instead, the detection of statistically significant features almost automatically leads to conclusions about certain characteristics of a person. Statistical discrimination puts prefabricated personality profiles over the persons concerned, which they

[88] "Every person shall have the right to free development of his personality insofar as he does not violate the rights of others or offend against the constitutional order or the moral law." Article 2 para. 1 GG (translation by Deutscher Bundestag 2019).

are largely defencelessly at the mercy of (Britz 2008: 124f., translated from German original). If, in addition, misjudgements are made in statistical differentiation, the persons concerned are unjustifiably attributed a certain characteristic, "[...] without being able to defend themselves against this in the process of the development of this personality profile on their own (counter-) representation." (Britz 2008: 180, translated from the German original).

In this way, the **right to self-expression** of the persons concerned is taken away, which derives from the right to free development of personality.[89] According to Britz, self-representation is the means by which an individual can influence how other people see him or her (Britz 2008: 179-207). The right to self-expression serves the free development of personality in two ways:

(1) Through self-representation it can be achieved that others get a "favourable" image of an individual and thus make decisions favourable to the preservation of his or her scope for decision and action (**external development**). This is true especially considering that the image of an individual is always decisive when their scope for action depends on the willingness of others to cooperate. The image of the individual determines whether the willingness to cooperate is shown to the individual at all and whether this opens up room for manoeuvre for them, e.g. whether they are offered a contract or membership at all. Likewise, the individual can anticipatively limit their freedom of action if they do not know what external images of the individual have been created. If one has no influence on what data and information is included in the public image, the mere anticipation of public images can have a prohibitive effect (Britz 2008: 190f.).

(2) Furthermore, in the sense of **inner development**, the individual can, with "self-representation", ensure for him or herself "a sufficient share in the mutual process of constituting his or her identity in order to be able to understand his or herself as a voluntarily chosen personality" (Britz 2008: 195, translated from the German original). Although personality development in social contexts always takes place in interactive processes of external expectations and attributions (external images) on the one hand and one's own self-images, ideas and desires on the other, the core guarantee of the right of personality is to "provide mechanisms that integrate the individual into the processes of the constitution of personality in such a way that he or she can understand his or her personality as freely chosen [...]" (Britz 2008: 191, translated from the German original).

This is primarily a matter of protection against increased forms of heteronomy, which make the development of personality inhibited. This happens when intensive external images, i.e. those characterised by a special quality and density, are "imposed" on a person, thus depriving them of the options to form their own ideas about themselves. This can (a) be the case with comprehensive data-based personality profiles, where the evaluators' side is so comprehensively informed about the personality of the individual concerned that there is no possibility left for the own interpretation of roles in social contexts. This danger is also addressed in particular by the right to informational self-determination, in which self-representation is also seen as a condition for the development and preservation of personality (see also Section 6.4.2).[90] Also (b) the use of a single surrogate variable in the case of statistical discrimination does not perceive the persons concerned themselves, but a stereotypically constructed personality (Britz 2008: 193f.).

[89] For the formulation of the right to free development of personality and the right to self-expression see also Britz (2007).

[90] Cf. Britz (2008:193), Hoffmann-Riem (1998), Trute (2003), (1998), Britz (2010) and Albers (2017).

In order to protect the development of personality, prohibitions of discrimination are derived not only from the fundamental principles of equality, but above all from the right to self-expression and the underlying guarantee of the free development of personality. Prohibitions of discrimination are thus also to be understood as protection against inadmissible external images and attributions by others (Britz 2008: pp. 200ff., 204).

5.2.6 Creation of structural advantage

Algorithmic methods of analysis, usually based on artificial intelligence procedures, can increasingly identify personality traits, character traits and emotional states automatically (see Section 2.2.2). They could be used to identify and exploit a person's affinity with a product, service, resource or position, thereby increasing the structural superiority of those offering it. The effect can be enhanced if the providers also have access and control points of large volumes of personal data and personal profiles, which is an advantage in machine-learning methods. This is because the volume of data in this case also influences the quality of the generated models or algorithms. In addition, network effects,[91] especially in the case of online platforms or IT systems, can further increase the structural advantage of the providers, because network effects mean (sometimes prohibitively) high switching costs for the individual users. This reduces the number of choices and alternatives.

The risk of **structural advantage or disadvantage** is also increasingly taken into account for private circumstances, for example, with the Federal Constitutional Court's case law on structural inferiority, including the decision on guarantee agreements on structurally unequal negotiating power (BVerfGE 89, 214 (1993)) or the decision on stadium bans (BVerfGE 148, 267 (2018)). They may become relevant for areas of digitisation with large power asymmetries (Hoffmann-Riem 2017: 25; Schweighofer et al. in 2018: 7880; Härtel 2019).

Admittedly, the constitutionally protected 'private autonomy' still applies, according to which "[...] it is part of the freedom of every person to decide according to their own preferences with whom they want to conclude contracts and under what conditions" (BVerfG 2018: Guiding principles, translated from the German original). However, according to the decision, the scope of the principle of equality under Article 3 (1) GG can also extend to private sector areas for specific constellations. For example, private parties may not use their "[...] discretionary powers, which [...] might potentially arise from a monopoly or a position of structural advantage [...] to exclude specific persons from such events without factual reasons." (BVerfG 2018: point 41, translated by the Federal Constitutional Court). The events addressed are those events that mean "participation in social life" for the persons concerned (BVerfG 2018: point 41). From this, Schweighofer et al. (2018) derive the characteristics of structural advantage, i.e. the opening of the service to a wide range of traffic, the reliance on the service and the unilateral power of disposal of the offering company or person (ibid., p. 79).

In algorithmic decision-making systems, structural advantage is not presumed per se, but rather, "when the algorithmic judgement of persons is used to generate, reinforce or exploit the dependency of a person upon [...]" a product or service (Schweighofer et al. 2018: 79, translated from the German original). This may be the case in particular if the algorithmic procedure is used to identify those contractual partners who are dependent on the service (ibid.). According to Härtel, this could be the case with dynamic prices or price differentiation, credit scoring or online platforms with market power (Härtel 2019: 58). In the aforementioned constellations, this can result in a requirement for equal treatment in accordance with the principle of equality under Article 3 GG from the point of view of structural advantage with the main characteristic of

[91] For network effects, see p. 4.

being dependent on a product or service. However, the prerequisites and arrangements for giving sufficient consideration to structural advantage are still largely unclear (Schweighofer et al. 2018: 80).

6. Needs and Options for Action

Almost at the same time as discrimination risks of algorithms are beginning to be detected, the search for possible solutions has begun. Many recommendations or proposals for options for action or instruments emerge from the international arena and cannot be directly transferred to the local institutional framework. For example, it has been proposed that developers and practitioners carry out self-inspections for risks of discrimination, the implementation of human rights impact assessments (UN GA 2018; Yeung 2018: 65; Council of Europe 2019) or algorithmic impact assessments (Reisman et al. 2018). In the case of these instruments, however, the relationship with existing data protection law would have to be aligned, in particular with the provisions for data protection impact assessments (Art. 35 GDPR), which are necessary when processing special categories of personal data or data with particularly sensitive characteristics or characteristics prone to discrimination (according to Art. 9 GDPR), as well as the relationship with the requirements for automated individual decision-making (Art. 22 GDPR), which should also serve anti-discrimination law.

6.1 Transparency and proof of discrimination

A detailed discussion has developed on the transparency of algorithms.[92] Some authors emphasise that algorithms and computer systems are characterised by opacity and incomprehensibility or have so-called "black box" properties (e.g. Pasquale 2015; Castelvecchi 2016; Kitchin 2017). Almost as an antidote to this, the demand for transparency has arisen. The discussion revolves around questions about what, for which persons, for what purpose and in what form transparency should be created or should be avoided (Citron & Pasquale 2014; Mittelstadt et al. 2016; Ananny & Crawford 2018; de Laat 2017).

Demands for transparency can range from the explanation of the most important functions to the disclosure of the programme code or the possibility of inspection. In addition, transparency should fulfil very different functions, ranging from its function as an information instrument for consumer protection, to the creation of accountability in the sense of Algorithmic Accountability[93] to various stakeholders (Hacker & Petkova 2017).

In this context, different aspects have to be distinguished regarding the causes of the lack of transparency of algorithms and computer systems (according to Burrell 2016): (1) In many cases, the lack of transparency is due to the behaviour of developing and using entities who refuse to disclose algorithms, programme structures or even the programmed decision-making rules and criteria to external parties for reasons of protection of trade and business secrets, copyright protection, data protection (when computer systems contain personal data of third parties) or out of caution against targeted behavioural adjustments by the persons concerned ("gaming the system") (de Laat 2017).

(2) The different abilities and previous knowledge of the observer can also create an impression of opaqueness. A programme code that implements algorithms is not comprehensible without prerequisites

[92] Sometimes with different understandings of the term "transparency" from disclosure to explanation. See, for example, Castelluccia & Le Métayer (2019: 26-30).

[93] For the umbrella concept of "algorithmic accountability" with US-American origin, see Diakopoulos (2014), World Wide Web Foundation (2017), for classification in the European context Bush (2018) and EDPS (2018).

and in very few cases can be completely gauged and understood. Knowledge of the programming language is required as a basic prerequisite to be able to reproduce algorithms in a "raw" state. Therefore, disclosure of the programme code to persons concerned without prior knowledge would not be very productive, but disclosure for inspection by specialists can be useful (see below on testing software systems).

(3) A differentiated view emerges with regard to the lack of transparency caused by the technical properties of algorithms and software systems. For example, it is pointed out that the opaqueness can increase with the increasing complexity of algorithms and software systems (e.g. Yeung 2018: 15; Wischmeyer 2018: 47). This includes above all machine-learning algorithms, as well as adaptive or dynamic systems whose rules are constantly adapted through the continuous analysis of data streams (Desai & Kroll 2017). In contrast, the view is also held that algorithms are fundamentally comprehensible technical elements and that inscrutability arises primarily from the interest and power structures of the development processes of software systems (Kroll 2018). Since the algorithms programme the decision-making rules and these rules must be specifically formulated for this purpose, Kleinberg et al. (2019) even consider the rules to be in principle more comprehensible than rules enforced by humans. This is because with simulations, the outcomes of the decision-making rules can be assessed more clearly than in cases of human decision-making.

6.1.1 Technical options for transparency, traceability and non-discrimination

Numerous proposals are currently being made to create transparency and traceability. Of the very rapidly developing research and development trends, with partly unclear conceptual delimitations and wide overlaps, only selective excerpts can be reproduced here.[94]

For the **technical analysis** of algorithms, in particular of machine learning and thus the generation of explainability, different approaches can be distinguished: (1) Approaches with "open" systems (a "white box approach"), where it is possible to analyse the programme code. (2) From this approach the "closed" system approaches ("black box approach") have to be distincted, in which the behaviour of a system is analysed without taking note of the programme code. Explanations are constructed by observing both the input and the output. (3) In addition, a distinction can be made between the "constructive approach", which aims to implement explainability as early as during the development of the programme code (Castelluccia & Le Métayer 2019: 47-54). The research initiatives of "explainable AI" (e.g. Dosilovic, Brcic & Hlupic 2018) should also be seen in this context. For example, Ehsan et al. (2019) present an IT system that is supposed to be able to explain its steps in natural language ("automated rational generation").

With regard to technical analysis, Schweighofer emphasises the possibilities of testing software systems, which have existed for a long time as a component of "System and Software Engineering", and which also serve as standard procedures, in particular for quality assurance. They have similarities to auditing (see below). During testing, a software system receives a pre-defined input and aims to generate an output from it (Schweighofer et al. 2018: 5864).

Currently, some software systems for testing machine learning and automated decision-making systems are available mostly as open source programmes (Sanchez-Monedero & Dencik 2018: 12f.). Examples

[94] Overviews are provided, for example, by Guidotti et al. (2018), Schweighofer et al. (2018), Castelluccia & Le Métayer (2019) or Dosilovic, Brcic & Hlupic (2018), each with a different structure.

include the Themis system for testing fairness in software (Galhotra, Brun & Meliou 2017), the FairTest tool for investigating relationships between application outcomes and sensitive or protected characteristics of users (Tramèr et al. 2017) or the comprehensive "AI Fairness 360" system, which integrates a whole range of tools (Bellamy et al. 2018).

Kleinberg et al. (2019) point to the special ability to check decision-making rules for discrimination when they are implemented in algorithms. Since relevant decision-making rules are programmed in the algorithms, experiments and simulations are possible to investigate the effects of the decision-making rules on affected groups of people, e. g. by varying the data inputs or the decision-making rules themselves. However, the prerequisite for verification is access to the algorithms and the data sets. In contrast to algorithms, humans are the "ultimate black box" (ibid., p. 10). Therefore, the authors demand that in particular the decisions made by humans in the development and application process, such as the selection of data sets or influencing variables, be documented. This would make the effects of algorithms more comprehensible and discrimination could even be proven more easily in court proceedings, compared to (conventional) proof with statistics (Kleinberg et al. 2019).

Tests and analyses are associated with technical approaches to **prevention of discrimination**, which start with the design and use of the systems. In data mining, Romei and Ruggieri differentiate between (1) the naive approach of removing protected characteristics, although they also point out potential problems with this.[95] They also mention (2) the controlled disturbance or modification of the training data set (the "pre-processing" approach), (3) the modification of the learning algorithm for classification during training (the "in-processing" approach), (4) the modification of the model for classification after training (the "post-processing" approach) and (5) corrective interventions in the application of the prediction algorithm or model (Romei & Ruggieri 2014: 622-624).[96] However, Case 36 illustrates that some of these corrective options in ML procedures, which make discriminatory risk prognoses in juvenile justice, would not lead to satisfactory outcomes or would trigger new discrimination.

In addition, non-discriminatory machine-learning algorithms are also being developed that can be used for data sets containing legally protected characteristics. This often results in a conflict of objectives between fairness or avoidance of discrimination (measured by fairness criteria, see below) on the one hand and accuracy on the other. Various discrimination-reducing algorithms perform differently in terms of fairness or accuracy (overview and test in Friedler et al. 2019).

Metrics are also being developed to measure fairness (**fairness criteria/measures or metrics**). According to Berk et al., the following fairness criteria can be distinguished, (1) overall accuracy equality, (2) statistical parity, (3) conditional procedure accuracy equality, (4) conditional use accuracy equality, and (5) treatment equality (Berk 2018; Schweighofer et al. 2018: pp. 39f.).[97] However, Chouldechova points out, using the example of the algorithms for computing recidivism rates in the judicial system (Case 35), that the fairness criteria cannot be fulfilled simultaneously (Chouldechova 2017).

Which fairness criteria should be used in which situations and for which differentiation purposes cannot be decided in computer science or by applying entities, but requires political treatment and decisions (Berk et al. 2018; Castelluccia & Le Métayer 2019: 55). Furthermore, when considering fairness criteria, the application of the system itself is not questioned, but is instead a given assumption. Societal risks of discrimination that arise from the application of differentiation systems themselves are not solved in this way.

[95] See also Section 5.1.2.

[96] See also Friedler et al. (2019) or Castelluccia & Le Métayer (2019: 46-47).

[97] A further overview is given in Verma & Rubin (2018).

6.1.2 Improving the evidence of discrimination

6.1.2.1 Empirical investigations and evidence

Empirical investigations and statistics have long played an important role in the discovery and proof of discrimination (e.g. Supik 2017). They are also available for decisions of algorithmic differentiations to capture and evaluate the consequences and outcomes. With the increasing digitisation of administrative and private interactions, there is potentially much more data available for statistical analysis to detect and prove discrimination.

Romei and Ruggieri (2014) provide a comprehensive bibliographical overview of the most common empirical analyses of discrimination with references to exemplary studies. In doing so, they differentiate the studies according to the possibilities that the researchers have to influence the influencing variables (or independent variables) in the statistical analyses. (1) In observational studies, researchers have no control over the influencing variables. They collect data from observations by means of surveys or interviews on specific situations, conditions, structures of economic or life areas, such as labour or credit markets, or the treatment of groups of people who are prone to discrimination. In (2) quasi-experimental studies, researchers only have control over some influencing variables. This type of study includes (2a) auditing studies in which individuals are sent as test pairs in decision-making situations and discrimination can be derived from comparisons of treatments. (2b) In situationtestings, the subjects have contact with the decision-maker and can make (hidden) records of possible unequal treatment. (2c) Correspondence-testings attempt to identify discriminatory behaviour in the responses, mainly through written questions, e.g. fake applications and CVs. This type of study in particular is used in the online sector, e.g. to investigate online recruitment services. (3) In experimental studies, researchers have control over all influencing variables. A distinction can be made between laboratory experiments and natural experiments (Romei & Ruggieri 2014: 591-621). They also point out that data mining is also suitable for detecting discrimination (Romei & Ruggieri 2014: 621-624) e.g. to detect gender discrimination in research proposals (Romei, Ruggieri & Turini 2013).

6.1.2.2 Algorithm audits

Algorithm audits are methods and tools designed to enable researchers and protective institutions to investigate systems with algorithmic and data-based differentiations and to help understand the effects of algorithms, including discrimination, on all types of persons concerned (Sandvig et al. 2014; Hannák et al. 2017: 2; Schweighofer et al. in 2018: 64-73). In part, they correspond to the research methods of testing software systems or the “classical” empirical studies of discrimination (see above). According to Sandvig’s classification (2014), which refers to online platforms, there are (1) code audits, which correspond to the testing of the programme code with full inspection,[98] (2) Another form collects data on the interactions of platform users and does not require insight into the code of the system (“non-invasive user audits”). (3) Furthermore, data is collected with repeated requests to a platform (“scraping audits”). (4) Fictitious test persons generated by computer programmes can also use the service under investigation to collect data by carrying out repeated uses (“sock puppet audit”). (5) Finally, test persons hired via crowdsourcing services can use the service to generate relevant data (“crowdsourced audit” or “collaborative audit”).

[98] For comparison see Schweighofer et al. (2018: 70).

In some of the cases described in Chapter 4, algorithm audits were used to detect unequal treatment on websites, online platforms and online marketplaces, e.g. for online labour markets (Case 2), for price differentiation in retail (Case 12) (see also Mikians et al. 2012, 2013) or for online facial recognition services (Case 46).

The procedure in the above-mentioned cases for the detection of unequal treatment and discrimination with the help of algorithm audits has the character of scientific investigations that require expertise and, above all, resources for the empirical investigation of interactions in the online sector (e.g. use of crawlers, handling of fictitious accounts) and, above all, statistical analysis. They mostly use information that can be obtained on the internet. For this purpose, the necessary data is partly collected with web crawlers, partly by extensive automated queries to search engines or online platforms with the help of software or "crowdworkers" (e.g. via Amazon Mechanical Turk).[99] It is also becoming apparent that a growing number of researchers are making the identification and investigation of discrimination – in particular in the online sector – the subject of their research and are further developing the methodological toolkit for this purpose.

Case 16, which highlights the "Sunlight" system and Case 17, which highlights the "AdFisher" system, also provide tools for automated data collection and the analysis of online transactions (partly with machine-learning methods), which can be used to investigate personalisation or potentially discriminatory differentiations. Such tools can be used to perform algorithm audits (in particular type 2 and type 3).

Conclusions

- Research results and cases show that proof of unequal treatment and discrimination based on algorithmic and data-based differentiations can also be achieved without the access and direct inspection of the algorithms. This is made possible by observing, recording and evaluating the outcomes and consequences of the differentiation applications, often by taking on the role of test users, and can also be done via automated uses and queries.

- Similarly, cases[100] show that, in particular in automated decision support systems, it may be important to look at the overall outcome, i.e., which actual consequences are triggered by human decisions in light of computer recommendations. This is also relevant for the many cases where AI systems are less able to detect certain groups of persons with protected characteristics (in Section 4.12). In those cases, discrimination would only become apparent if certain practices and decisions based on poorer detection were considered (and these would lead, for example, to disproportionately more police or border controls, disproportionately fewer staff being recruited or inadequate consideration in marketing strategies). However, the form of evidence of identifying outcomes also requires expertise, particularly in statistics and programming, as well as financial and human resources to carry out the investigations. Therefore, this form of providing evidence is unlikely to be suitable for affected individuals without specialist knowledge. For equality bodies, particularly in cooperation with research institutes, the forms of verification can be suitable instruments and are also used by them in the beginning.[101]

[99] See, for example, Case 2, Case 8, Case 12, or Case 38.

[100] See cases on employment services in Poland (Case 29) and Austria (Case 30).

[101] See, for example, Case 47.

- Another requirement is that the relevant communication and transactions relating to the differentiations must also be statistically recordable, as seems to be possible for many online applications and online platforms that have public offerings. However, such option of proof can no longer be acquired wherever these transactions and communications are not publicly recordable, such as in the case of closed administrative procedures or exclusively individualised commercial offers.

- Furthermore, these methods cannot reveal the exact causes of discrimination when several causes may lie within a complex system or in the complex interplay between data sets, algorithms and human decision-makers.[102] This can be done by investigating the practices of data generation used (as in Case 34) and the possibilities of testing software systems and simulating variations of input data or components of decision-making rules, and also by investigating how human decision-makers deal with computer recommendations.

6.1.3 Legal situation

6.1.3.1 Data protection information obligations and rights of access

Data protection law includes various information obligations for the controllers or operators of data processing vis-à-vis the persons concerned or data subjects respectively (Articles 12, 13 and 14 GDPR) as well as information rights, which the data subjects can assert vis-à-vis the controllers (Article 15 GDPR). The purpose of the information requirements is to ensure that data subjects are aware of the data processing so that they can effectively exercise their rights. They should also form the basis of the consent of the data subjects, which (among other reasons) determines the lawfulness of the data processing (Art. 6 para. 1 GDPR). This is why it is also referred to as "**informed consent**". With the rights of access, data subjects have the right to obtain information on the purpose and scope of the data processing in order to ensure that they can check whether the data has been processed lawfully (Busch 2018: 37-41). Extended information requirements apply to the existence of automated decisions (see Section 6.2.3).

In practice, the information obligations and the concept of informed consent are usually specified in the ”privacy policies" (also called data protection statements), although their design is criticised. Critique is directed at the ambiguities of the terms used, the use of legal language that is not easy to understand, the inadequacies of the information provided, cognitive barriers to understanding it and the effort and time constraints that prevent reading and understanding the privacy policies (Milne & Culnan 2004; Solove 2013; Cate & Mayer-Schönberger 2013; Reidenberg et al. 2015; Reidenberg et al. 2016; McDonald & Cranor 2008; Van Alsenoy, Kosta & Dumortier 2014; Martin 2013; Moll et al. 2018; Kamp & Rost 2013; Orwat & Schankin 2018; Kettner, Thorun & Kleinhans 2018; Hänold 2019).

While the privacy policies must specify the purpose of the data processing,[103] it is doubtful that this information can be used to assess the consequences of the differentiation decisions based on the data processing in terms of unequal treatment by the data subjects. Since the purpose is only to provide information on the existence of an intended purpose of processing and not on its consequences, it can be

[102] This is shown, for example, in Case 17.
[103] According to Art. 13 para. 1 lit. c GDPR.

assumed that this legal instrument does not provide the necessary evidence for an anti-discrimination action.

6.1.3.2 Burden of proof and circumstantial evidence under the General Equal Treatment Act (AGG)

Section 22 AGG provides for the burden of proof to be eased for the persons concerned by placing the burden of proof on the party accused of discrimination to prove that there has been no violation of the provisions on protection against discrimination. According to Ebert, the facilitation of evidence is linked to three conditions: (1) The person claiming to have been discriminated against must prove that they have been treated differently from other persons and (2) the person claiming must prove that they differ with regard to one of the protected characteristics (pursuant to Section 1 AGG). (3) Furthermore, the person must produce evidence which shows with a substantial degree of probability that the characteristic referred to in Section 1 was the cause of the discrimination (Ebert 2019: Section 22 AGG point 1 and 2). With regard to risks of algorithmic discrimination, these requirements are **problematic**:

From the perspective of the persons concerned, the poor traceability of the effects of algorithms makes it difficult or even impossible for the persons concerned to demonstrate that they have suffered discrimination due to algorithms. In the case of personalisation and the targeted, exclusive offering of information, services or products – in particular in online offers and on online platforms – a single, potentially affected person may have difficulty in proving that he or she is treated less favourably than comparable persons. This is even more difficult to prove if the offers change dynamically. Still more serious is that it is impossible for an individual without in-depth expertise to prove unequal treatment with regard to the legally protected characteristic, precisely through the unintentionally or intentionally concealed use of surrogate variables or proxies that correlate with the protected characteristic (Section 5.1). Thus, the individual can hardly provide the necessary evidence. This is indicated, for example, by the extensive empirical studies and algorithm audits in the cases, which are necessary for discrimination to be proven at all (Chapter 4).[104] In addition to these problems, (anticipated) dependencies (such as in the selection of applicants) or the high risk of legal costs may be obstacles to taking action against any discrimination perceived by the persons concerned.

[104] See similarly Hänold (2019) with regard to algorithms, profiling and scoring in the insurance industry. See also Fröhlich & Spiecker aka Döhmann (2018).

Conclusion

One possible solution would be **collective redress** with the class action suit, which has long been demanded for the anti-discrimination action (Berghahn et al. 2016: pp. 141ff., 159-162; Ponti & Tuchtfeld 2018; Straker & Niehoff 2018). However, the need remains for someone to perceive potential discrimination as possible damage in the first place who can then take the initiative to bring a collective action. For the barely perceivable risks of algorithmic discrimination, (supplementary) regular surveys carried out by equality, research or similar institutions or specialised authorities (in particular also in collaborations) would be more suitable in the absence of any concrete damage. Moreover, legally safeguarded information rights for equality bodies can facilitate such a procedure for the provision of evidence, which is not yet available.

6.1.3.3 Documentation

From the reversal of the burden of proof, Dzida and Groh conclude for the field of labour that in the case of a dispute in which a court suspects discrimination, even those who implement algorithms have difficulty in proving that there is no discrimination or that the use of the system is proportionate to the differentiation task, such as personnel selection (Dzida & Groh 2018). For example, while the proportionality test may be able to demonstrate the suitability of the system for the differentiation task, e.g. by means of scientific studies, the differentiation objective in question may not be achieved by "other equally appropriate but less intervening means and the procedures may not unduly prejudice the legitimate interests of disadvantaged persons" (ibid., pp. 1920f., translated from the German original).

For the duty to prove that there is no discrimination, it may be necessary for applying entities to take precautions to ensure the comprehensibility of algorithms or artificial intelligence, so that in cases of litigation it can be proven how a decision and its outcomes or consequences for persons concerned came about, or according to which decision-making criteria and weightings differentiation decisions were made. According to Yeung, these requirements may also result from the principle of procedural fairness, according to which persons in court proceedings have a right to know the reasons for decisions that adversely and significantly affect them (Yeung 2017: 23).[105]

For example, protocol obligations regarding the programme procedures or characteristics used in the differentiation decisions are proposed to serve as evidence in cases of dispute (Martini 2017; Ernst 2017: 1032; Brauneis & Goodman 2018). Furthermore, the provisions on the records of processing activities (according to Article 30 GDPR) as well as on the codes of conduct and certification (Articles 40-43 GDPR) should be further developed to take account of the specific features of algorithmic decision-making procedures.

[105] See also the legal proceedings in the US; see AI Now Institute (2018).

6.2 More detailed regulation of algorithmic decision-making rules

In view of the problems of information duties and rights of access, which do not allow for sufficient self-protection against disadvantages and discrimination, and given how difficult, if not impossible, it is for persons concerned to prove discrimination (without expertise), consideration should be given to stricter regulation of differentiation decisions.[106] Regulations of differentiation decisions based on the analysis of personal data are not new. They can be found in the law relevant to anti-discrimination and informational self-determination and its institutional implementation in relevant authorities and institutions. For example, the AGG can also be understood as a form of regulation of decision-making rules, in which the use of certain decision-making characteristics is excluded.

6.2.1 Prohibitions of discrimination and legally protected characteristics

The General Equal Treatment Act (Allgemeines Gleichbehandlungsgesetz, AGG) was intended to prevent, in principle, decisions about persons on the basis of certain widespread generalisations and differentiations according to characteristics that are particularly prone to generalisation and that can lead to discrimination (Britz 2008: 4). Thus, in principle, the direct use of the protected characteristics is inadmissible under Sections 7 para. 1 and 19 para. 1 AGG, even in differentiation decisions using algorithms or computer systems.

Similarly, indirect discrimination, which is common in algorithmic decisions, i.e. where an apparently neutral feature is used but persons are disadvantaged with regard to the protected characteristics, is prohibited pursuant to Section 3 AGG unless the use of the feature is justified by a legitimate aim and the means of achieving that objective are appropriate and necessary (proportionality test) (according to Ernst 2017: 1032).

The existing catalogues of protected characteristics of the AGG[107] should be reviewed in order to determine whether they cover the characteristics that can be identified by (new) algorithmic methods and whether new grounds of discrimination are created. This is because systems with machine learning and other forms of AI make it possible to identify and differentiate according to characteristics that are not yet included in the catalogues of anti-discrimination law (cf. also Zuiderveen Borgesius 2018: 20). The systems can include analysis options for detecting sentiments, naivety and suggestibility, identifying cognitive weaknesses or psychological and emotional states (such as depression), the respective social status or character traits (see Section 2.2.2). These can be, for example, the characteristics “biometric features”, “political opinion” or “state of health”, which are regulated in the GDPR as a special category of personal data (see below), but not in the AGG.

At the moment, the societal consequences can only be conjectured, since knowledge about the actual use of the systems and the changes triggered is still largely lacking. To this end, research is needed to examine potential applications of such systems, mechanisms of action and potential risks for equal treatment and free personal development. However, there is already a risk that such identified personality traits could be

[106] See also Raabe & Wagner (2019 in progress)

[107] For discussion of the catalogues of characteristics, see e.g. Däubler (2018: AGG Section 1 point 6-10).

exploited to differentiate persons according to their dependence on a good, a resource or a position (e.g. a job), thus the increase in structural advantage[108] of the providers of such goods, positions, etc. can be difficult to dismiss.

Anti-discrimination law could be used with the extension of the catalogue of protected characteristics to protect against abuse of structural advantage.

Furthermore, it must be examined in this context whether the provisions of Article 9 GDPR are sufficient for protection against algorithmic discrimination. This provision prohibits the **processing of sensitive personal data**, which reveals "[...] racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation [...]" (Article 9 para. 1 GDPR) to be prohibited, unless one of the numerous exemptions allows it. For example, the processing of sensitive data is permitted if the data subject gives his or her explicit consent (Article 9 para. 2 lit. a GDPR). Here, too, the data subject would in principle have to estimate the partly long-term individual consequences, also in terms of unequal treatment, at the time of giving consent, which can be a major challenge. It is doubtful whether a realistic chance of self-protection of the persons concerned from algorithmic discrimination will arise from this.

Summary and conclusion

The catalogues of protected characteristics laid down in the GG and AGG are to be examined to see whether new methods of analysis, in particular with artificial intelligence algorithms, for the automated identification of personality characteristics require their extension. This also makes characteristics identifiable and accessible for differentiation, which can be used to identify and use the dependency on a good, resource or position in order to establish or increase structural advantage. The still largely unknown connections between technically feasible and potentially endangered protection objectives should be researched and their legitimacy should be assessed from a societal perspective.

6.2.2 Exemptions justified on objective grounds or by the use of recognised methods

Pursuant to Section 20 para. 1 AGG, there is no violation of the prohibition of discrimination if there is an objective ground (or objectively justified or factual reason). The vague term "objective ground" is clarified in its dimension by the examples given in the following sentences, but not exhaustively (Schrader & Schubert 2018: AGG Section 3 points 68ff.). According to this, an objective ground exists, for example, if (a) it is a matter of avoiding risks, preventing damage or the like, e.g. if certain groups of persons are excluded from using certain equipment or vehicles due to safety obligations. However, it is debatable to what extent this also includes an economic threat, e.g. in the form of loss of sales.[109] (b) Similarly, there may be an objective ground if certain persons are excluded in order to protect the private life or personal security of other persons (e.g. if the opening hours of swimming pools or saunas are separated according to gender).

[108] See Section 5.2.6.

[109] Franke & Schlichtmann (2018: Section 20 AGG point 17) interpret this provision as allowing only a few differentiations by this.

(c) Another example of an objective reason is special advantages or benefits and where there is no interest in enforcing equal treatment, such as discounts granted for social reasons (e.g. for students) or favourable sales promotions that only affect certain groups of people (e.g. lower prices for men in dance courses where there is a surplus of women, or vice versa) (Franke & Schlichtmann 2018: AGG Section 20 points 12-21). If none of these objective grounds exist, the only remaining option is to clarify the matter on a case-by-case basis, whereby a consideration must be made according to the principles of proportionality (Schrader & Schubert 2018: AGG Section 3 points 68ff.). The legal situation, however, provides for uncertainties of interpretation in advance, for example, when the design of decision-making systems is concerned with which features can or cannot be used.

Section 20 para. 2 Sentence 2 AGG regulates unequal treatment in insurance contracts with regard to the characteristics of religion, disability, age and sexual identity (differentiation according to all other protected characteristics is prohibited by Section 19 AGG). It is permissible if it is carried out according to “recognised principles of risk-adequate calculations” (Section 20 para. 2 Sentence 4 AGG, translated by the Federal Anti-Discrimination Agency 2009), in particular if it is based on an “assessment of risk based on relevant and accurate actuarial and statistical data.” (Section 20 para. 2 Sentence 2 AGG, translated by the Federal Anti-Discrimination Agency 2009).

Schiek (2000: Section 20 AGG point 8) believes insurance discrimination, with the continuing formation of stereotypes and prejudices through the collection of statistics linked to legally protected characteristics, to be an objectively unjustified breach of the right to equal treatment. She considers an extended application of this form of legitimation to other areas of life, such as banking services, to be inadmissible. Berghahn et al. also demand the restriction of this form of unequal treatment (Berghahn et al. 2016: 122ff.). Nevertheless, in the practice of scoring, e.g. in online trading, legally protected characteristics with reference to economic interests are used as objective grounds (e.g. Moos & Rothkegel 2016).

Conclusions

With the increasing spread of algorithmic computations and the implementation of differentiation, including scoring, it should be examined whether the admissibility of the calculation methods, the justifications, the characteristics, the fields of application and differentiation purposes should not be regulated more clearly and made mandatory. This should also apply to the recognition procedures for the “recognised principles of risk-adjusted calculation” or “scientifically recognised mathematical-statistical procedure”. In view of the rapidly increasing number of algorithms, machine-learning procedures, concretisation of “sufficient” forecast accuracy, fairness criteria and quality or error measures, it is likely that assessments of their respective suitability and adequacy may also diverge, so that it is no longer possible to speak of a generally recognised method. Clearer and generally mandatory clarifications in this regard could create stable expectations among developing and using entities and, where appropriate, persons concerned.

6.2.3 Prohibition of automated decisions

One of the most important regulations to protect against risks of algorithmic discrimination is the prohibition of automated decisions in data protection law. The purpose of the provision is already found in the Data Protection Directive (Directive 95/46/EC) and the old Federal Data Protection Act (Bundesdatenschutzgesetz, BDSG old version), where it serves to protect human individuality as an element of the right to free development of personality and autonomous shaping of one's own life (Ernst 2017: 1030; Martini 2018: GDPR Art. 22 point 8; Hoeren & Niehoff 2018: 53). According to Article 22 para. 1 GDRP, a data subject has the right not to be subject to a decision based solely on automated processing, which produces legal effects for him or her or significantly affects him or her in a similar way. It is not clear from the GDPR which types of automated decisions are actually covered. This can only be derived indirectly from the wording of the legal norm:

(1) On the one hand, this are automated decisions with decisions based exclusively on automated data processing. This is interpreted to mean that this is the case if a natural person has not made a substantive assessment and decision based on it, or if the natural person involved has no final decision-making authority (Ernst 2017: pp. 1029f., 1031; Bush 2018: 31). Martini stresses that it is decisive whether a person has an influence on the decision, i.e. its content. In doing so, people can prepare decisions manually, because whether the prohibition applies does not depend on the preparation but on the decision itself. If a person has a substantive power of decision, they actually exercise this power of decision, and if there is regular intervention, i.e. there is no random control and moreover no single case of human intervention, it is no longer possible to speak of an exclusively automated decision and the prohibition does not apply (Martini 2018: GDPR Art. 22 points 17-19).[110]

(2) On the other hand, all those types of automated decisions are covered, which "produce legal effects or significantly affect the person in a similar manner" (Scholz 2019: GDPR Art. 22 points 31-37, translated from the German original). The nature of the effects is therefore decisive.[111] The "legal effect" is to be assumed if the legal position of the data subject changes, such as when a contract is terminated, and a "significant impairment" is always given if the economic and personal development of the data subject is significantly disturbed, such as when a favourable interest rate fails to materialise (Busch 2018: 33).

If there are automated decisions that are permitted under the above-mentioned articles of the GDPR, then further regulations must be complied with (Weichert 2018: 131; Hoeren & Niehoff 2018: 54f.): Article 14 para. 2 lit. g GDPR regulates the **right to be informed**, meaning that where automated decision-making exists, the controller must provide the data subject with "[...] meaningful information about the logic

[110] See also Weichert (2018: 128-135, in particular pp. 133f.), for whom banned automated decisions are considered to be also situations in which the natural person only examines documents prior to the decision or is purely formally involved in the decision-making process. Similarly Hoeren & Niehoff (2018: 53) as well as Scholz claim, "An exclusively automated decision is to be assumed not only if no review by a human being is intended from the outset and no such review takes place, but also if the human being – without making any considerations of their own – merely confirms or accepts the automated prescription" (2019: GDPR Art. 22 point 26, translated from the German original).

[111] Weichert (2018) provides another interpretation with reference to Buchner (2018: Art. 22 point 18) and Reichwald & Pfisterer (2016: pp. 211f.), which sees the prohibition as being determined in particular by the degree of lack of transparency and lack of influence for data subjects. Thus, only automated decision-making systems in which the decision-making process is no longer manageable for data subjects and which lack controllability and revisability for data subjects are covered by the probition. This may be the case if the algorithms are not fully documented or in the case of automated decisions by means of learning algorithms or artificial intelligence (Weichert 2018: 130). However, as shown in the examples of discrimination, even comparatively "simple" algorithms, without the involvement of people and on a contractual basis, can have legally disadvantageous or significantly detrimental effects that would not be covered by the prohibition according to this interpretation.

involved, as well as the significance and the envisaged consequences of such processing for the data subject." (see below for further details). The **rights of access** of the data subject are regulated in Article 15 para. 1 lit. h GDPR and provide, with the same wording, for information on the logic and consequences involved.

In addition, for automated decisions, a **data protection impact assessment**[112] is required pursuant to Article 35 para. 3 lit. a GDPR if a "systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person". This makes it clear that the legislator ascribes a high risk to automated decisions. The controller must use the data protection impact assessment to assess these risks in advance, as well as whether the processing operations are necessary and proportionate, and the remedial measures planned to deal with the risks (Article 35 para. 7 lit. b to d GDPR). This means, for example, that the controller also refrains (or must refrain) from using machine learning or artificial intelligence procedures if the risk is assessed as disproportionate to the purpose of processing and less risky alternatives to the decision-making procedure are available or the procedure in question is not absolutely necessary for the (differentiation) purpose. If high risks are identified, a report must be submitted to the supervisory authority (pursuant to Article 36 GDPR). In such cases, the supervisory authority may prohibit the processing (pursuant to Article 58 para. 3 lit. f GDPR).

6.2.3.1 Exemptions

Exemptions to the prohibition of automated individual decisions are regulated in Paragraph 2. They exist when the automated decision is necessary to conclude or perform a contract, is permitted by legislation of the European Union or the member states, or by **explicit consent**[113] of the data subject. However, according to Article 22 para. 4 GDPR, these exemptions do not apply if decisions on the processing of **special categories of personal data** of Article 9 GDPR ("sensitive data"), which is explicitly intended to serve anti-discrimination (Buchner 2018: GDPR Art. 22 point 44). However, this prohibition is again restricted if the data subject has given his or her explicit consent (Article 9 para. 2 lit. a GDPR), or if the processing is necessary for reasons of substantial public interest (Article 9 para. 2 lit. g GDPR).

If one of these two exemptions applies to the processing of particularly sensitive data, the admissibility of the automated decision also depends on whether the exemptions of Article 22 para. 2 apply, i.e. whether it is necessary for the conclusion or performance of a contract or whether the data subject has given his or her explicit consent (Buchner 2018: GDPR Art. 22 point 45f.; Bush 2018: 35). In this context, von Ernst points out a potential dilemma here, highlighting that although consent to data processing is possible pursuant to Article 22 para. 2 lit. c GDPR, this conflicts with the AGG, which excludes discrimination even if consent to unequal treatment has been given (Ernst 2017: 1033; Schrader & Schubert 2018: Section 3 AGG point 47).

[112] More detailed provisions on data protection impact assessment are provided by the Article 29 Data Protection Working Party, WP29 (2017b).

[113] The term "explicit consent" is not explained in the GDPR. The Article 29 Data Protection Working Party's Directive on Consent provides guidance on this; see WP29 (2017a: 18f.). Scholz elaborates on this: "Even if, formally speaking, this does not involve consent to individual data processing steps, but rather the use of a data processing procedure, this consent will also have to be measured against the requirements of Art. 4 No. 11 and Art. 7. [...] From the perspective of the data subjects, the need for protection is comparable. Consent is therefore only effective if it is given unambiguously, voluntarily, specifically and in an informed way [...]. The latter presupposes that the data subject must receive all information necessary to correctly assess the reason, purpose and consequences of the processing before giving consent" Scholz (2019: GDPR Art. 22 point 52, translated from the German original).

6.2.3.2 Appropriate measures

Also, in the exceptional cases of Article 22 para. 2 lit. a and c GDPR, i.e. in the presence of "contract" and "consent", i.e. the situations in which automated decision-making is permitted, the automated decision should be made with appropriate measures taken by the controller to safeguard the rights, freedoms and legitimate interests of the data subjects. These include at least the right of the data subjects to obtain direct intervention by a person of the controlling authority, as well as the right to express one's point of view and to contest the decision (Article 22 para. 3 GDPR).[114] The objective of these regulations is not only to protect against discriminatory automated decisions, but also to ensure transparency and fairness in the decision-making process itself (Scholz 2019: GDPR Art. 22 points 3, 56).

As regards the right of the data subject to "**human intervention**" or the resulting possibility to object to the automated decision at any time ("opt out"), is, however, still unclear. Martini and Nink and also Busch, for example, interpret this so narrowly that the intervention of a person can only be demanded if there are justified reasons and in individual cases (Martini & Nink 2017; Busch 2018: 36). However, the paragraphs could also be interpreted differently. Direct intervention and contesting are the preliminary stage for the subsequent exercise of the rights to present one's point of view and to obtain a review of the decision. However, Recital 71 GDPR states that these rights must be granted "in any event".

With the option to **express his or her point of view**, the data subject should be given the opportunity to present the specificities of the individual case from his or her point of view, which would not be taken into account in an automated decision. Martini and Nink explain that the controller is obliged to actually take the aspects presented into account so that the authority does not degenerate into a "meaningless phrase" (Martini & Nink 2017: 4). The controller is then required to review the decision and deal with the content of the aspects brought forward (ibid.). However, the rights of direct intervention and the right to express one's point of view are weakened in that data subjects must be aware of the situation and of the potential harm that automated decision-making may cause.

6.2.3.3 Information requirements

The right to put forward one's point of view implies the need to provide the data subject of the decision-making procedure with information, either in advance or during the decision-making process, in sufficient detail to enable the data subject to express a meaningful opinion. This is ensured by the information requirements of Article 13 para. 2 lit. f and Article 14 para. 2 lit. g GDPR. "According to this provision, the controller must both inform at an early stage about the existence of automated decision-making and provide meaningful information on the logic involved and the scope and intended impact of such processing on the data subject [...]. According to Article 12 para. 1, this information must also be provided in a precise, transparent, comprehensible and easily accessible form in clear and simple language [...]" (Scholz 2019: GDPR Art. 22 point 58, translated from the German original).

For the "**logic involved**", Scholz specifies: "The term 'logic' shall be understood to mean information on the organisation, structure and operation of automated data processing [...]. The information must therefore include the basic functional principles of the application programmes and the basic decision-making

[114] Also described in Recital 71 GDPR: "In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision."

criteria. However, the technical details of the (analysis) software used or the source code do not have to be communicated. In this respect, processors can generally rely on the protection of their business and trade secrets. [...] However, the data subject must be able to understand how certain ratings and classifications are derived and what meaning and weighting these values have for the automated decision (Scholz 2019: GDPR Art. 22 point 54, translated from the German original).

Bäcker makes similar comments, claiming: "This information refers to the methods and criteria used in data processing, such as the functioning of the algorithm used to calculate a score" (Bäcker 2018: Art. 13 points 53-54, translated from the German original). Hoeren and Niehoff (2018: 56f.) also argue for the disclosure of the algorithm in the form of the presentation of "[...] rules of action and programme sequences with the corresponding weightings [...]" (ibid., p. 57, translated from the German original). This would also not jeopardise the business secrets[115] of the controller, since the source code would also have to be available for such a threat. With regard to the special forms of automated decision-making with artificial intelligence systems, they point out that the actual decision-making criteria are often not comprehensible in the case of AI procedures. However, additional technical procedures can make it possible to have the criteria that are particularly relevant to the decision revealed, which can then be presented to data subjects. However, these are approximate values, so that data subjects would also have to be informed about the uncertainties of the procedures (Hoeren & Niehoff 2018: 57-60). Further information and claims, which would mean a detailed explanation in each individual case of a decision, cannot, according to Wischmeyer (2018: pp. 51f.), be derived from the provisions of the GDPR.[116]

Nevertheless, the article is valuable in terms of anti-discrimination in that it clarifies that, in addition to the obligation to provide information on the functioning or 'logic', it also clarifies "the significance and the envisaged consequences of such processing" (Article 13 para. 2 lit. f GDPR and Article 15. para. 1 lit. h GDPR) or information must be provided. According to this provision, a controller must describe "[...] what is to be decided on the basis of the data processing, what choices are available and what processing outcomes lead or may lead to which decision" (Bäcker 2018: Art. 13 point 55, translated from the German original).

[115] In this context, the authors (ibid., pp. 56f.) refer to the so-called SCHUFA ruling of the Federal Court of Justice (BGH) rejecting an action for disclosure of the score formula used by the company SCHUFA to calculate creditworthiness with the prevailing importance of business secrecy over transparency requirements (BGH ruling of 28 January 2014, BGH (2014)).

[116] For a discussion on the so-called "right to explanation", see Goodman & Flaxman (2017), Wachter, Mittelstadt & Floridi (2017), Wischmeyer (2018), Edwards & Veale (2017).

Conclusion

In view of such a duty to provide information where automated decisions exist, it should in future be required to provide information not only on the decision-making rules but also on the scope and effects of the differentiation decision, including the risks of discrimination. Since the information must be provided ex ante, i.e. before the decision is made, data subjects potentially affected by the risks would have the option to refuse consent. As a result, the controller would also have to address the risks of discrimination in the first place in order to be able to provide information about them.

It should also be investigated whether such an obligation to provide information about the scope and effects of automated decisions could effectively supplement insufficient rights to information under the AGG, such as in situations involving job applications in the field of human resources.

6.2.3.4 Criticism and need for further development

Overall, Scholz criticises the provisions of Article 22 GDPR, noting that “This provision neither regulates the highly relevant question of whether and under what conditions a personal profile may be created and used, nor does it generally provide for the non-discriminatory and transparent use of algorithms” (Scholz 2019: GDPR Art. 22 points 8-11, translated from the German original). Martini is equally critical of the fact that the legislator has placed greater emphasis on the exploitation of potentials of value creation and economic innovation than on the protection of privacy. Automated decisions that “make the individual the mere object of an algorithmic analysis carried out without human intervention” are only prohibited to an extremely limited extent (Martini 2018: GDPR Art. 22 point 8, translated from the German original).

Furthermore, it should be considered that if algorithmic procedures are used by the controller and it is no longer possible to comprehend the path of the decision, this should be interpreted as an exclusively automated decision, even if a human being were involved in the decision-making process. The prohibition would apply in these cases. This would mean considering whether the criteria set out in the GDPR, according to which a decision is exclusively automated, should also be supplemented and substantiated to include the ability of the decision-makers to comprehend the decision recommendations of the computer system and to be able to explain them to the data subjects.

Summary and conclusions

- Overall, many crucial points of the prohibition of automated decisions under the GDPR remain unclear from a legal point of view, in particular the extent and scope of the exemptions, and above all with regard to cases where legally protected characteristics are used as well as the specific information obligations. The legal provision does stipulate that information must be provided in a precise, transparent, comprehensible and easily accessible form in clear and simple language. This also refers to the so-called involved logic, which is interpreted as the structure and procedure of automated data processing. Furthermore, information would also have to be provided on the scope and intended impact of such processing, which would also have to provide information on the risks, including risks of discrimination. Further mandatory clarifications provided in advance would be helpful in this respect and would serve to reduce ambiguities of interpretation before any court rulings in the event of damages.

- The criteria for applying the prohibition should take into account whether the lack of traceability and explanation of “decisions” of the computer systems is substantiated by the controller as the criterion for the existence of an automated decision.
- It can be seen as particularly problematic that automated decisions based on protected characteristics are also permitted if data subjects have given their explicit consent. However, doubts are increasing as to the effectiveness of consent in data protection.[117]

6.2.4 Communicative processes in differentiation decisions

On the grounds of the objective to protect the free development of the personality and the right to self-representation,[118] any differentiation decisions concerning the development of personality must be designed as communicative processes (Trute 1998: 825; Britz 2008: 185). The risks of statistical or algorithmic discrimination could be mitigated through communicative processes in differentiation decisions. According to this, instead of a unilateral assessment of persons and the attribution of characteristics, the communicative processes should provide the option, in accordance with the right to self-representation, of enabling one’s own self-image to contribute to the process of generating the image of a person and of creating possibilities for comparing and correcting the external images.

Neither anti-discrimination law nor the right of informational self-determination currently operationalised in data protection law[119] establish an adequate basis for ensuring that data subjects in decision-making situations always have the opportunity to shape the personality profile in a communicative process. The prevailing approach in both discrimination and data protection law, namely to have the opportunity to correct external images only after damages have been perceived or errors have been discovered, if necessary by means of laborious self-protection measures or legal proceedings, cannot comply with the right to self-representation in order to safeguard the right to free development of personality.[120] This is illustrated, for example, by Case 22 on multiple discrimination in online lending, where the data subject had no opportunity to provide information on his or her actual ability to repay before the credit decision was taken.

The basis for communicative processes is the understanability and **comprehensibility of the decision-making rules**, which include the criteria of the decision and the relations between the criteria and the conclusions drawn from them. This is required in order for the data subject to know whether and in what way the criteria apply to his or her life situation and whether completely different circumstances or criteria in his or her life situation do fulfil the differentiation objective, thus allowing him or her to recognise that this must be communicated to the decision-making body. The data subject must also have the chance to recognise not only that he or she has the option in principle to present his or her own point of view, but also when and why it is important to do so. Decision-making procedures would have to provide for the possibility that the data subject can also ensure that criteria within communicative decision-making procedures are relativised, supplemented, adapted or revised before a decision is taken. This fundamental possibility of reducing the disadvantages of the phenomenon of statistical discrimination promoted by algorithms should be maintained and also implemented electronically.

[117] See Section 6.1.3.1.

[118] See Section 5.2.5.

[119] Also despite the right to contribute one’s own point of view in automated decisions, see Section 6.2.3.2.

[120] This is made even more difficult by the fact that many offers are subject to non-negotiable contractual and data processing conditions unilaterally defined by the providers, to which the data subjects must agree or waive the service or product altogether.

In order to maintain the efficiency gains, such communicative processes can also be automated. Instead of using automation to automatically generate external images of an individual, even by means of a quasi **secret**[121] enforcement of decisions on an individual, IT-supported facilitation of the presentation of one's own point of view or self-representation can also be sought. Likewise, the introduction of the self-image can be improved by allowing self-selection or assigning the persons concerned to differentiation categories. In terms of the protection of personality, it would make more sense to use openly comprehensible, differentiated typologies of persons or categories of products and services to which customers, etc. can assign themselves, instead of secret identification and classification into customer segments.

6.2.5 Design of online platforms

It can generally be assumed that online platforms could also limit the risks of discrimination via algorithmic systems that they harbour (see Section 5.1.3), and that they could achieve even better results than in conventional markets and exchange relationships. The operators of online platforms are able to centrally and efficiently control unnecessary or undesired information flows, e.g. about certain characteristics of persons (e.g. gender or ethnic background), in the interactions and transactions taking place on the platform (Edelman & Luca 2014: 10) and act as a neutral intermediary (Hannák et al. 2017: 1914). While the direct visibility of the characteristics of persons in conventional face-to-face trading and exchanges can promote stereotypical behaviour, online markets could in principle also conceal characteristics that are sensitive or particularly prone to discrimination.[122]

Using the example of the online platform Airbnb, Edelman et al. make **suggestions** on how to improve online platforms with the aim of reducing the risks of discrimination, such as preventing the name of participants from being displayed or avoiding checking people before booking (Edelman et al. 2017: 117ff.). Similarly, Hannák et al. propose that online marketplaces for employment services should function without demographic data, i.e. requests for work services should only be addressed to individuals and to selected groups. They also state that the operators of the online platforms could make adjustments in order to offset biased ratings (Hannák et al. 2017).

Chen et al. also conclude from their research results (see Case 3) that online employment services could play an active role in overcoming structural inequalities in labour markets by having their ranking algorithms present outcomes according to the criterion of group fairness, i.e. according to the distribution of the relevant groups (e.g. women and men) in the population, rather than reflecting structural inequalities (Chen et al. 2018: 10). The tasks of the equality bodies can also be based on the design of the algorithmic differentiation rules of online platforms (see below).

[121] In principle, under the law of informational self-determination, the secret collection and processing of data are associated with certain dangers. "On the one hand, secrecy deprives person concerned of the possibility of avoiding the disclosure of information and the associated risks of disadvantage by adapting their behaviour in order to protect themselves. On the other hand, the possibility of subsequent legal protection, in particular the subsequent correction of incorrect information, is excluded." Britz (2010: 579, translated from the German original).

[122] Levy and Barocas (2017) present additional possibilities and examples of online platform design that can reduce discriminatory behaviour by users.

6.3 Possibilities for equality bodies

6.3.1 Mission and competencies

European anti-discrimination directives require Member States to designate bodies to promote the principle of non-discriminatory equal treatment of all persons[123] and, additionally, to analyse, observe and provide support.[124] The German legislator has decided to set up a central body, the Federal Anti-Discrimination Agency, to implement these directives. The relevant guidelines include three core tasks, which Germany has addressed with Sections 25ff. AGG, in particular Section 27 AGG: (1) to provide independent assistance to victims of discrimination in pursuing their complaints about discrimination; (2) to conduct independent surveys on discrimination; (3) to publish independent surveys and make recommendations based on them.[125] In contrast to some other EU countries (e.g. the United Kingdom, Belgium, Romania), the AGG does not provide for a right of action for the German anti-discrimination body nor for a right of action by associations. Moreover, it has no own authority to conduct investigations or demand disclosures.

6.3.2 Possibilities for investigations and evidence

It is worth considering whether the technical transparency and testing instruments described above[126] are suitable, useful or necessary for the work of equality bodies. Within this context, it must be noted that the direct testing of algorithms and computer systems requires a high level of expertise in computer science. If the testing for risks of algorithmic discrimination is to be carried out externally, there may be a case for concentrating such knowledge. However, knowledge of facts, the actors and persons affected and their interests, the conditions of the relevant sectors or industries, the considerations, regulations and provisions already established for differentiations are equally important. In particular, knowledge of previous unequal treatment, situations at risk of discrimination and groups at risk of discrimination play a role, as do practices and (economic) motivations for differentiation and potential abuse. This expertise is necessary, for example, to understand and verify the assumptions underlying the selection and application of algorithms, models and criteria of differentiation or to question or confirm the legitimacy of the approach. In addition to the knowledge base, issues concerning legally authorised access to the necessary data, algorithms and systems are a prerequisite.

Irrespective of the question of direct inspection and access to the systems, equality bodies or researchers can also use “classical” empirical investigations and discrimination analyses for the application of algorithms, which can be supplemented with specialised algorithm audits.[127] These examine mainly the **outcomes and consequences** for persons concerned in terms of unequal treatment or generated inequalities resulting from algorithmic and data-based differentiation decisions.

[123] According to Article 13 of Directive 2000/43/EC.

[124] Pursuant to Article 12 of Directive 2004/113/EC and Article 20 of Directive 2006/54/EC.

[125] Furthermore, Article 20 of Directive 2006/54/EC adds as a description of tasks the exchange of available information at the appropriate level with the relevant European bodies. Similarly, under Article 12 of Directive 2000/43/EC, Article 14 of Directive 2000/78/EC, Article 11 of Directive 2004/113/EC and Article 22 of Directive 2006/54/EC Member States are required to promote dialogue with appropriate non-governmental organisations that have a legitimate interest in the categories of discrimination covered by the Directives and are involved in combating discrimination.

[126] See Section 6.1.1 above.

[127] See Section 6.1.2 above.

In principle, as communication, interactions and transactions in the private and public sectors become increasingly computerised and networked (or “digitised”), the volume of data for statistical studies on inequality and unequal treatment can also be expected to grow. Within this context, it is more a question of access rules to the data, i.e. to what extent they can be made available for investigations by equality bodies. This could include specific or extended obligations for the provision of information by the applying entities to recognised equality bodies. For the public sector, the possibilities arising from the Federal Act Governing Access to Information held by the Federal Government could also be taken into account and, if necessary, expanded (e.g. Fink 2018).

In many of the examples in Chapter 4, online platforms, e.g. for housing rentals or job advertisements, were examined in terms of their outcomes and consequences and the resulting unequal treatments. They demonstrate that means such as empirical investigations and audit studies appear suitable for examining online platforms with their algorithm-based transaction rules and decision-making rules for discriminatory practices. It can be assumed that, given the techniques available for online procedures, such as automatically repeated queries or the uses of web crawlers or the uses with fictitious persons or accounts,[128] it may even be easier to obtain data compared to an offline procedure. Some example investigations[129] have also shown that, in addition to the users, the algorithms of the online platforms have also given rise to risks of discrimination, which were revealed “from the outside” without direct access to the programme code. However, Case 26 of the case study of the transport service provider Uber shows that the investigation of online platforms can also encounter access problems to the relevant data.

Furthermore, some algorithms and machine learning and artificial intelligence systems can be used as online services and tested with different data sets, as in Case 46, which describes the facial recognition services of Microsoft, IBM and Face++, or in Case 47 describing the facial recognition system of Amazon. The direct inspection of the algorithm and the programme code did not seem to be necessary, but unequal treatment or discrimination was identified by analysing the outcomes of the online services.

However, it is not (yet) possible to draw general conclusions from the cases, as they are too heterogeneous and do not originate from a systematic survey.[130] This is because other examples show that unequal treatment and discrimination could only be identified and (partially) proven in legal disputes, where the procedure of data analysis and evaluation of the decision-making criteria had to be (partially) disclosed (e.g. Case 2 and Case 31). Case 30 of the system of the Austrian Arbeitsmarktservice also demonstrated that risks of discrimination were made accessible to a public discussion by publishing documentation on computation formulas.

6.3.3 Experiences and suggestions from other equality bodies

Equality bodies in the EU countries have little experience in dealing with risks of algorithmic and data-based discrimination. There are also only a few specific cases of discrimination to date. A number of EU

[128] See Section 6.1.2.

[129] See, for example, Case 17, Case 19 or Case 20.

[130] The fact that there are many research results on online platforms and online systems may be because, compared to direct access to the programme code, online systems and online platforms may relatively well to investigate, for example, because problems of direct access to algorithms or programme code did not arise (e.g. due to protection of trade secrets or copyright issues) or the research effort with statistical analyses may still have been justifiable.

equality bodies with some previous experience in the field were asked about the possibilities of investigating and authority to do so:[131]

- Knowledge is required of how algorithms work and in which areas of life and conditions they are used and have an impact on protected groups of people. In this regard, personnel with knowledge of data, data use and anti-discrimination law are needed, more so than technical tools. Instead of expertise in computer science or data science within the body, the body could benefit from structural partnerships with computer or data scientists to avoid discrimination, uncover cases of discrimination or obtain evidence on the cases. Audit studies and discrimination testing are considered to be suitable for detecting unequal treatments involving computer systems and algorithms.[132]

- One example of a case of discriminatory insurance rates in the Czech Republic highlights the importance of access to statistical data and actuarial methods of the insurance companies. Without access and the obligation to provide data and information on the request of the Czech equality body, it would not be possible to handle the case and assess whether or not discrimination had occurred.[133]

- The case of suspected discrimination in job advertisements in the “social” online network Facebook (see Case 4) underscored that it was not possible for the persons concerned to become aware of the unequal treatment because they could not see the advertisements. Other persons who were able to see the advertisements informed the equality body, and this is the only reason the body was able to process the case. As the protected characteristics “age” and “gender” were used in the case, it was relatively easy to discover, examine and produce evidence. In order to achieve a comprehensive understanding and be able to take legal action against the use of certain algorithms, computer specialists would be needed to analyse systems and algorithms in detail. The exact analysis would also only be possible if the algorithms were made fully accessible for investigation, which is considered unlikely due to the protection of trade and business secrets. The statistical studies commonly used in anti-discrimination research and analysis of discrimination cases are also considered suitable for studying inequalities caused by algorithms and computer systems.[134]

- Statistics, and in particular equality data, are considered a key element in the analysis of alleged cases of discrimination using algorithms and computer systems. This may include reforms of the legal framework aimed at alleviating structural problems, such as lack of data or lack of consequences when fails to respond to requests by the equality body. Changes in the legal framework could include, for example, the obligation for certain public bodies to collect equality data, the obligation to report to equality bodies on automated decision-making systems, mandatory cooperation between entities using the systems and bodies responsible for equality data and legal consequences if entities using algorithms do not provide the data to the equality bodies on request.[135]

[131] A number of European equality bodies were asked about this issue by email.

[132] Information provided by employees of the Belgian equality body Unia, by email to the author, March 2019.

[133] Information provided by employees of the Czech equality body Office of the Public Defender of Rights, by email to the author, April 2019.

[134] Information provided by employees of the equality body The Danish Institute for Human Rights, by email to the author, March 2019.

[135] Information provided by employees of the Slovenian equality body Advocate of the Principle of Equality, by email to the author, April 2019.

6.3.4 Preventive approach and cooperation possibilities

In many instances, the legal instruments of discrimination and data protection are based on the fact that harm or injustice has already occurred. As explained above, algorithmic discrimination can also take place unnoticed or be deliberately disguised, occur unintentionally via correlations to protected characteristics, or it can be extremely difficult to prove discrimination.

For entities developing and using algorithms and computer systems, the legal framework also offers insufficient guidance in the design of non-discriminatory algorithms and systems, as there are too many uncertainties and possible interpretations that could only be clarified by court rulings, if at all. Uncertainties about legality can cause misinvestment or prevent innovations from being realised. As such, there is a strong case for a preventive approach. To this end, equality bodies or supervisory institutions could assume a variety of tasks, but they need the appropriate equipment.

The author asked certain equality bodies whether a preventive approach to risks of discrimination through the use of algorithms and computer systems is more appropriate or even necessary (such as studies initiated by the institution). The equity bodies responded with the following **proposals** (notes of the author in brackets):

- Awareness must be raised at an early stage among the developers and the responsible controllers when creating algorithms in order to bring together legal, ethical and technical points of view. This can also be achieved, for example, by increasing employee diversity at the developers' organisations so that thought and argumentation processes are not carried out exclusively by a majority group. As expertise in the field of algorithms continues to grow, and on the basis of possible audit studies, a strategy to raise awareness about the risks of discrimination through algorithms may become important. In addition to the preventive approach, it is important to continue the work on high-quality equality data,[136] as the quality of the algorithms may also depend on the availability of non-biased and accurate data, in particular data representing groups at risk of discrimination.[137] (In addition to the statements of the equality body Unia, it should be noted that equality data can also help to better identify risks of discrimination in relation to specific groups. This can also support the proof of discrimination in the context of algorithms.)

- In view of the increasing use of automated decision-making tools by providers of goods and services, the preventive approach is considered important, particularly in terms of increasing awareness and knowledge of ethical handling. The best way to do this is by informing the persons and companies preparing to use such techniques in their business and practices. On the other hand, complaint mechanisms should also be accessible to people affected by discrimination through automatic decision-making systems.[138] (In addition to the statements of the equality body Advocate of the Principle of Equality, it can be noted that also in the workplace, internal complaints procedures under Section 13 AGG should give employees the opportunity to address risks of discrimination through algorithms.)

[136] See also Baumann, Egenberger & Supik (2018).

[137] Information provided by employees of the Belgian equality body Unia, by email to the author, March 2019.

[138] Information provided by employees of the Slovenian equality body Advocate of the Principle of Equality, by email to the author, April 2019.

For example, the Belgian equality body Unia provides a website where companies can perform a relatively quick scan to uncover any problems in terms of diversity and developments using a questionnaire to be filled in at their premises ("Quick scan of diversity").[139] Taking into account specific challenges in terms of methods, technical issues, workload and ethical questions, a similar tool could also be offered for decisions on the use of algorithms, for example.[140]

It also appears worthwhile for equality bodies and online platforms to cooperate on a preventive approach (see also Section 6.2.5). In this context, a balance can be sought between, on the one hand, the objective of the platforms to provide as much information as possible about users and, on the other, the protection of the bearers of protected characteristics. The latter would be achieved not only by avoiding direct use of the characteristics, but also by preventing conclusions from being drawn about surrogate variables and correlations. Through the mass impact of online platforms, non-discriminatory practices could thus be implemented for comparatively many people at once. Case 9 shows a cooperative approach between the National Fair Housing Alliance (NFHA) and the company Facebook, where the NFHA offers a training programme for the company.

Furthermore, Zuiderveen Borgesius proposes cooperation between data protection institutions and equality bodies, as well as for public bodies wishing to use artificial intelligence algorithms an obligation to consult equality bodies in advance and to involve them in public procurement processes (Zuiderveen Borgesius 2018: 31).

Conclusion

A preventive, cooperative approach between equality bodies and entities developing and using algorithms and computer systems could include several elements, such as (1) advice on the legitimate or prohibited use of protected characteristics depending on decision-making situations and groups of persons concerned, (2) interpretation and advice on the use of proxies, surrogate information or variables with correlations to protected characteristics or of apparently neutral criteria in the case of indirect discrimination, or (3) interpretation and investigation of the possibilities of implementing justice and fairness criteria for different differentiation situations.

[139] See website in French: https://www.ediv.be/site/fr/ediv-quickscan-non-discrimination-et-egalite-des-chances or in Dutch: https://www.ediv.be/site/nl/edivquickscannondiscriminatieengelijkekansen (last retrieved on 28 August 2019).

[140] Information provided by employees of the Belgian equality body Unia, by email to the author, March 2019.

6.3.5 Proposals for the Federal Anti-Discrimination Agency

The following proposals for the Federal Anti-Discrimination Agency (FADA) can be derived from the preceding findings:

- Anti-discrimination law is relevant to all those algorithmic and data-based differentiation decisions that can lead to worse treatment on the basis of a characteristic protected in the AGG in the areas of working life and access to goods and services. The AGG already prohibits these types of discriminations, regardless of whether or not they are made using algorithms.[141] A further-reaching legal prohibition therefore does not appear necessary within the scope of labour and civil law. However, developments in algorithms, in particular artificial intelligence, need to be further observed and researched in order to make adjustments to the protected characteristics of the AGG where necessary.

- However, the AGG has weaknesses, as it is limited to the individual enforcement of rights. As such, discriminatory practice cannot be prohibited, only individual victims can claim damages or compensation in civil lawsuits. The approach of a merely selective procedure in individual cases does not seem appropriate in view of the potentially systematic worse treatment of many persons concerned by algorithmic differentiations. The right of collective action by associations would be a necessary response to the mass phenomenon[142] of possible algorithmic discrimination and the poorer perceptibility and provability of algorithmic discrimination.

- In order to be able to prove discrimination through algorithm-based decisions (in court) and justify claims for repayment for damages or compensation, documentation obligations should be imposed on artificial intelligence systems or systems that are particularly discriminatory. In specific cases of suspected discrimination, the equality body should be given access to such documentation for identification purposes. Access to or release of the documentation would have to be regulated by law in such cases. Furthermore, it would be worth considering whether access or release could also apply to data sets[143] and algorithms themselves and how the protection of secrets could then be safeguarded.

- From the perspective of the persons concerned with regard to complaints and legal support, ambiguities about jurisdiction can lead to a situation where different responsibilities and procedural channels would be established for algorithmic and non-algorithmic differentiation decisions. This is especially true given the increasing difficulty of making this distinction precisely in practice. Different responsibilities should also be avoided, as this could lead to different levels of protection.

[141] See also Section 6.2.1.

[142] See p. 23 for more on the mass phenomenon.

[143] The many described cases of risks of discrimination arising from the (further) use of data sets reflecting previous unequal treatment, particularly in the development of risk assessment systems (see, for example, Case 27 or Case 34), would support this. A basic understanding of the practices pertaining to how and by whom the data is generated and analysed appears necessary in order to be able to assess the risks of discrimination arising from the use of the systems. However, this presupposes the high skill requirements described in Section 6.1.2.

- The legal mandate for a national equality body to be responsible for the implementation of the European anti-discrimination directives imposes a responsibility for investigations and evaluations under discrimination law. In order to fulfil the legal mandate, technical expertise should be developed or acquired through cooperation with research institutions. Moreover, this legal mandate should be considered and embedded in the further design of the regulatory framework for algorithmic differentiations and decision-making systems.

- In order to fulfil its statutory research mandate, the equality body should also conduct reviews without suspicion (testing or algorithmic audits) of the outcomes of differentiation decisions, e.g. for online platforms. It could also do this in cooperation with research institutions (see above).

- The equality body should also make preventive offers to avoid algorithm-based risks of discrimination (see above) and cooperate with companies to this end. Other useful measures include mandatory provisions for consultation of the equality body in the procurement and prior to the use of computer systems by public authorities that are prone to discrimination, such as certain AI systems.

6.4 Need for societal considerations and decisions

Beyond the more concrete needs for action and options for the purpose of optimising existing regulatory structures, there is a need for societal considerations and political and legislative decisions that fundamentally call into question the suitability of existing regulatory and institutional approaches in view of the developments in algorithms, applications and practices. This concerns the approach of self-protection and the resulting burdens on the individual and the legitimacy of algorithmic and data-based differentiations with regard to the weighing of advantages and societal risks.

6.4.1 Burdens on the affected individuals

Both the right to informational self-determination and anti-discrimination place the burden of responsibility on the affected individual to identify and take action against unlawful data processing and unjustified unequal treatment. However, questions arise as to whether these basic legal concepts are still appropriate at all, given the increasing amount of data and algorithm-based and automated decision-making procedures and their special characteristics. This is because such burdens of responsibility require very high knowledge cognitive and temporal prerequisites on the part of the affected individual in order to (a) be able to perceive the many situations involving data processing and differentiation at all, (b) process the information resulting from the information obligations of data protection (such as the “logic involved” in automated decisions) as well as information, correction or deletion rights and, above all, (c) assess the individual consequences resulting from data processing and diverse (potential) differentiation decisions for themselves and to detect the risk of possible discrimination for themselves.

For example, many data processing operations only lead to differentiation decisions after a long time, which requires high prognostic abilities from the individual. In any case, it is nearly impossible for the individual to assess the consequences of using data for external decision-making purposes and in other contexts. The scepticism is further reinforced by the increasingly perceived inadequacy of informed consent,[144] which requires the above-mentioned assessments of the individual at the time of consent, and by the difficulty in detecting and proving discrimination by the (potentially) affected individuals.[145] A number of instances in which personal data are collected (such as web tracking) during the use of smartphones and their apps or the analysis of communication in “social” online networks are carried out more or less without the knowledge of the users; as a result, the possibilities for self-protection are also severely limited.

The limited possibilities for self-protection of the (potential) persons concerned must also be taken into account in proposals for regulating risks of algorithmic discrimination, such as the proposal for a labelling requirement when using automated decisions. As this would only point to the existence of automated decision-making systems, but not to their consequences, the instrument would not improve the problem of insufficient self-protection.

In accordance with the principle of subsidiarity, this may lead to calls for a more representative approach by protective institutions such as equality bodies, consumer protection and data protection bodies or specialised authorities, either with (further) help for individuals or an approach instead of individuals, which seems even more effective. Another shift of responsibility for avoidance measures to the entities developing and using algorithmic and data-based differentiations (which, however, requires supervision) would also be debatable. To this end, for example, the obligations under data protection law to document, to carry out data protection impact assessments or appoint a data protection officer could be expanded or supplemented internally by anti-discrimination provisions.

The procedure by representative institutions can also be needed, in particular by the increasing structural advantage[146] of the entities developing and using algorithmic differentiation against persons concerned. This can arise particularly through the reinforcement of dependency on a certain service or product, which in turn can be caused by network effects of online platforms or other monopolisation tendencies or can be increased by the fact that personality traits can be more thoroughly researched and exploited. Structural advantage can also be increased by reducing the number of alternatives in the form of offline or analogue alternatives.

In addition, there is the recurrent problem that data protection law is conceptually based on the individual, but algorithmic differentiations often relate to groups without necessarily having to identify an individual unambiguously by name or in any other way. In cases of doubt or legal disputes about the personal reference, the right could thus prove to be “toothless” (Barocas & Nissenbaum 2014; Mantelero 2016; Zuiderveen Borgesius 2016). This would require clarification through clear legal provisions. The problem also justifies action by representative institutions instead of the affected individuals.

6.4.2 Legitimacy of differentiations

As described in Chapter 5, algorithmic and data-based differentiations generate societal risks in addition to technical risks. With regard to societal risks, technical improvements to algorithms or data sets are of no help. Instead, it is for societal considerations and decisions to determine the significance of societal values and objectives as well as the desired or undesired practices of differentiation, in order to avoid societal risks. The handling of statistical discrimination has until now also been the subject of societal considerations and decisions, above all through the design of anti-discrimination law. However, due to a number of

[144] See Section 6.1.3.1.
[145] See Section 6.1.3.2.
[146] See Section 5.2.6.

developments in data processing and algorithmic applications, the advantages and disadvantages need to be re-evaluated by means of societal consideration processes:

(1) The legitimacy of the algorithmic differentiations is partly justified by cost considerations and efficiency goals, which justify the use of surrogate information or proxies in differentiation decisions primarily with the efficient solution of information deficits. The alternative case-by-case individual assessment may be too costly for many differentiation decisions or may itself lead to problems of privacy protection or stigmatisation (Britz 2008). Abandoning forms of statistical or algorithmic differentiation can result in societal costs in the form of relative inefficiencies or lost instrumental benefits. In concrete terms, this would be expressed in the form of the costs of individual case assessments (Schauer 2018: 50).

On the other hand, there are the risks of injustice by generalisation due to inappropriate or incorrect surrogate information.[147] The necessity and appropriateness of algorithmic differentiation must therefore be considered when weighing the advantages and disadvantages, as well as the question of whether less risky alternatives are available. Aspects such as clear evidence of consistent improvement in the accuracy of predictions or objectivity of decisions, determinations of fairness and error rates acceptable for different risks, contexts of use and areas of life[148] as well as provisions on scientifically recognised procedures[149] of data analysis and algorithm-based decision-making procedures must also be clarified in this context.

However, these considerations also raise questions about how efficiency gains and societal costs or risks of algorithmic differentiation are distributed in society, e.g. whether they affect all or only a few, and whether those who benefit from efficiency gains are also those who bear the risks or whether the risks are externalised. For example, one should bear in mind that in statistical differentiation by private companies, resources are not conserved for the collective benefit, but merely in one's own self-interest (Britz 2008: 49f.). Furthermore, the distribution of efficiency gains must also be assessed in light of the increasing concentration of the relevant markets and the dominance of a small number of companies.

(2) Due to the fact that data on the affiliation to certain categories of persons as well as algorithms and software systems containing them are now available comparatively easily and cheaply, it is likely that such data and decision-making procedures based on them are "overused" (Schauer 2003). As a result, algorithmic and data-based differentiations are being realised in many areas where equal treatment previously prevailed, or even individual assessments and other decision-making regulations can be systematically pushed back, even if they would be possible with reasonable effort:

(a) Algorithmic differentiations can penetrate into areas where differentiations were previously unwanted from a view of societal equality objectives or socio-political goals.[150]

(b) Algorithmic differentiation is increasingly being used for decisions that have a significant impact on human dignity and the free development of personality (such as imprisonment, the extent of state controls, and access to housing, jobs, education opportunities or credit). In actual automation

[147] See Section 5.2.1.

[148] See Section 6.1.1.

[149] See Section 6.2.2.

[150] See Section 5.2.3.

practices, people can increasingly be treated only as mere means, because the persons concerned are effectively no longer able to consent to the practices.[151]

(3) Algorithmic and data-based differentiations are in many cases based on the collection and analysis of large amounts of (increasingly sensitive) personal data, which can undermine the right of **informational self-determination**.

The protective goals of the free development of personality are primarily realised through anti-discrimination law and the right to informational self-determination. They are (according to Britz 2010),[152]

(a) to guarantee external freedom of development by ensuring freedom of conduct and protection against adverse decisions by others, by ensuring that the potentially restrictive decisions of others can be influenced by the persons concerned in such a way that they are in their favour to the greatest possible extent,

(b) to guarantee inner freedom of development through the fact that the personality development in interactive processes can still be perceived by the individual as free and belonging to oneself, and that individuals can still assert themselves against public images and,

(c) to protect the uninhibitedness of individual behaviour by reducing the freedom-inhibiting effects of "abstract uncertainty" or avoiding intimidation. The latter arise from information inventories and data processing purposes that are no longer clear to the individual.[153]

For example, algorithmic differentiations can increasingly be based on comprehensive and detailed personality profiles[154] which, as an enhanced form of heteronomy, are suitable for "imposing" an external image on a person, without the person concerned having a chance to develop their own personality and role interpretation in social contexts.[155] Therefore, developments of concern in data protection also need to be further examined in terms of their impact on protection against algorithmic and data-based discrimination. These include the problematic merging of personal data records, in particular through data trading or data brokerage, or the (company internal) merging of data and the softening of the purpose limitation of the data use[156] and the transfer to new purposes. Furthermore, societal considerations must take into account who (the entities using algorithmic differentiation or the persons concerned) will benefit from increased tracking, data analysis and differentiation and to whom disadvantages will be externalised.

Considerations, decisions and implementation in regulatory measures should be based for individual decision-making situations and contexts in each case according to the specific level of risk of injustice by

[151] See Section 5.2.4.

[152] See also Section 5.2.5.

[153] These intimidation effects are expected not only from incomprehensible data bases but also from the misuse of personal data. For example, people may refrain from using applications that are actually beneficial if they have to fear that the data collected when using the application and related to them could be used in other contexts and for other purposes that are not in their interest (see Yeung 2018: 33). These intimidation and self-restraint effects are also examined under the term "chilling effects"; see, for example, Baruh (2007), Schwartz (1999), Das & Kramer (2013), Lang & Barton (2015), Marder et al. (2016), Marthews & Tucker (2017), Penney (2016), (2017), Orwat & Schankin (2018).

[154] See Section 2.2.2.

[155] See Section 5.2.5.

[156] See Raabe & Wagner (2016) on purpose-limitation in the GDPR.

generalisation, on the extent of the threat to human dignity and the free development of personality, as well as on the possibilities and limits of self-protection and the possible overburdening of individuals. In instrumental terms, this can involve the detailed regulation of decision-making procedures[157] and also – depending on the extent of the risk – prohibitions[158] or contain regulations on the use of certain algorithms, data processing and decision-making procedures, forms of differentiation or decision-making criteria. Further instruments can be the legal requirement to assess the societal risks for objectives of euqality and personal development by developing and implementing entities or their obligation to take protective measures in terms of avoiding discrimination. These can extend to strengthening the competencies and authority of specialised institutions for discrimination and data protection.

[157] See Section 6.2.

[158] An example of a ban on algorithmic systems is the ban on facial recognition systems imposed by the city of San Francisco. See Conger, Fausset & Kovaleski (2019).

4. Bibliography

Acquisti, Alessandro; Taylor, Curtis; Wagman, Liad (2016): The Economics of Privacy; in: Journal of Economic Literature, Vol. 54, No. 2, pp. 442-492.

Agrawal, Ajay; Gans, Joshua; Goldfarb, Avi (2016): The Simple Economics of Machine Intelligence; in: Harvard Business Review, Vol. 17, Nov. Issue, pp. 2-5.

Agrawal, Ajay; Gans, Joshua; Goldfarb, Avi (2018): Prediction Machines. The Simple Economics of Artificial Intelligence; Boston: Harvard Business Review Press.

AI Now Institute (2018): Litigating Algorithms: Challenging Government Use of Algorithmic Decision Systems. An AI Now Institute Report, in collaboration with Center on Race, Inequality, and the Law and Electronic Frontier Foundation; New York: New York University, AI Now Institute.

Albers, Marion (2017): Informationelle Selbstbestimmung als vielschichtiges Bündel von Rechtsvorschriften und Rechtspositionen; in: Michael Friedewald, Jörn Lamla and Alexander Roßnagel (eds.): Informationelle Selbstbestimmung im digitalen Wandel; Wiesbaden: Springer, pp. 11-35.

Ali, Muhammad; Sapiezynski, Piotr; Bogen, Miranda; Korolova, Aleksandra; Mislove, Alan; Rieke, Aaron (2019): Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes; arXiv e-prints.

Allhutter, Doris (2019): AMS-Algorithmus am Prüfstand. ITA-Dossier No. 43; Vienna: Institut für Technikfolgen-Abschätzung (ITA).

Alpaydin, Ethem (2016): Machine Learning; Cambridge, London: The MIT Press.

Altenburger, Kristen M.; Ho, Daniel E. (2018): When Algorithms Import Private Bias into Public Enforcement: The Promise and Limitations of Statistical Debiasing Solutions; in: Journal of Institutional and Theoretical Economics (JITE), Vol. 175, No. 1, pp. 98-122.

an der Heiden, Iris; Wersig, Maria (2017): Preisdifferenzierung nach Geschlecht in Deutschland – Forschungsbericht. Eine Studie im Auftrag der Antidiskriminierungsstelle des Bundes; Baden-Baden: Nomos.

Ananny, Mike; Crawford, Kate (2018): Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability; in: New Media & Society, Vol. 20, No. 3, pp. 973-989.

Angrave, David; Charlwood, Andy; Kirkpatrick, Ian; Lawrence, Mark; Stuart, Mark (2016): HR and analytics: why HR is set to fail the big data challenge; in: Human Resource Management Journal, Vol. 26, No. 1, pp. 1-11.

Angwin, Julia; Larson, Jeff; Mattu, Surya; Kirchner, Lauren (2016): Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks, ProPublica, online article dated 23 May 2016, available at https://www.propublica.org/article/machine-bias-riskassessments-in-criminal-sentencing (last retrieved on 27 August 2019).

Angwin, Julia; Scheiber, Noam; Tobin, Ariana (2017): Dozens of Companies Are Using Facebook to Exclude Older Workers From Job Ads, ProPublica, online article dated 20 December 2017, available at Https://www.propublica.org/article/facebook-ads-age-discrimination-targeting (last retrieved on 28 August 2019).

Angwin, Julia; Tobin, Ariana; Varner, Madeleine (2017): Facebook (Still) Letting Housing Advertisers Exclude Users by Race, ProPublica, online article dated 17 November 2017, available at https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin (last retrieved on 28 August 2019).

Arentz, Christine; Rehm, Rebekka (2016): Behavior-based Tariffs in Health Insurance Compatibility with the German System; Cologne: Otto Wolff Institut für Wirtschaftsordnung.

Arrow, Kenneth J. (1973): The Theory of Discrimination; in: Orley Ashenfelter and Albert Rees (eds.): Discrimination in Labor Markets; Princeton: Princeton University Press, pp. 3-33.

Arrow, Kenneth J. (1998): What has economics to say about racial discrimination?; in: Journal of Economic Perspectives, Vol. 12, No. 2, pp. 91-100.

Bäcker, Matthias (2018): DS-GVO Art. 13 Informationspflicht bei Erhebung von personenbezogenen Daten bei der betroffenen Person; in: Jürgen Kühling and Benedikt Buchner (eds.): Datenschutz-Grundverordnung, Bundesdatenschutzgesetz: DS-GVO/BDSG, legal commentary, 2nd edition; Munich: Beck.

Barocas, Solon; Hood, Sophie; Ziewitz, Malte (2013): Governing Algorithms: A Provocation Piece; New York: New York University, Department of Media, Culture, and Communication.

Barocas, Solon; Nissenbaum, Helen (2014): Big Data's End Run around Anonymity and Consent; in: Julia Lane, Victoria Stodden, Stefan Bender and Helen Nissenbaum (eds.): Privacy, Big Data, and the Public Good. Frameworks for Engagement; New York: Cambridge University Press, pp. 44-75.

Barocas, Solon; Selbst, Andrew D. (2016): Big data's disparate impact; in: California Law Review, Vol. 104, pp. 671-732.

Bartlett, Robert P.; Morse, Adair; Stanton, Richard; Wallace, Nancy (2018): Consumer lending discrimination in the FinTech era, UC Berkeley Public Law Research Paper.

Baruh, Lemi (2007): Read at your own risk: shrinkage of privacy and interactive media; in: New Media & Society, Vol. 9, No. 2, pp. 187-211.

Baumann, Anne-Luise; Egenberger, Vera; Supik, Linda (2018): Erhebung von Antidiskriminierungsdaten in repräsentativen Wiederholungsbefragungen. Bestandsaufnahme und Entwicklungsmöglichkeiten; Berlin: Antidiskriminierungsstelle des Bundes (ADS).

Beck, Susanne; Grunwald, Armin; Jacob, Kai; Matzner, Tobias (2019): Künstliche Intelligenz und Diskriminierung. Herausforderungen und Lösungsansätze; Munich: Lernende Systeme – Die Plattform für Künstliche Intelligenz.

Becker, Gary S. (1957/1971): The Economics of Discrimination (2nd edition); Chicago: University of Chicago Press.

Bellamy, Rachel K. E.; Dey, Kuntal; Hind, Michael; Hoffman, Samuel C.; Houde, Stephanie; Kannan, Kalapriya; Lohia, Pranay; Martino, Jacquelyn; Mehta, Sameep; Mojsilovic, Aleksandra (2018): AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias; arXiv preprint arXiv:1810.01943.

Berghahn, Sabine; Egenberger, Vera; Klapp, Micha; Klose, Alexander; Liebscher, Doris; Supik, Linda; Tischbirek, Alexander (2016): Evaluation des Allgemeinen Gleichbehandlungsgesetzes (on behalf of the Federal Anti-Discrimination Agency); Berlin: Antidiskriminierungsstelle des Bundes, published by Nomos Verlag.

Berghahn, Sabine; Klose, Alexander; Lewalter, Sandra; Liebscher, Doris; Spangenberg, Ulrike; Wersig, Maria (2014): Handbuch „Rechtlicher Diskriminierungsschutz“; Berlin: Antidiskriminierungsstelle des Bundes (ADS).

Berk, Richard; Heidari, Hoda; Jabbari, Shahin; Kearns, Michael; Roth, Aaron (2018): Fairness in criminal justice risk assessments: The State of the Art; in: Sociological Methods & Research, Online first, pp. 1-42.

BGH (2014): Judgment of the Federal Court of Justice’s Civil Panel VI dated 28 January 2014, VI ZR 156/13 (SCHUFA judgement); Karlsruhe: Federal Court of Justice (BGH).

Bitter, Philip; Uphues, Steffen (2017): Big Data und die Versicherungsgemeinschaft „Entsolidarisierung“ durch Digitalisierung?, ABIDA-Dossier; Münster: University of Münster, Institute for Information, Telecommunication and Media Law.

Blodgett, Su Lin; Green, Lisa; O’Connor, Brendan (2016): Demographic Dialectal Variation in Social Media: A Case Study of African-American English; in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 1119-1130.

Blodgett, Su Lin; O’Connor, Brendan (2017): Racial disparity in natural language processing: A case study of social media african-american english; arXiv preprint arXiv:1707.00061.

Blömeke, Eva; Clement, Michel (2009): Selektives Demarketing Management von unprofitablen Kunden; in: Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung, Vol. 61, No. 7, pp. 804-835.

Bogen, Miranda; Rieke, Aaron (2018): Help wanted. An Examination of Hiring Algorithms, Equity, and Bias; Washington, D. C., Upturn.

Bolukbasi, Tolga; Chang, Kai-Wei; Zou, James Y.; Saligrama, Venkatesh; Kalai, Adam T. (2016): Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings; in: Advances in Neural Information Processing Systems, pp. 4349-4357.

Brantingham, P. Jeffrey; Valasik, Matthew; Mohler, George O. (2018): Does Predictive Policing Lead to Biased Arrests? Results From a Randomized Controlled Trial; in: Statistics and Public Policy, Vol. 5, No. 1, pp. 1-6.

Brauneis, Robert; Goodman, Ellen P. (2018): Algorithmic transparency for the smart city; in: Yale Journal of Law and Technology, Vol. 20, pp. 103-176.

Breiman, Leo (2001): Statistical modeling: The two cultures; in: Statistical science, Vol. 16, No. 3, pp. 199-231.

Brey, Philip (2000): Disclosive Computer Ethics; in: Computer and Society ACM SIGCAS, Vol. 30, No. 4, pp. 10-16.

Brey, Philip (2009): Values in Technology and Disclosive Computer Ethics; in: Luciano Floridi (ed.): The Cambridge Handbook of Information and Computer Ethics; Cambridge: Cambridge University Press, pp. 41-58.

Britz, Gabriele (2007): Freie Entfaltung durch Selbstdarstellung. Eine Rekonstruktion des allgemeinen Persönlichkeitsrechts aus Art. 2 I GG; Tübingen: Mohr Siebeck.

Britz, Gabriele (2008): Einzelfallgerechtigkeit versus Generalisierung. Verfassungsrechtliche Grenzen statistischer Diskriminierung; Tübingen: Mohr Siebeck.

Britz, Gabriele (2010): Informationelle Selbstbestimmung zwischen rechtswissenschaftlicher Grundsatzkritik und Beharren des Bundesverfassungsgerichts; in: Wolfgang Hoffmann-Riem (ed.): Offene Rechtswissenschaft; Tübingen: Mohr Siebeck, pp. 561-596.

Brown, Ian; Marsden, Christopher T. (2013): Regulating Code. Good Governance and Better Regulation in the Information Age; Cambridge, MA: MIT Press.

Bruce, Margaret; Adam, Alison (1989): Expert systems and women's lives: a technology assessment; in: Futures, Vol. 21, No. 5, pp. 480-497.

Brundage, Miles; Avin, Shahar; Clark, Jack; Toner, Helen; Eckersley, Peter; Garfinkel, Ben; Dafoe, Allan; Scharre, Paul; Zeitzoff, Thomas; Filar, Bobby; Anderson, Hyrum; Roff, Heather; Allen, Gregory C.; Steinhardt, Jacob; Flynn, Carrick; Ó hÉigeartaigh, Seán; Beard, Simon; Belfield, Haydn; Farquhar, Sebastian; Lyle, Clare; Crootof, Rebecca; Evan, Owain; Page, Michael; Bryson, Joanna; Yampolskiy, Roman; Amodei, Dario (2018): The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation; Oxford et al.: Future of Humanity Institute, University of Oxford, Centre for the Study of Existential Risk, University of Cambridge, Center for a New American Security, Electronic Frontier Foundation, OpenAI.

Buchner, Benedikt (2018): DS-GVO Art. 22 Automatisierte Entscheidungen im Einzelfall einschließlich Profiling, Jürgen Kühling and Benedikt Buchner (eds.): Datenschutz-Grundverordnung, Bundesdatenschutzgesetz: DS-GVO/BDSG, legal commentary, 2nd edition.; Munich: Beck.

Buolamwini, Joy; Gebru, Timnit (2018): Gender shades: Intersectional accuracy disparities in commercial gender classification; in: Conference on Fairness, Accountability and Transparency, No. 81, pp. 77-91.

Burdon, Mark; Harpur, Paul (2014): Re-conceptualising privacy and discrimination in an age of talent analytics; in: University of New South Wales Law Journal, Vol. 37, No. 2, pp. 679-712.

Burrell, Jenna (2016): How the machine 'thinks': Understanding opacity in machine learning algorithms; in: Big Data and Society, Vol. 3, No. 1, pp. 1-12.

Busch, Christoph (2018): Algorithmic Accountability; Expert report in the context of the “Assessing Big Data” (ABIDA) project; Münster, Karlsruhe: University of Münster, Karlsruhe Institute of Technology.

BVerfG (1983): Judgement date 15 December 1983, case 1 BvR 209/83 et al. (”Census decision”), BVerfGE 65, 1; Karlsruhe: Federal Constitutional Court (BVerfG).

BVerfG (1993): Order of the First Senate of 19 October 1993 – 1 BvR 567, 1044/89 (”Guarantee contracts”), BVerfGE 89, 214; Karlsruhe: Federal Constitutional Court (BVerfG).

BVerfG (2018): Order of the First Senate of 11 April 2018 – 1 BvR 3080/09 (”Stadium ban”; indirect horizontal effects of the right to equality in private law relations), BVerfGE 148, 267-290; Karlsruhe: Federal Constitutional Court (BVerfG).

Cabitza, Federico; Rasoini, Raffaele; Gensini, Gian Franco (2017): Unintended consequences of machine learning in medicine; in: JAMA, Vol. 318, No. 6, pp. 517-518.

Calders, Toon; Custers, Bart (2013): What Is Data Mining and How Does It Work?; in: Bart Custers, Toon Calders, Bart Schermer and Tal Zarsky (eds.): Discrimination and Privacy in the Information Society; Studies in Applied Philosophy, Epistemology and Rational Ethics, Vol. 3; Berlin, Heidelberg: Springer, pp. 27-42.

Calders, Toon; Žliobaitė, Indrė (2013): Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures; in: Bart Custers, Toon Calders, Bart Schermer and Tal Zarsky (eds.): Discrimination and Privacy in the Information Society; Studies in Applied Philosophy, Epistemology and Rational Ethics, Vol. 3; Berlin, Heidelberg: Springer, pp. 43-57.

Caliskan, Aylin; Bryson, Joanna J.; Narayanan, Arvind (2017): Semantics derived automatically from language corpora contain human-like biases; in: Science, Vol. 356, No. 6334, pp. 183-186.

Caruana, Rich; Lou, Yin; Gehrke, Johannes; Koch, Paul; Sturm, Marc; Elhadad, Noemie (2015): Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission, Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1721-1730, published by ACM.

Castelluccia, Claude; Le Métayer, Daniel (2019): Understanding algorithmic decision-making: Opportunities and challenges. Study for Panel for the Future of Science and Technology; Brussels: European Parliament, European Parliamentary Research Service, Scientific Foresight Unit (STOA).

Castelvecchi, Davide (2016): The Black Box of AI; in: Nature, Vol. 538, October 6, 2016, pp. 20-23.

Cate, Fred H.; Mayer-Schönberger, Viktor (2013): Notice and consent in a world of Big Data; in: International Data Privacy Law, Vol. 3, No. 2, pp. 67-73.

Chamorro-Premuzic, Tomas; Akhtar, Reece; Winsborough, Dave; Sherman, Ryne A. (2017): The datafication of talent: how technology is advancing the science of human potential at work; in: Current Opinion in Behavioral Sciences, Vol. 18,Dec., pp. 13-16.

Chamorro-Premuzic, Tomas; Winsborough, Dave; Sherman, Ryne A.; Hogan, Robert (2016): New Talent Signals: Shiny New Objects or a Brave New World?; in: Industrial and Organizational Psychology, Vol. 9, No. 3, pp. 621-640.

Chen, Le; Ma, Ruijun; Hannák, Anikó; Wilson, Christo (2018): Investigating the Impact of Gender on Rank in Resume Search Engines, Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 651, published by ACM.

Chen, Min; Mao, Shiwen; Liu, Yunhao (2014): Big Data: A Survey; in: Mobile Networks and Applications, Vol. 19, No. 2, pp. 171-209.

Chouldechova, Alexandra (2017): Fair prediction with disparate impact: A study of bias in recidivism prediction instruments; in: Big data, Vol. 5, No. 2, pp. 153-163.

Chouldechova, Alexandra; Benavides-Prado, Diana; Fialko, Oleksandr; Vaithianathan, Rhema (2018): A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions; in: Proceedings of the 1st Conference on Fairness, Accountability and Transparency; Proceedings of Machine Learning Research: PMLR.

Christl, Wolfie (2017): Corporate Surveillance in Everyday Life. How Companies Collect, Combine, Analyze, Trade, and Use Personal Data on Billions; Vienna: Cracked Labs.

Christl, Wolfie; Spiekermann, Sarah (2016): Networks of Control. A Report on Corporate Surveillance, Digital Tracking, Big Data & Privacy; Vienna: Facultas Verlags- und Buchhandels AG.

Citron, Danielle Keats (2016): (Un)Fairness of Risk Scores in Criminal Sentencing, Forbes, online article dated 13 July 2016, available at https://www.forbes.com/sites/daniellecitron/2016/07/13/unfairness-of-riskscores-in-criminal-sentencing/#4774e7c24479 (last retrieved on 28 August 2019).

Citron, Danielle Keats; Pasquale, Frank (2014): The scored society: due process for automated predictions; in: Washington Law Review, Vol. 89, No. 1, pp. 101-133.

Conger, Kate; Fausset, Richard; Kovaleski, Serge F. (2019): San Francisco Bans Facial Recognition Technology, New York Times, online article dated 14 May 2019, available at https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html (last retrieved on 28 August 2019).

Constantiou, Ioanna D.; Kallinikos, Jannis (2015): New games, new rules: big data and the changing context of strategy; in: Journal of Information Technology, Vol. 30, No. 1, pp. 44-57.

Cormen, Thomas H.; Leiserson, Charles E.; Rivest, Ronald; Stein, Clifford; Molitor, Paul (2010): Algorithmen – Eine Einführung, third edition; Munich: Oldenbourg Verlag.

Council of Europe (2019): Unboxing Artificial Intelligence: 10 steps to protect Human Rights; Straßbourg: Council of Europe, Commissioner for Human Rights.

Courtland, Rachel (2018): Bias detectives: the researchers striving to make algorithms fair. As machine learning infiltrates society, scientists are trying to help ward off injustice; in: Nature, Vol. 558, June 20, 2018, pp. 357-360.

Crawford, Kate (2013): The Hidden Biases in Big Data; in: Harvard Business Review, online article dated 1 April 2014, available at https://hbr.org/2013/04/the-hidden-biases-in-big-data?autocomplete=true (last retrieved on 29 August 2019).

Crawford, Kate; Whittaker, Meredith; Elish, Madeleine Clare; Barocas, Solon; Plasek, Aaron; Ferryman, Kadija (2016): The AI Now Report. The Social and Economic Implications of Artificial Intelligence, Technologies in the Near-Term. A summary of the AI Now public symposium, hosted by the White House and New York University's Information Law Institute, July 7th, 2016; New York: New York University, AI Now Institute.

Cummings, Mary (2004a): Automation bias in intelligent time critical decision support systems, AIAA 1st Intelligent Systems Technical Conference, S. Paper AIAA-6113.

Cummings, Mary L. (2004b): Creating moral buffers in weapon control interface design; in: IEEE Technology and Society Magazine, Vol. 23, No. 3, pp. 28-33.

Custers, Bart (2013): Data Dilemmas in the Information Society: Introduction and Overview; in: Bart Custers, Toon Calders, Bart Schermer and Tal Zarsky (eds.): Discrimination and Privacy in the Information Society; Studies in Applied Philosophy, Epistemology and Rational Ethics, Vol. 3; Berlin, Heidelberg: Springer, pp. 3-26.

Das, Sauvik; Kramer, Adam D. I. (2013): Self-Censorship on Facebook, Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media (ICWSM), pp. 120-127, published by Association for the Advancement of Artificial Intelligence.

Dastin, Jeffrey (2018): Amazon scraps secret AI recruiting tool that showed bias against women, Reuters Business News, online article published by Reuters dated 2018-10-10, available at https://www.reuters.com/article/us-amazon-com-jobsautomation-insight/amazon-scrapssecret-ai-recruiting-tool-that-showed-bias-against-women -idUSKCN1MK08G (last retrieved on 28 August 2019).

Datta, Amit; Datta, Anupam; Makagon, Jael; Mulligan, Deirdre. K.; Tschantz, Michael Carl (2018): Discrimination in Online Personalization: A Multidisciplinary Inquiry; in: Proceedings of Machine Learning Research, Vol. 81, pp. 1-15.

Datta, Amit; Tschantz, Michael Carl; Datta, Anupam (2015): Automated experiments on ad privacy settings; in: Proceedings on Privacy Enhancing Technologies, Vol. 2015, No. 1, pp. 92-112.

Däubler, Wolfgang (2018): AGG §1 Ziel des Gesetzes; in: Wolfgang Däubler and Martin Bertzbach (eds.): Allgemeines Gleichbehandlungsgesetz – Handkommentar, 4th edition; Baden-Baden: Nomos.

De-Arteaga, Maria; Romanov, Alexey; Wallach, Hanna; Chayes, Jennifer; Borgs, Christian; Chouldechova, Alexandra; Geyik, Sahin; Kenthapadi, Krishnaram; Kalai, Adam Tauman (2019): Bias in Bios: A Case Study of Semantic Representation Bias in a High-Stakes Setting; in: Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 120-128, published by ACM.

de Laat, Paul B. (2017): Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?; in: Philosophy & Technology, pp. 1-17.

DerStandard (2019): Volksanwaltschaft kritisiert AMS-Algorithmus, DerStandard, online article dated 10 March 2019, available at https://apps. derstandard.de/privacywall/story/2000099270837/volksanwaltschaftkritisiert-ams-algorithmus-in-der-krikik-der (last retrieved on 28 August 2019).

Desai, Deven R.; Kroll, Joshua A. (2017): Trust but Verify: A Guide to Algorithms and the Law; in: Harvard Journal of Law & Technology, Vol. 31, No. 1, p. 1.

Deutscher Ethikrat (2017): Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung, Stellungnahme; Berlin: Deutscher Ethikrat.

Dewenter, Ralf; Lüth, Hendrik (2018): Datenhandel und Plattformen; Expert report in the context of the "Assessing Big Data" (ABIDA) project; Münster, Karlsruhe: University of Münster, Karlsruhe Institute of Technology.

Diakopoulos, Nicholas (2014): Algorithmic Accountability Reporting: On the Investigation of Black Boxes; New York: Columbia University Academic Commons.

Dieterich, William; Mendoza, Christina; Brennan, Tim (2016): COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity. Performance of the COMPAS Risk Scales in Broward County. Technical report, July 2016: Northpointe Inc.

Dixon, Pam; Gellman, Robert (2014): The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future; Washington, D. C.: World Privacy Forum.

Domingos, Pedro (2012): A Few Useful Things to Know About Machine Learning; in: Communications of the ACM, Vol. 55, No. 10, pp. 78-87.

Dorfleitner, Gregor; Hornuf, Lars (2018): Neue digitale Akteure und ihre Rolle in der Finanzwirtschaft. Eine Analyse des deutschen Marktes unter besonderer Berücksichtigung von Datenschutzaspekten; Expert report in the context of the "Assessing Big Data" (ABIDA) project; Münster, Karlsruhe: University of Münster, Karlsruhe Institute of Technology.

Dosilovic, F. K.; Brcic, M.; Hlupic, N. (2018): Explainable artificial intelligence: A survey, 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2018 Proceedings, pp. 210-215.

Dwork, Cynthia; Mulligan, Deirdre K. (2013): It's Not Privacy, And It's Not Fair; in: Stanford Law Review Online, Vol. 66, pp. 35-40.

Dzida, Boris (2017): Big Data im Arbeitsrecht; in: Neue Zeitschrift für Arbeitsrecht (NZA), Vol. 34, No. 9, pp. 541-546.

Dzida, Boris; Groh, Naemi (2018): Diskriminierung nach dem AGG beim Einsatz von Algorithmen im Bewerbungsverfahren; in: Neue Juristische Wochenschrift (NJW), Vol. 71, No. 27, pp. 1917-1922.

Ebert, Ina (2019): Allgemeines Gleichbehandlungsgesetz (AGG), Reiner Schulze (ed.): Bürgerliches Gesetzbuch – Handkommentar, 10th edition; Baden-Baden: Nomos.

Eckhouse, Laurel; Lum, Kristian; Conti-Cook, Cynthia; Ciccolini, Julie (2019): Layers of Bias: A Unified Approach for Understanding Problems With Risk Assessment; in: Criminal Justice and Behavior, Vol. 46, No. 2, pp. 185-209.

Edelman, Benjamin G.; Luca, Michael (2014): Digital discrimination: The case of Airbnb.com, Harvard Business School NOM Unit Working Paper, no. 14-054; Boston: Harvard Business School.

Edelman, Benjamin; Luca, Michael; Svirsky, Dan (2017): Racial discrimination in the sharing economy: Evidence from a field experiment; in: American Economic Journal: Applied Economics, Vol. 9, No. 2, pp. 1-22.

EDPS (2017): Opinion 4/2017 on the Proposal for a Directive on certain aspects concerning contracts for the supply of digital content; Brussels: European Data Protection Supervisor (EDPS).

EDPS (2018): Guidelines on the protection of personal data in IT governance and IT management of EU institutions; Brussels: European Data Protection Supervisor (EDPS).

Edwards, Lilian; Veale, Michael (2017): Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for; in: Duke Law & Technology Review, Vol. 16, No. 1, pp. 18-84.

Ehsan, Upol; Tambwekar, Pradyumna; Chan, Larry; Harrison, Brent; Riedl, Mark (2019): Automated rationale generation: a technique for explainable AI and its effects on human perceptions; arXiv preprint arXiv:1901.03729.

Epp, Clayton; Lippold, Michael; Mandryk, Regan L. (2011): Identifying emotional states using keystroke dynamics, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 715-724, published by ACM.

Ernst, Christian (2017): Algorithmische Entscheidungsfindung und personenbezogene Daten; in: Juristenzeitung, Vol. 72, No. 21, pp. 1026-1036.

Eschholz, Stefanie (2017): Big Data-Scoring unter dem Einfluss der Datenschutz-Grundverordnung; in: Datenschutz und Datensicherheit – DuD, Vol. 41, No. 3, pp. 180-185.

Eubanks, Virginia (2017): Automating Inequality. How High-Tech Tools Profile, Police, and Punish the Poor; New York: St. Martin’s Press.

Ezrachi, Ariel; Stucke, Maurice E. (2016): Virtual Competition. The Promise and Perils of the Algorithm-Driven Economy; Cambridge, London: Harvard University Press.

Fang, Hanming; Moro, Andrea (2011): Theories of statistical discrimination and affirmative action: A survey; in: Jess Benhabib, Matthew O. Jackson and Alberto Bisin (eds.): Handbook of Social Economics, Vol. 1A; Amsterdam: North-Holland (Elsevier), pp. 133-200.

Fanta, Alexander (2018): Österreichs Jobcenter richten künftig mit Hilfe von Software über Arbeitslose, Netzpolitik.org, online article dated 18 October 2018, available at https://netzpolitik.org/2018/oesterreichsjobcenter-richten-kuenftig-mit-hilfe-von-software-ueber-arbeitslose/ (last retrieved on 28 August 2019).

Favaretto, Maddalena; De Clercq, Eva; Elger, Bernice Simone (2019): Big Data and discrimination: perils, promises and solutions. A systematic review; in: Journal of Big Data, Vol. 6, No. 1, pp. 1-27.

Ferguson, Andrew Guthrie (2017): The Truth About Predictive Policing and Race, The Appeal, online article dated 07 December 2017, available at https://theappeal.org/the-truth-about-predictive-policing-and-raceb87cf7c070b1/ (last retrieved on 28 August 2019).

Ferretti, Federico (2017): Not-So-Big and Big Credit Data Between Traditional Consumer Finance, FinTechs, and the Banking Union: Old and New Challenges in an Enduring EU Policy and Legal Conundrum, Global Jurist, Vol. 18, No. 1, 1-41.

Fink, Katherine (2018): Opening the government's black boxes: freedom of information and algorithmic accountability; in: Information, Communication & Society, Vol. 21, No. 10, pp. 1453-1471.

Forbrukerrådet (2018): Deceived by design. How tech companies use dark patterns to discourage us from exercising our rights to privacy; Oslo: Forbrukerrådet.

FRA (2018): #BigData: Discrimination in data-supported decision making, FRA Focus; Vienna: European Union Agency for Fundamental Rights (FRA).

FRA (2019): Data quality and artificial intelligence mitigating bias and error to protect fundamental rights, FRA Focus; Vienna: European Union Agency for Fundamental Rights (FRA).

Franke, Bernhard; Schlichtmann, Gisbert (2018): AGG §20 Zulässige unterschiedliche Behandlung; in: Wolfgang Däubler and Martin Bertzbach (eds.): Allgemeines Gleichbehandlungsgesetz – Handkommentar, 4th edition; Baden-Baden: Nomos.

Freeman, Katherine (2016): Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis; in: North Carolina Journal of Law & Technology, Vol. 18, No. 5, pp. 75-106.

Friedler, Sorelle A.; Scheidegger, Carlos; Venkatasubramanian, Suresh; Choudhary, Sonam; Hamilton, Evan P.; Roth, Derek (2019): A comparative study of fairness-enhancing interventions in machine learning, FAT* '19 Proceedings of the Conference on Fairness, Accountability, and Transparency, January 29 - 31, 2019, Atlanta, GA, USA, pp. 329-338, published by ACM.

Friedman, Batya; Nissenbaum, Helen (1996): Bias in Computer Systems; in: ACM Transactions on Information Systems, Vol. 14, No. 3, pp. 330-347.

Fröhlich, Wiebke; Spiecker genannt Döhmann, Indra (2018): Können Algorithmen diskriminieren?; in: Verfassungsblog (VerfBlog), online article dated 26 December 2018, available at https://verfassungsblog.de/koennenalgorithmen-diskriminieren/ (last retrieved on 27 August 2019).

FTC (2016): Big Data. A Tool for Inclusion or Exclusion?; Washington, D. C.: Federal Trade Commission (FTC).

Fu, S.; He, H.; Hou, Z. (2014): Learning Race from Face: A Survey; in: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 36, No. 12, pp. 2483-2509.

Galhotra, Sainyam; Brun, Yuriy; Meliou, Alexandra (2017): Fairness testing: testing software for discrimination, Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering, pp. 498-510, published by ACM.

Gandy Jr., Oscar (2009): Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage; Farnham, Burlington: Ashgate.

Gandy Jr., Oscar (2010): Engaging rational discrimination: exploring reasons for placing regulatory constraints on decision support systems; in: Ethics and Information Technology, Vol. 12, No. 1, pp. 1-14.

Garg, Nikhil; Schiebinger, Londa; Jurafsky, Dan; Zou, James (2018): Word embeddings quantify 100 years of gender and ethnic stereotypes; in: Proceedings of the National Academy of Sciences, Vol. 115, No. 16, pp. E3635-E3644.

Garvie, Clare; Bedoya, Alvaro M.; Frankle, Jonathan (2016): The perpetual line-up: Unregulated police face recognition in America; Washington, D. C.: Georgetown Law, Center on Privacy & Technology.

Géron, Aurélien (2018): Praxiseinstieg Machine Learning mit Scikit-Learn und TensorFlow: Konzepte, Tools und Techniken für intelligente Systeme (translated by Kristian Rother); Heidelberg: O'Reilly.

Gilheany, John; Wang, David; Xi, Stephen (2015): The model minority? Not on Airbnb.com: A hedonic pricing model to quantify racial bias against Asian Americans, Technology Science, online article dated 1 September 2015, available at https://techscience.org/a/2015090104/ (last retrieved on 28 August 2019).

Gillum, Jack; Tobin, Ariana (2019): Facebook Won't Let Employers, Landlords or Lenders Discriminate in Ads Anymore, ProPublica, online article dated 19 March 2019, available at https://www.propublica.org/article/facebook-ads-discrimination-settlement-housing-employment-credit (last retrieved on 28 August 2019).

Goldfarb, Avi; Tucker, Catherine (2017): Digital Economics, NBER Working Paper 23684; Cambridge, MA: National Bureau of Economic Research.

Goodman, Bryce; Flaxman, Seth (2017): European Union regulations on algorithmic decision-making and a "right to explanation"; in: AI Magazine, Vol. 38, No. 3, pp. 50-57.

Goodman, Bryce W. (2016): Economic Models of (Algorithmic) Discrimination, Machine Learning and the Law, NIPS Symposium, 8 December, 2016, Barcelona, Spain.

Grimmelmann, James (2005): Regulation by Software; in: Yale Law Journal, Vol. 114, No. 7, pp. 1721-1758.

Guidotti, R.; Monreale, A.; Ruggieri, S.; Turini, F.; Giannotti, F.; Pedreschi, D. (2018): A survey of methods for explaining black box models; in: ACM Computing Surveys, Vol. 51, No. 5, pp. 1-42.

Gurovich, Yaron; Hanani, Yair; Bar, Omri; Nadav, Guy; Fleischer, Nicole; Gelbman, Dekel; Basel-Salmon, Lina; Krawitz, Peter M.; Kamphausen, Susanne B.; Zenker, Martin (2019): Identifying facial phenotypes of genetic disorders using deep learning; in: Nature medicine, Vol. 25, No. 1, p. 60.

Hacker, Philipp; Petkova, Bilyana (2017): Reining in the Big Promise of Big Data: Transparency, Inequality, and New Regulatory Frontiers; in: Northwestern Journal of Technology and Intellectual Property, Vol. 15, No. 1.

Hand, David J. (2006): Classifier Technology and the Illusion of Progress; in: Statistical Science, Vol. 21, No. 1, pp. 1-14.

Hannák, Anikó.; Soeller, Gary; Lazer, David; Mislove, Alan; Wilson, Christo (2014): Measuring price discrimination and steering on e-commerce web sites; in: Proceedings of the 2014 conference on internet measurement conference, pp. 305-318, published by ACM.

Hannák, Anikó.; Wagner, Claudia; Garcia, David; Mislove, Alan; Strohmaier, Markus; Wilson, Christo (2017): Bias in online freelance marketplaces: Evidence from taskrabbit and fiverr; in: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, pp. 1914-1933, published by ACM.

Hannák, Anikó.; Wagner, Claudia; Garcia, David; Strohmaier, Markus; Wilson, Christo (2016): Bias in online freelance marketplaces: Evidence from taskrabbit;; in: Proceedings DAT Workshop, pp. 1914-1933, published by ACM.

Hänold, Stefanie (2018): Profiling and Automated Decision-Making: Legal Implications and Shortcomings; in: Ugo Pagallo, Marcelo Corrales, Mark Fenwick and Nikolaus Forgó (eds.): Robotics, AI and the Future of Law; Singapore: Springer Singapore, pp. 123-153.

Hänold, Stefanie (2019): Profiling und automatisierte Einzelentscheidungen im Versicherungsbereich. report in the context of the "Assessing Big Data" (ABIDA) procect; Hannover: Institutionelles Repositorium der Leibniz Universität Hannover.

Hargittai, Eszter (2015): Is bigger always better? Potential biases of big data derived from social network sites; in: The ANNALS of the American Academy of Political and Social Science, Vol. 659, No. 1, pp. 63-76.

Härtel, Ines (2019): Digitalisierung im Lichte des Verfassungsrechts – Algorithmen, Predictive Policing, autonomes Fahren; in: Landes- und Kommunalverwaltung, Vol. 29, No. 2, pp. 49-60.

Helberger, Natali (2016): Profiling and Targeting Consumers in the Internet of Things A new Challenge for Consumer Law; in: Reiner Schulze and Dirk Staudenmayer (eds.): Digital Revolution: Challenges for Contract Law in Practice; Baden-Baden: Nomos, pp. 135-165.

Hellman, Deborah (1998): Two Types of Discrimination: The Familiar and the Forgotten; in: California Law Review, Vol. 86, No. 2, pp. 315-361.

Hellman, Deborah (2008): When is Discrimination Wrong?; Cambridge, London: Harvard University Press.

Hildebrandt, Mireille (2018): Algorithmic regulation and the rule of law; in: Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, Vol. 376, No. 2128, p. 20170355.

Hildebrandt, Mireille; Gutwirth, Serge (eds.) (2008): Profiling the European Citizen. Cross-Disciplinary Perspectives; Dordrecht: Springer.

Hill, Robin K. (2016): What an algorithm is; in: Philosophy & Technology, Vol. 29, No. 1, pp. 35-59.

Hinz, Thomas; Ausprung, Katrin (2017): Diskriminierung auf dem Wohnungsmarkt; in: Albert Scherr, Aladin El-Mafaalani and Gökçen Yüksel (eds.): Handbuch Diskriminierung; Wiesbaden: Springer VS, pp. 387-406.

Hoeren, Thomas; Kolany-Raiser, Barbara (eds.) (2018): Big Data in Context. Legal, Social and Technological Insights, Springer Briefs in Law; Cham: Springer Open.

Hoeren, Thomas; Niehoff, Maurice (2018): KI und Datenschutz Begründungserfordernisse automatisierter Entscheidungen; in: Rechtswissenschaft (RW) – Zeitschrift für rechtswissenschaftliche Forschung, Vol. 9, No. 1, pp. 47-66.

Hoffmann-Riem, Wolfgang (1998): Informationelle Selbstbestimmung in der Informationsgesellschaft Auf dem Wege zu einem neuen Konzept des Datenschutzes; in: Archiv des Öffentlichen Rechts, Vol. 123, No. 4, pp. 513-540.

Hoffmann-Riem, Wolfgang (2017): Verhaltenssteuerung durch Algorithmen Eine Herausforderung für das Recht; in: Archiv des öffentlichen Rechts, Vol. 142, No. 1, pp. 1-42.

Holl, Jürgen; Kernbeiß, Günter; Wagner-Pinter, Michael (2018): Das AMS-Arbeitsmarktchancen-Modell. Dokumentation zur Methode; Vienna: Synthesis Forschung.

Hurley, Mikella; Adebayo, Julius (2016): Credit Scoring in the Era of Big Data; in: Yale Journal of Law & Technology, Vol. 18, No. 1, pp. 148-216.

Illinois Attorney General (2017): Madigan Probes National Job Search Sites Over Potential Age Discrimination. Attorney General Madigan Calls on Career Search Sites to Explain Potential Age Discrimination Violations Against Older Job Seekers, online article dated 02 March 2017, available at http://www.illinoisattorneygeneral.gov/pressroom/2017_03/20170302. html (last retrieved on 28 August 2019).

Ingold, David; Soper, Spencer (2016): Amazon Doesn't Consider the Race of Its Customers. Should It?, Bloomberg, online article dated 21 April 2016, updated 01 May 2016, available at https://www.bloomberg.com/graphics/2016amazon-same-day/ (last retrieved on 28 August 2019).

Isaac, William; Lum, Kristian (2018): Setting the Record Straight on Predictive Policing and Race, online article published by The Appeal dated 3 January 2018, available at https://the appeal.org/setting-the-recordstraight-on-predictive-policing-and-race-fe588b457ca2/ (last retrieved on 28 August 2019).

Jernigan, Carter; Mistree, Behram F.T. (2009): Gaydar: Facebook friendships expose sexual orientation; in: First Monday, Vol. 14, No. 10.

Jordan, M. I.; Mitchell, T. M. (2015): Machine learning: Trends, perspectives, and prospects; in: Science, Vol. 349, No. 6245, pp. 255-260.

Just, Natascha; Latzer, Michael (2016): Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet; in: Media, Culture & Society, Vol. 39, No. 2, pp. 238-258.

Kallinikos, Jannis (2011): Governing through Technology: Information Artefacts and Social Practice; Basingstoke: Palgrave Macmillan.

Kamp, Meike; Rost, Martin (2013): Kritik an der Einwilligung. Ein Zwischenruf zu einer fiktiven Rechtsgrundlage in asymmetrischen Machtverhältnissen; in: Datenschutz und Datensicherheit, Vol. 37, No. 2, pp. 80-84.

Kamp, Meike; Weichert, Thilo (2005): Scoringsysteme zur Beurteilung der Kreditwürdigkeit Chancen und Risiken für Verbraucher; Kiel: Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein (ULD).

Kant, Immanuel (1786/1977): Grundlegung zur Metaphysik der Sitten. In: Immanuel Kant: Werke in zwölf Bänden. Vol. 7; initial print: Riga (Hartknoch) 1785. The text is printed in the 2nd (revised) edition, Riga (Hartknoch) 1786; Frankfurt am Main: public domain edition via Zeno.org, http://www.zeno.org/nid/20009189599 (last retrieved on 29 August 2019).

Kasperkevic, Jana (2015): Google says sorry for racist auto-tag in photo app; in: The Guardian, online article dated 1 July 2015, available at https://www.theguardian.com/technology/2015/jul/01/google-sorry-racist-auto-tag-photo-app (last retrieved on 28 August 2019).

Kay, Matthew; Matuszek, Cynthia; Munson, Sean A. (2015): Unequal representation and gender stereotypes in image search results for occupations; in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 3819-3828, published by ACM.

Kettner, Elisa; Thorun, Christian; Kleinhans, Jan-Peter (2018): Big Data im Bereich Heim und Freizeit. Smart Living: Status Quo und Entwicklungstendenzen; Expert report in the context of the "Assessing Big Data" (ABIDA) project; Münster, Karlsruhe: University of Münster, Karlsruhe Institute of Technology.

Kim, Pauline T. (2016): Data-driven discrimination at work; in: William & Mary Law Review, Vol. 58, No. 3, pp. 857-936.

Kiritchenko, Svetlana; Mohammad, Saif M. (2018): Examining gender and race bias in two hundred sentiment analysis systems; arXiv preprint arXiv:1805.04508.

Kitchin, Rob (2017): Thinking critically about and researching algorithms; in: Information, Communication & Society, Vol. 20, No. 1, pp. 14-29.

Klare, Brendan F.; Burge, Mark J.; Klontz, Joshua C.; Bruegge, Richard W. Vorder; Jain, Anil K. (2012): Face recognition performance: Role of demographic information; in: IEEE Transactions on Information Forensics and Security, Vol. 7, No. 6, pp. 1789-1801.

Klebert, Florian; Shirazi, Fatemeh; Simo, Hervais; Wüchner, Tobias; Buchmann, Johannes; Pretschner, Alexander; Waidner, Michael (2012): State of Online Privacy: A Technical Perspective; in: Johannes Buchmann (ed.): Internet Privacy. Eine multidisziplinäre Bestandsaufnahme / A multidisciplinary analysis (acatech STUDIE); Heidelberg u. a.: Springer Verlag, pp. 189-279.

Kleinberg, Jon; Ludwig, Jens; Mullainathan, Sendhil; Sunstein, Cass R. (2019): Discrimination in the Age of Algorithms; Cambridge, MA: National Bureau of Economic Research.

Klingele, Cecelia (2015): The promises and perils of evidence-based corrections; in: Notre Dame Law Review, Vol. 91, No. 2, pp. 101-151.

Kolany-Raiser, Barbara; Heil, Reinhard; Orwat, Carsten; Hoeren, Thomas (eds.) (2018): Big Data und Gesellschaft. Eine multidisziplinäre Annäherung; Wiesbaden: Springer VS.

Kornwachs, Klaus (2018): Arbeit 4.0 – People Analytics. Führungsinformationssysteme: Soziologische, psychologische, wissenschaftsphilosophischethische Überlegungen zum Einsatz von Big Data in Personalmanagement und Personalführung; Expert report in the context of the "Assessing Big Data" (ABIDA) project; Argenbühl-Eglofs: Büro für Kultur und Technik.

Kosinski, Michal; Stillwell, David; Graepel, Thore (2013): Private traits and attributes are predictable from digital records of human behavior; in: Proceedings of the National Academy of Sciences, Vol. 110, No. 15, pp. 5802-5805.

Kroll, Joshua A. (2018): The fallacy of inscrutability; in: Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, Vol. 376, No. 2133, pp. 1-14.

Lambrecht, Anja; Tucker, Catherine E. (2019): Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads, Management Science (Articles in Adance).

Lang, Caroline; Barton, Hannah (2015): Just untag it: Exploring the management of undesirable Facebook photos; in: Computers in Human Behavior, Vol. 43, Feb., pp. 147-155.

Lecuyer, Mathias; Spahn, Riley; Spiliopolous, Yannis; Chaintreau, Augustin; Geambasu, Roxana; Hsu, Daniel (2015): Sunlight: Fine-grained targeting detection at scale with statistical confidence; in: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, pp. 554-566, published by ACM.

Lehr, David; Ohm, Paul (2017): Playing with the Data: What Legal Scholars Should Learn About Machine Learning; in: UC Davis Law Review, Vol. 51, pp. 653-717.

Leis, Miriam; Petzka, Henning; Rüping, Stefan; Voss, Angelika (2018): Maschinelles Lernen Einordnung, Konzepte, Methoden und Grenzen, Inga Döbel, Miriam Leis, Manuel Molina Vogelsang, Dmitry Neustroev, Henning Petzka, Stefan Rüping, Angelika Voss, Martin Wegele and Juliane Welz (eds.): Maschinelles Lernen Kompetenzen, Anwendungen und Forschungsbedarf: Fraunhofer IAIS, Fraunhofer IMW, Fraunhofer Zentrale, pp. 7-52.

Lerman, Jonas (2013): Big data and its exclusions; in: Stanford Law Review Online, Vol. 66, No. Sep., pp. 55-63.

Lessig, Lawrence (1999): Code and other laws of cyberspace; New York: Basic Books.

Lessig, Lawrence (2006): Code Version 2.0; New York: Basic Books.

Levy, Karen; Barocas, Solon (2017): Designing against discrimination in online markets; in: Berkeley Technolgy Law Journal, Vol. 32, No. 3, pp. 1183-1237.

Linoff, Gordon S.; Berry, Michael J. A. (2011): Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management; Indianapolis: Wiley.

Lippert-Rasmussen, Kasper (2007): Nothing personal: On statistical discrimination; in: Journal of Political Philosophy, Vol. 15, No. 4, pp. 385-403.

Lischka, Konrad; Klingel, Anita (2017): Wenn Maschinen Menschen bewerten. Internationale Fallbeispiele für Prozesse algorithmischer Entscheidungsfindung – Working paper –; Gütersloh: Bertelsmann Stiftung.

Lorenz, Wilhelm (1993): Diskriminierung; in: Bernd-Thomas Ramb and Manfred Tietzel (eds.): Ökonomische Verhaltenstheorie; Munich: Vahlen, pp. 119-147.

Lum, Kristian; Isaac, William (2016): To predict and serve?; in: Significance, Vol. 13, No. 5, pp. 14-19.

Mantelero, Alessandro (2016): Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection; in: Computer Law and Security Review, Vol. 32, No. 2, pp. 238-255.

Marder, Ben; Joinson, Adam; Shankar, Avi; Houghton, David (2016): The extended 'chilling' effect of Facebook: The cold reality of ubiquitous social networking; in: Computers in Human Behavior, Vol. 60, Jul., pp. 582-592.

Marler, Janet H.; Boudreau, John W. (2017): An evidence-based review of HR Analytics; in: The International Journal of Human Resource Management, Vol. 28, No. 1, pp. 3-26.

Marthews, Alex; Tucker, Catherine E. (2017): Government Surveillance and Internet Search Behavior; Cambridge, MA: Digital Fourth and MIT Sloan School of Management.

Martin, Kirsten (2013): Transaction costs, privacy, and trust: The laudable goals and ultimate failure of notice and choice to respect privacy online; in: First Monday, Vol. 18, No. 12, online article dated 2 December 2013, available at https://firstmonday.org/ojs/index.php/fm/article/view/4838/3802 (last retrieved on 27 August 2019).

Martini, Mario (2017): Algorithmen als Herausforderung für die Rechtsordnung; in: Juristenzeitung, Vol. 72, No. 21, pp. 1017-1025.

Martini, Mario (2018): DS-GVO Art. 22 Automatisierte Entscheidungen im Einzelfall einschließlich Profiling; in: Boris P. Paal and Daniel A. Pauly (eds.): Datenschutz-Grundverordnung Bundesdatenschutzgesetz DSGVO BDSG, commentary, 2nd edition; Munich: Beck.

Martini, Mario; Nink, David (2017): Wenn Maschinen entscheiden ... – vollautomatisierte Verwaltungsverfahren und der Persönlichkeitsschutz; in: Neue Zeitschrift für Verwaltungsrecht Extra, Vol. 36, No. 10, pp. 1-14.

Mathur, Arunesh; Acar, Gunes; Friedman, Michael; Lucherini, Elena; Mayer, Jonathan; Chetty, Marshini; Narayanan, Arvind (2019): Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites; Princeton: Princeton University.

Matsakis, Louise (2019): Facebook's Ad System Might Be Hard-Coded for Discrimination; in: Wired, online article dated 6 April 2019, available at https://www.wired.com/story/facebooks-ad-systemdiscrimination/ (last retrieved on 28 August 2019).

Matz, S. C.; Kosinski, M.; Nave, G.; Stillwell, D. J. (2017): Psychological targeting as an effective approach to digital mass persuasion; in: Proceedings of the National Academy of Sciences (PNAS), Vol. 114, No. 48, pp. 12714-12719.

Matz, Sandra C.; Netzer, Oded (2017): Using Big Data as a window into consumers' psychology; in: Current Opinion in Behavioral Sciences, Vol. 18, No. Dec., pp. 7-12.

Matzat, Lorenz; Zielinski, Lukas; Cocco, Miriam; Penner, Kristina; Spielkamp, Matthias; Gießler, Sebastian; Lang, Sebastian; Thiel, Veronika (2019): Atlas der Automatisierung: Automatisierung und Teilhabe in Deutschland; Berlin: AW AlgorithmWatch gGmbH.

Mayer, Jonathan; Mutchler, Patrick; Mitchell, John C. (2016): Evaluating the privacy properties of telephone metadata; in: Proceedings of the National Academy of Sciences, Vol. 113, No. 20, pp. 5536-5541.

McDonald, Aleecia M.; Cranor, Lorrie Faith (2008): The cost of reading privacy policies; in: I/S: A Journal of Law and Policy for the Information Society, Vol. 4, No. 3, pp. 543-568.

Medina, E. (2015): Rethinking algorithmic regulation; in: Kybernetes, Vol. 44, No. 6-7, pp. 1005-1019.

Merz, Christina (2016): Predictive Policing Polizeiliche Strafverfolgung in Zeiten von Big Data; Dossier of the "Assessing Big Data" (ABIDA) project; Karlsruhe: Institute for Technology Assessment and Systems Analysis (ITAS), Karlsruhe Institute of Technology.

Meyer, Robinson (2015): Could a Bank Deny Your Loan Based on Your Facebook Friends?; in: The Atlantic, online article dated 25 September 2015, available at https://www.theatlantic.com/technology/archive/2015/09/facebooks-new-patent-and-digital-redlining/407287/ (last retrieved on 28 August 2019).

Mikians, Jakub; Gyarmati, László; Erramilli, Vijay; Laoutaris, Nikolaos (2012): Detecting price and search discrimination on the internet; in: Proceedings of the 11th ACM Workshop on Hot Topics in Networks, pp. 79-84, published by ACM.

Mikians, Jakub; Gyarmati, László; Erramilli, Vijay; Laoutaris, Nikolaos (2013): Crowd-assisted search for price discrimination in e-commerce: First results; in: Proceedings of the ninth ACM conference on Emerging networking experiments and technologies, pp. 1-6, published by ACM.

Miller, Akiva A. (2014): What Do We Worry About When We Worry About Price Discrimination? The Law and Ethics of Using Personal Information for Pricing; in: Journal of Technology Law & Policy, Vol. 19, No. 1, pp. 41-104.

Milne, George R.; Culnan, Mary J. (2004): Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices; in: Journal of Interactive Marketing, Vol. 18, No. 3, pp. 15-29.

Mittelstadt, Brent Daniel; Allo, Patrick; Taddeo, Mariarosaria; Wachter, Sandra; Floridi, Luciano (2016): The ethics of algorithms: Mapping the debate; in: Big Data & Society, Vol. 3, No. 2, pp. 1-21.

Moll, Ricarda; Horn, Marco; Scheibel, Lisa; Rusch-Rodosthenous, Miriam (2018): Soziale Medien und die EU-Datenschutzgrundverordnung. Informationspflichten und datenschutzfreundliche Voreinstellungen; Düsseldorf: Marktwächter Digitale Welt, Verbraucherzentrale NRW e. V.

Moos, Flemming; Rothkegel, Tobias (2016): Nutzung von Scoring-Diensten im Online-Versandhandel. Scoring-Verfahren im Spannungsfeld von BDSG, AGG und DS-GVO; in: Zeitschrift für Datenschutz, Vol. 6, No. 12, pp. 561-568.

Mullainathan, Sendhil; Spiess, Jann (2017): Machine learning: an applied econometric approach; in: Journal of Economic Perspectives, Vol. 31, No. 2, pp. 87-106.

NFHA (2019): National Fair Housing Alliance Settles Lawsuit with Facebook: Transforms Facebook's Ad Platform Impacting Millions of Users, online article published by National Fair Housing Alliance dated 18 March 2019, available at https://nationalfairhousing.org/2019/03/18/ national-fair-housing-alliance-settles-lawsuit-with-facebook-transformsfacebooks-ad-platform-impacting-millions-of-users/ (last retrieved on 28 August 2019).

Niklas, Jędrzej (2018): Profiling the Unemployed, Digital Society Blog, Humboldt Institut für Internet und Gesellschaft, online article dated 16 January 2018, available at https://www.hiig.de/profiling-von-arbeitslosen/ (last retrieved on 28 August 2019).

Niklas, Jędrzej (2019): Polen: Regierung schafft umstrittenes Scoring-System für Arbeitslose ab, online article published by AlgorithmWatch dated 16 April 2019, available at: https://algorithmwatch.org/story/polnische-regierung-schafft-umstrittenes-scoring-system-fuer-arbeitslose-ab (last retrieved on 28 August 2019).

Niklas, Jędrzej; Sztandar-Sztanderska, Karolina; Szymielewicz, Katarzyna (2015): Profiling the unemployed in Poland: Social and political implications of algorithmic decision making; Warsaw: Fundacja Panoptykon.

Noble, Safiya Umoja (2018): Algorithms of Oppression. How Search Engines Reinforce Racism; New York: New York University Press.

Obermeyer, Ziad; Mullainathan, Sendhil (2019): Dissecting Racial Bias in an Algorithm that Guides Health Decisions for 70 Million People, Proceedings of the Conference on Fairness, Accountability, and Transparency, p. 89, published by ACM (1 September 2019).

Orwat, Carsten; Bless, Roland (2016): Values and Networks – Steps Toward Exploring their Relationships; in: Computer Communication Review (ACM SIGCOMM), Vol. 46, No. 2, pp. 25-31.

Orwat, Carsten; Raabe, Oliver; Buchmann, Erik; Anandasivam, Arun; Freytag, Johan-Christoph; Helberger, Natali; Ishii, Kei; Lutterbeck, Bernd; Neumann, Dirk; Otter, Thomas; Pallas, Frank; Reussner, Ralf; Sester, Peter; Weber, Karsten; Werle, Raymund (2010): Software als Institution und ihre Gestaltbarkeit; in: Informatik-Spektrum, Vol. 33, No. 6, pp. 626-633.

Orwat, Carsten; Schankin, Andrea (2018): Attitudes towards big data practices and the institutional framework of privacy and data protection – A population survey, KIT Scientific Report 7753; Karlsruhe: KIT Scientific Publishing.

Parasuraman, Raja; Riley, Victor (1997): Humans and automation: Use, misuse, disuse, abuse; in: Human factors, Vol. 39, No. 2, pp. 230-253.

Pasquale, Frank (2015): The Black Box Society: the Secret Algorithms that Control Money and Information; Cambridge, London: Harvard University Press.

Penney, Jonathon W. (2016): Chilling Effects: Online Surveillance and Wikipedia Use; in: Berkeley Technology Law Journal, Vol. 31, No. 1, pp. 117-182.

Penney, Jonathon W. (2017): Internet surveillance, regulation, and chilling effects online: a comparative case study; in: Internet Policy Review, Vol. 6, No. 2, pp. 1-39.

Phelps, Edmund S. (1972): The Statistical Theory of Racism and Sexism; in: American Economic Review, Vol. 62, No. 4, pp. 659-661.

Ponti, Sarah; Tuchtfeld, Erik (2018): Zur Notwendigkeit einer Verbandsklage im AGG; in: Zeitschrift für Rechtspolitik (ZRP), Vol. 51, No. 5, pp. 139-141.

Pope, Devin G.; Sydnor, Justin R. (2011): What's in a Picture? Evidence of Discrimination from Prosper.com; in: Journal of Human Resources, Vol. 46, No. 1, pp. 53-92.

Powles, Julia; Nissenbaum, Helen (2018): The Seductive Diversion of 'Solving' Bias in Artificial Intelligence; in: Medium, online article dated 7 December 2018, available at https://medium.com/s/story/the-seductive-diversionof-solving-bias-in-artificial-intelligence-890df5e5ef53 (last retrieved on 28 August 2019).

Puri, Ruchir (2018): Mitigating Bias in AI Models; in: IBM Research Blog, online article dated 6 February 2018, available at https://www.ibm.com/blogs/ research/2018/02/mitigating-bias-ai-models/ (last retrieved on 28 August 2019).

Raabe, Oliver; Wagner, Manuela (2016): Die Zweckbindung: Ein Überblick über die aktuelle Rechtslage und Harmonisierung durch die EU-Datenschutzgrundverordnung; in: Smart-Data-Begleitforschung (ed.): Die Zukunft des Datenschutzes im Kontext von Forschung und Smart Data. Datenschutzgrundprinzipien im Diskurs; Berlin: Smart-Data-Begleitforschung, pp. 16-22.

Raabe, Oliver; Wagner, Manuela (2019 forthcoming): Daten, Informationen, Wissen, Entscheiden und Steuern – Ein Referenzmodell für den zukunftsfähigen Datenschutz in wissensbasiert steuernden digitalen Ökosystemen; Karlsruhe: Karlsruhe Institute of Technology, Zentrum für Angewandte Rechtswissenschaft (ZAR).

Raji, Inioluwa Deborah; Buolamwini, Joy (2019): Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products; in: AAAI/ACM Conf. on AI Ethics and Society.

Reichwald, Julian; Pfisterer, Dennis (2016): Autonomie und Intelligenz im Internet der Dinge. Möglichkeiten und Grenzen autonomer Handlungen; in: Computer und Recht, Vol. 32, No. 3, pp. 208-212.

Reidenberg, Joel R. (1998): Lex Informatica: The Formulation of Information Policy Rules Through Technology; in: Texas Law Review, Vol. 76, No. 3, pp. 553-584.

Reidenberg, Joel R.; Bhatia, Jaspreet; Breaux, Travis D.; Norton, Thomas B. (2016): Ambiguity in privacy policies and the impact of regulation; in: The Journal of Legal Studies, Vol. 45, No. S2, pp. S163-S190.

Reidenberg, Joel R.; Russell, N. Cameron; Callen, Alexander J.; Qasir, Sophia; Norton, Thomas B. (2015): Privacy harms and the effectiveness of the notice and choice framework; in: I/S: A Journal of Law and Policy for the Information Society, Vol. 11, No. 2, pp. 485-524.

Reisman, Dillon; Schultz, Jason; Crawford, Kate; Whittaker, Meredith (2018): Algorithmic impact assessments: a practical framework for public agency accountability; New York: AI Now Institute.

Richardson, Rashida; Schultz, Jason; Crawford, Kate (2019): Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice; in: New York University Law Review Online.

Robinson, David; Koepke, Logan (2016): Stuck in a Pattern. Early evidence on "predictive policing" and civil rights; Washington, D.C.: Upturn.

Romei, Andrea; Ruggieri, Salvatore (2014): A multidisciplinary survey on discrimination analysis; in: The Knowledge Engineering Review, Vol. 29, No. 5, pp. 582-638.

Romei, Andrea; Ruggieri, Salvatore; Turini, Franco (2013): Discrimination discovery in scientific project evaluation: A case study; in: Expert Systems With Applications, Vol. 40, No. 15, pp. 6064-6079.

Rosenblat, Alex; Kneese, Tamara; Boyd, Danah (2014): Networked Employment Discrimination, Open Society Foundations' Future of Work Commissioned Research Papers 2014; New York: Data & Society Research Institute.

Rosenblat, Alex; Levy, Karen E. C.; Barocas, Solon; Hwang, Tim (2017): Discriminating Tastes: Uber's Customer Ratings as Vehicles for Workplace Discrimination; in: Policy & Internet, Vol. 9, No. 3, pp. 256-279.

Ru, Hong; Schoar, Antoinette (2016): Do credit card companies screen for behavioral biases?; Washington, D.C.: National Bureau of Economic Research.

Salzburger Nachrichten (2019): Kopf sieht keine Diskriminierung durch AMS-Algorithmus; in: Salzburger Nachrichten, online article dated 18 January 2019, available at https://www.sn.at/wirtschaft/oesterreich/kopf-sieht-keine-diskriminierung-durch-ams-algorithmus-64301929 (last retrieved on 28 August 2019).

Sanchez-Monedero, Javier; Dencik, Lina (2018): How to (partially) evaluate automated decision systems; Cardiff: Data Justice Lab.

Sandvig, Christian; Hamilton, Kevin; Karahalios, Karrie; Langbort, Cedric (2014): Auditing algorithms: Research methods for detecting discrimination on internet platforms; in: Data and discrimination: converting critical concerns into productive inquiry; a preconference at the 64th Annual Meeting of the International Communication Association. May 22, 2014; Seattle, WA, USA.

Schaber, Peter (2012): Menschenwürde; Stuttgart: Reclam.

Schauer, Frederick (2003): Profiles, probabilities, and stereotypes; Cambridge: Harvard University Press.

Schauer, Frederick (2018): Statistical (and non-statistical) discrimination; in: Kasper Lippert-Rasmussen (ed.): The Routledge Handbook of the Ethics of Discrimination; London: Routlegde, pp. 42-53.

Scherr, Albert (2016): Diskriminierung/Antidiskriminierung – Begriffe und Grundlagen; in: Aus Politik und Zeitgeschichte, Vol. 66, No. 9, pp. 3-10.

Schiek, Dagmar (2000): Differenzierte Gerechtigkeit: Diskriminierungsschutz und Vertragsrecht; Baden-Baden: Nomos.

Schinzel, Britta (2017): Algorithmen sind nicht schuld, aber wer oder was ist es dann?; in: FIfF-Kommunikation, No. 2, pp. 5-9.

Schneider, Ingrid; Ulbricht, Lena (2018): Ist Big Data fair? Normativ hergestellte Erwartungen an Big Data; in: Barbara Kolany-Raiser, Reinhard Heil, Carsten Orwat and Thomas Hoeren (eds.): Big Data und Gesellschaft. Eine multidisziplinäre Annäherung; Wiesbaden: Springer VS, pp. 198-207.

Scholz, Philip (2019): DSGVO Art. 22 Automatisierte Entscheidungen im Einzelfall einschließlich Profiling; in: Spiros Simitis, Gerrit Hornung and Indra Spiecker aka Döhmann (eds.): Datenschutzrecht. DSGVO und BDSG; Baden-Baden: Nomos.

Schrader, Peter; Schubert, Jens (2018): AGG §3 Begriffsbestimmungen; in: Wolfgang Däubler and Martin Bertzbach (eds.): Allgemeines Gleichbehandlungsgesetz – Handkommentar, 4th edition; Baden-Baden: Nomos.

Schwaiger, Manfred; Hufnagel, Gerrit (2018): Handel und elektronische Bezahlsysteme; Expert report in the context of the "Assessing Big Data" (ABIDA) project; Munich: Ludwig-Maximilians-Universität Munich, Institute for Market-based Management.

Schwartz, Paul M. (1999): Privacy and democracy in cyberspace; in: Vanderbilt Law Review, Vol. 52, No. 6, pp. 1607-1702.

Schweighofer, Erich; Sorge, Christoph; Borges, Georg; Schäfer, Burkhard; Waltl, Bernhard; Grabmair, Matthias; Krupka, Daniel (2018): Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren. Gutachten der Fachgruppe Rechtsinformatik der Gesellschaft für Informatik e. V. im Auftrag des Sachverständigenrats für Verbraucherfragen; Berlin: Advisory Council for Consumer Affairs (SVRV) at the Federal Ministry of Justice and Consumer Protection.

Selbst, Andrew D. (2017): Disparate Impact in Big Data Policing; in: Georgia Law Review, Vol. 52, No. 1, pp. 109-195.

Selbst, Andrew D.; Barocas, Solon (2018): The Intuitive Appeal of Explainable Machines; in: Fordham Law Review, Vol. 87, No. 3, pp. 1085-1139.

Selke, Stefan; Biniok, Peter; Achatz, Johannes; Späth, Elisabeth (2018): Ethische Standards für Big Data und deren Begründung; Expert report in the context of the "Assessing Big Data" (ABIDA) project; Münster, Karlsruhe: University of Münster, Karlsruhe Institute of Technology.

Shah, Rajiv C.; Kesan, Jay P. (2010): Software as Governance; in: Hans J. Scholl (ed.): E-Government Information, Technology, and Transformation; Armonk: M.E. Sharpe, pp. 125-140.

Sherwin, Galen; Bhandari, Esha (2019): Facebook Settles Civil Rights Cases by Making Sweeping Changes to Its Online Ad Platform, online article published by American Civil Liberties Union (ACLU) dated 19 March 2019, availale at https://www.aclu.org/blog/womens-rights/womens-rightsworkplace/facebook-settles-civil-rights-cases-making-sweeping (last retrieved on 28 August 2019).

Silver, Joe (2013): Is Your Turn-By-Turn Navigation Application Racist?, online article published by American Civil Liberties Union (ACLU) dated 2 October 2013, available at https://www.aclu.org/blog/national-security/your-turn-turn-navigation-application-racist (last retrieved on 27 August 2019).

Smeddinck, Ulrich; Bornemann, Basil (2018): Verkehr, Mobilität, Nudging Zugleich zum Stand von Regulieren durch Anstoßen in Deutschland; in: Die Öffentliche Verwaltung, Vol. 71, No. 13, pp. 513-523.

Snow, Jacob (2018): Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots, online article published by American Civil Liberties Union (ACLU) dated 26 July 2018, available at https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazonsface-recognition-falsely-matched-28 (last retrieved on 28 August 2019).

Solove, Daniel J. (2013): Privacy Self-Management and the Consent Dilemma; in: Harvard Law Review, Vol. 126, No. 7, pp. 1880-1903.

Spielkamp, Mathias (ed.) (2019): Automating Society. Taking Stock of Automated Decision-Making in the EU. A report by AlgorithmWatch in cooperation with Bertelsmann Stiftung, supported by the Open Society Foundations; Berlin: AW AlgorithmWatch gGmbH.

Starr, Sonja B. (2014): Evidence-based sentencing and the scientific rationalization of discrimination; in: Stanford Law Review, Vol. 66, No. 4, pp. 803-872.

Steppe, Richard (2017): Online price discrimination and personal data: A General Data Protection Regulation perspective; in: Computer Law & Security Review, Vol. 33, No. 6, pp. 768-785.

Straker, Christian; Niehoff, Maurice (2018): ABIDA-Fokusgruppe – Diskriminierung durch Algorithmen und KI im eRecruiting; in: ZD-Aktuell, No. 06252.

Strauß, Stefan (2018): From Big Data to Deep Learning: A Leap Towards Strong AI or 'Intelligentia Obscura'?; in: Big Data and Cognitive Computing, Vol. 2, No. 3, pp. 1-19.

Sunstein, Cass R. (2014): Nudging: A Very Short Guide; in: Journal of Consumer Policy Vol. 37, No. 4, pp. 583-588.

Supik, Linda (2017): Statistik und Diskriminierung; in: Albert Scherr, Aladin El-Mafaalani and Gökçen Yüksel (eds.): Handbuch Diskriminierung; Wiesbaden: Springer VS, pp. 191-207.

SVRV (2018): Verbrauchergerechtes Scoring. Gutachten des Sachverständigenrats für Verbaucherfragen; Berlin: Sachverständigenrat für Verbaucherfragen (SVRV).

Swedloff, Rick (2014): Risk classification's big data (r) evolution; in: Connecticut Insurance Law Journal, Vol. 21, No. 1, pp. 339-373.

Sweeney, Latanya (2013): Discrimination in Online Ad Delivery; in: Communication of the ACM, Vol. 56, No. 5, pp. 44-54.

Szigetvari, András (2018): Beruf, Ausbildung, Alter, Geschlecht: Der Algorithmus des AMS; in: DerStandard, online article dated 15 October 2018, available at https://derstandard.at/2000089325546/Beruf-Ausbildung-AlterGeschlecht-Das-sind-die-Zutaten-zum-neuen (last retrieved on 28 August 2019).

Tatman, Rachael (2017): Gender and dialect bias in YouTube's automatic captions; in: Proceedings of the First ACL Workshop on Ethics in Natural Language Processing, pp. 53-59.

ten Have, Marieke (2013): Lifestyle Differentiation in Health Insurance. An Overview of the Ethical Arguments, Monitoring Report on Ethics and Health 2013; The Hague: Netherlands Centre for Ethics and Health.

The Royal Society (2017): Machine learning: the power and promise of computers that learn by example; London: The Royal Society.

The White House (2014): Big Data: Seizing Opportunities, Preserving Values; Washington, D. C.: The White House, Executive Office of the President.

Thelwall, Mike (2018): Gender bias in sentiment analysis; in: Online Information Review, Vol. 42, No. 1, pp. 45-57.

Tillmann, Tristan Julian; Vogt, Verena (2018a): Personalisierte Preise – Diskriminierung 2.0?, ABIDA-Dossier; Münster, Karlsruhe: "Assessing Big Data" (ABIDA) project.

Tillmann, Tristan Julian; Vogt, Verena (2018b): Personalisierte Preise im Big-Data-Zeitalter; in: Verbraucher und Recht (VuR), Vol. 33, No. 12, pp. 447-455.

Tolan, Songül (2018): Fair and Unbiased Algorithmic Decision Making: Current State and Future Challenges; Seville: European Commission, Joint Research Centre.

Tolan, Songül; Miron, Marius; Gómez, Emilia; Castillo, Carlos (2019): Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia; in: ICAIL '19, June 17–21, 2019, Montreal, QC, Canada.

Tramèr, Florian; Atlidakis, Vaggelis; Geambasu, Roxana; Hsu, Daniel; Hubaux, Jean-Pierre; Humbert, Mathias; Juels, Ari; Lin, Huang (2017): FairTest: Discovering unwarranted associations in data-driven applications; in: 2017 IEEE European Symposium on Security and Privacy (EuroS&P), pp. 401-416, published by IEEE.

Trute, Hans-Heinrich (1998): Der Schutz personenbezogener Informationen in der Informationsgesellschaft; in: Juristenzeitung, Vol. 53, No. 17, pp. 822-831.

Trute, Hans-Heinrich (2003): Verfassungsrechtliche Grundlagen, Alexander Roßnagel (ed.): Handbuch Datenschutzrecht; Munich: C.H. Beck, pp. 156-187.

ULD; GP Forschungsgruppe (2014): Scoring nach der Datenschutz-Novelle 2009 und neue Entwicklungen. Final report; Kiel, Munich: Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein (ULD); GP Forschungsgruppe.

UN GA (2018): Promotion and protection of the right to freedom of opinion and expression. Seventy-third session, 29 August 2018; New York: United Nations, General Assembly (UN GA).

UNESCO (1951): Statement on Race; Paris: United Nations Educational, Scientific and Cultural Organization (UNESCO).

United States District Court for the Northern District of California (2018): Communications Workers of America v. T-Mobile US Inc, First Amended Class and Collective Action Complaint – Demand for Jury Trial. Case no. 17-cv-07232-BLF, online article available at https://www.onlineagediscrimination.com/sites/default/files/documents/og-cwa-complaint.pdf (last retrieved on 28 August 2019).

US CEA (2015): Big Data and Differential Pricing; Washington, D. C.: Council of Economic Advisers (CEA), Executive Office of the President of the United States.

US HUD (2019a): Charge of Discrimination. HUD versus Facebook; Washington, D. C.: United States of America, Department of Housing and Urban Development, Office of Administrative Law Judges.

US HUD (2019b): HUD charges Facebook with housing discrimination over company's targeted advertising practices, online article published by the U. S. Department of Housing and Urban Development (HUD) dated 28 March 2019, available at https://www.hud.gov/press/press_releases_media_advisories/HUD_No_19_035 (last retrieved on 28 August 2019).

Van Alsenoy, Brendan; Kosta, Eleni; Dumortier, Jos (2014): Privacy notices versus informational self-determination: Minding the gap; in: International Review of Law, Computers and Technology, Vol. 28, No. 2, pp. 185-203.

Van Otterlo, Martijn (2013): A machine learning view on profiling; in: Mireille Hildebrandt and Katja de Vries (eds.): Privacy, Due Process and the Computational Turn. The philosophy of law meets the philosophy of technology; Abingdon: Routledge, pp. 67-90.

Varian, Hal R. (2014): Beyond Big Data; in: Business Economics, Vol. 49, No. 1, pp. 27-31.

Varian, Hal R.; Farrell, Joseph; Shapiro, Carl (2004): The Economics of Information Technology: An Introduction; Cambridge et al.: Cambridge University Press.

Vercellis, Carlo (2011): Business Intelligence: Data Mining and Optimization for Decision Making; Chichester: Wiley.

Verma, Sahil; Rubin, Julia (2018): Fairness definitions explained; in: 2018 IEEE/ACM International Workshop on Software Fairness (FairWare), pp. 1-7, published by IEEE.

Volkova, Svitlana; Bachrach, Yoram (2015): On predicting sociodemographic traits and emotions from communications in social networks and their implications to online self-disclosure; in: Cyberpsychology, Behavior, and Social Networking, Vol. 18, No. 12, pp. 726-736.

von Grafenstein, Max; Hölzel, Julian; Irgmaier, Florian; Pohle, Jörg (2018): Nudging-Regulierung durch Big Data und Verhaltenswissenschaften; Expert report in the context of the „Assessing Big Data“ (ABIDA) project; Münster, Karlsruhe: University of Münster, Karlsruhe Institute of Technology.

Wachter, Sandra; Mittelstadt, Brent; Floridi, Luciano (2017): Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation; in: International Data Privacy Law, Vol. 7, No. 2, pp. 76-99.

Wang, Yilun; Kosinski, Michal (2018): Deep neural networks are more accurate than humans at detecting sexual orientation from facial images; in: Journal of Personality and Social Psychology, Vol. 114, No. 2, pp. 246-257.

Wei, Yanhao; Yildirim, Pinar; Van den Bulte, Christophe; Dellarocas, Chrysanthos (2016): Credit Scoring with Social Network Data; in: Marketing Science, Vol. 35, No. 2, pp. 234-258.

Weichert, Thilo (2013): Big Data und Datenschutz Chancen und Risiken einer neuen Form der Datenanalyse; in: Zeitschrift für Datenschutz, Vol. 3, No. 6, pp. 251-259.

Weichert, Thilo (2014): Scoring in Zeiten von Big Data; in: Zeitschrift für Rechtspolitik (ZRP), Vol. 47, No. 6, pp. 168-171.

Weichert, Thilo (2018): Big Data im Gesundheitsbereich; Expert report in the context of the “Assessing Big Data” (ABIDA) project; Münster, Karlsruhe: University of Münster, Karlsruhe Institute of Technology.

Wersig, Maria (2017): Fälle zum Allgemeinen Gleichbehandlungsgesetz (AGG). Eine Einführung in Theorie und Praxis des Antidiskriminierungsrechts in 23 Fällen; Opladen, Toronto: Verlag Barbara Budrich.

Wiegerling, Klaus (2016): Würde, Autonomie, Subsidiarität – Ist das alles? Ist das viel?, Die Werte des Westens. Wofür wir stehen und werben sollten; lecture at HTWG Konstanz - University of Applied Sciences, 30 May 2016.

Wiegerling, Klaus; Nerurkar, Michael; Wadephul, Christian (2018): Ethische und anthropologische Aspekte der Anwendung von Big-Data-Technologien; in: Barbara Kolany-Raiser, Reinhard Heil, Carsten Orwat and Thomas Hoeren (eds.): Big Data und Gesellschaft. Eine multidisziplinäre Annäherung; Wiesbaden: Springer VS, pp. 1-67.

Williams, Betsy Anne; Brooks, Catherine F.; Shmargad, Yotam (2018): How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications; in: Journal of Information Policy, Vol. 8, pp. 78-115.

Wilson, Benjamin; Hoffman, Judy; Morgenstern, Jamie (2019): Predictive Inequity in Object Detection; arXiv preprint arXiv:1902.11097.

Wimmer, Barbara (2018a): AMS-Chef: „Mitarbeiter schätzen Jobchancen pessimistischer ein als der Algorithmus". Interview given by Johannes Kopf, Director of Arbeitsmarktservice (AMS) on 21 October 2015 in Vienna; in: futurezone, online article dated 12 October 2018, available at https://futurezone.at/netzpolitik/ams-chef-mitarbeiter-schaetzen-jobchancen-pessimistischerein-als-der-algorithmus/400143839 (last retrieved on 28 August 2019).

Wimmer, Barbara (2018b): „AMS-Sachbearbeiter erkennen nicht, wann ein Programm falsch liegt"; in: futurzone, online article dated 18 October 2018, abrufbar unter: https://futurezone.at/netzpolitik/ams-sachbearbeitererkennen-nicht-wann-ein-programm-falsch-liegt/400147472 (last retrieved on 28 August 2019).

WIPO (2019): WIPO Technology Trends 2019 Artificial Intelligence; Genf: World Intellectual Property Organization (WIPO).

Wischmeyer, Thomas (2018): Regulierung intelligenter Systeme; in: Archiv des öffentlichen Rechts, Vol. 143, No. 1, pp. 1-66.

World Wide Web Foundation (2017): Algorithmic Accountability. Applying the Concept to Different Country Contexts; Washington, D. C.: World Wide Web Foundation.

WP29 (2017a): Guidelines on Consent under Regulation 2016/679; Brussels: Article 29 Data Protection Working Party (WP29).

WP29 (2017b): Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679; Brussels: Article 29 Data Protection Working Party (WP29).

Wu, Xiaolin; Zhang, Xi (2016): Automated inference on criminality using face images; arXiv preprint arXiv:1611.04135, pp. 4038-4052.

Yeung, Karen (2008): Towards an Understanding of Regulation by Design; in: Roger Brownsword and Karen Yeung (eds.): Regulating Technologies. Legal Futures, Regulatory Frames and Technological Fixes; Oxford and Portland: Hart, pp. 79-107.

Yeung, Karen (2017): Algorithmic regulation: A critical interrogation; in: Regulation & Governance, Vol. 12, No. 4, pp. 505-523.

Yeung, Karen (2018): A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework; Strasbourg: Council of Europe, Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence (MSI-AUT).

Yue, Lin; Chen, Weitong; Li, Xue; Zuo, Wanli; Yin, Minghao (2018): A survey of sentiment analysis in social media; in: Knowledge and Information Systems, Vol. 60, No. 2, pp. 617-663.

YVTltk (2018): Assessment of creditworthiness, authority, direct multiple discrimination, gender, language, age, place of residence, financial reasons, conditional fine. Plenary Session (voting), Register number:

216/2017, 21 March 2018; Finland, Government Publication: Yhdenvertaisuusja tasa-arvolautakunta/ National Non-Discrimination and Equality Tribunal of Finland.

Zander-Hayat, Helga; Reisch, Lucia A.; Steffen, Christine (2016): Personalisierte Preise: Eine verbraucherpolitische Einordnung; in: Verbraucher und Recht, Vol. 31, H. 11, pp. 403-409.

Zarsky, Tal Z. (2016): The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making; in: Science, Technology & Human Values, Vol. 41, No. 1, pp. 118-132.

Žliobaitė, Indrė; Custers, Bart (2016): Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models; in: Artificial Intelligence and Law, Vol. 24, No. 2, pp. 183-201.

Zuboff, Shoshana (2015): Big other: surveillance capitalism and the prospects of an information civilization; in: Journal of Information Technology, Vol. 30, No. 1, pp. 75-89.

Zuiderveen Borgesius, Frederik (2016): Singling out people without knowing their names Behavioural targeting, pseudonymous data, and the new Data Protection Regulation; in: Computer Law and Security Review, Vol. 32, No. 2, pp. 256-271.

Zuiderveen Borgesius, Frederik (2018): Discrimination, artificial intelligence, and algorithmic decision-making; Strasbourg: Council of Europe, Directorate General of Democracy.

Zuiderveen Borgesius, Frederik; Poort, Joost (2017): Online Price Discrimination and EU Data Privacy Law; in: Journal of Consumer Policy, Vol. 40, No. 3, pp. 347-366.

Zweig, Katharina (2019): Algorithmische Entscheidungen: Transparenz und Kontrolle, Bericht Analysen und Argumente. Digitale Gesellschaft, No. 338; Sankt Augustin: Konrad Adenauer Stiftung.

Zweig, Katharina; Fischer, Sarah; Lischka, Konrad (2018): Wo Maschinen irren können. Fehlerquellen und Verantwortlichkeiten in Prozessen algorithmischer Entscheidungsfindung; Gütersloh: Bertelsmann Stiftung.

Legal information

This publication is part of the public relations work of the Federal Anti-Discrimination Agency. It is distributed free of charge and is not intended for sale.

Publisher:
Federal Anti-Discrimination Agency
11018 Berlin, Germany
www.antidiskriminierungsstelle.de

Author:
Dr Carsten Orwat

Contact:
Tel.: +49 (0)30 185551855
Fax: +49 (0)30 1855541865
Initial legal consultation: Mon 1-3 pm, Wed and Fri 9 am-12 pm
Email: beratung@ads.bund.de
General inquiries: Mon to Fri, 9 am-12 pm and 1-3 pm
Email: poststelle@ads.bund.de
Design www.zweiband.de
As of: September 2019, first edition (German); English translation July 2020