Algorithmic Discrimination From the Perspective of Human Dignity

Orwat, Carsten

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Orwat, C. (2024). Algorithmic Discrimination From the Perspective of Human Dignity. *Social Inclusion*, *12*. https://doi.org/10.17645/si.7160

Social Inclusion
2024 • Volume 12 • Article 7160
https://doi.org/10.17645/si.v12.7160

cogitatio

ARTICLE

Open Access Journal

Algorithmic Discrimination From the Perspective of Human Dignity

Carsten Orwat

Institute for Technology Assessment and Systems Analysis, Karlsruhe Institute of Technology, Germany

Correspondence: Carsten Orwat (orwat@kit.edu)

Submitted: 8 May 2023 **Accepted:** 15 February 2024 **Published:** 13 May 2024

Issue: This article is part of the issue "Artificial Intelligence and Ethnic, Religious, and Gender-Based Discrimination" edited by Derya Ozkul (University of Warwick), fully open access at https://doi.org/10.17645/si.i236

Abstract

Applications of artificial intelligence, algorithmic differentiation, and automated decision-making systems aim to improve the efficiency of decision-making for differentiating persons. However, they may also pose new risks to fundamental rights, including the risk of discrimination and potential violations of human dignity. Anti-discrimination law is not only based on the principles of justice and equal treatment but also aims to ensure the free development of one's personality and the protection of human dignity. This article examines developments in AI and algorithmic differentiation from the perspective of human dignity. Problems addressed include the expansion of the reach of algorithmic decisions, the potential for serious, systematic, or structural discrimination, the phenomenon of statistical discrimination and the treatment of persons not as individuals, deficits in the regulation of automated decisions and informed consent, the creation and use of comprehensive and personality-constituting personal and group profiles, and the increase in structural dominance.

Keywords

algorithmic discrimination; artificial intelligence; automated decision-making; development of personality; generalisation; human dignity; informed consent; profiling; statistical discrimination

1. Introduction

Applications of artificial intelligence (AI), algorithmic differentiation, and automated decision-making systems (ADMs) are increasingly being used to support and automate decisions about the differentiation of persons in areas as diverse as lending, housing, recruitment, welfare benefits assessments, or judicial decision-making. Such differentiations then affect the availability and distribution of products, services,

positions, opportunities, benefits, or burdens, including those that are essential for personal development and the realisation of autonomy and freedom. Here, applications are used that employ machine learning as an analytical method in data mining, profiling, or predictive analytics, and the results of the analysis are used as models or algorithms in decision-making processes.

The article examines issues that arise when applying a human dignity perspective to algorithmic differentiation and discrimination. The concept of human dignity is not monolithic but includes different dimensions of concretisation (e.g., Mahlmann, 2012; Teo, 2023; von der Pfordten, 2023). Among these dimensions, the most relevant for the following are: the protection against instrumentalisation and the protection of self-determination in shaping one's own life by structuring and pursuing one's own interests, desires, and goals; the protection from false and unjustified degradation or humiliation and from being treated without equal moral worth; and the protection of essential conditions for exercising self-determination. The article discusses the constitutional anchoring of human dignity to gain insights into the impact of algorithmic differentiation on fundamental rights.

2. Algorithmic Differentiation and Discrimination

The causes of bias in the use of algorithms, especially of machine learning as a form of AI, are manifold. They result from human decisions about the database used and the development, deployment, or adaptation of algorithms. The most commonly cited causes of bias include training datasets contaminated with historical inequalities and unequal treatments, missing data that would represent certain groups, the selection of inappropriate labels, measurements, or algorithms, inappropriately decided technical trade-offs, or the application of algorithms in domains for which they have not been trained or optimised (e.g., Mehrabi et al., 2021; Pessach & Shmueli, 2022). Bias can lead to products and services not working equally well for different populations or to algorithmic models used to differentiate persons in decisions about access to and distribution of products, services, positions, or freedoms resulting in unequal treatment of those affected. Another problem is stereotyping in generative AI systems (such as ChatGPT), especially when decisions about humans are based on their output (e.g., in automated analysis or summaries of job applications).

Numerous research and development efforts are directed towards making algorithms less discriminatory or "fairer," especially by adapting and cleaning datasets or modifying algorithms. Mathematical fairness definitions or fairness metrics have been developed to express unequal treatment quantitatively and to optimise and compare systems. However, developers and providers must decide on various trade-offs. These include whether and which fairness definition to use, determining the residual risks that affected persons are exposed to, but also trade-offs between individual goals and metrics, for example, between accuracy in achieving differentiation goals and avoiding discrimination risks (e.g., Mehrabi et al., 2021; Pessach & Shmueli, 2022).

Some fairness metrics are ratios formed from the error rates of "false negatives" and "false positives" and the rates of correctly identified classifications related to certain population characteristics. However, it is still unclear how these error rates or fairness metrics will be dealt with in society, which of the fairness metrics should be used to meet certain societal expectations of justice in specific situations, and whether, how, and what levels of residual discrimination risks will have to be borne by society or in which cases residual risks are not acceptable at all. According to the European Union Agency for Fundamental Rights (FRA), when

using fairness metrics, it can only be decided on a case-by-case basis when there is sufficiently significant discrimination, not by setting an abstract threshold (FRA, 2022, p. 25).

The European Union AI Act (adopted text as of 13 March 2024, AI Act in the following; see European Parliament, 2024) refers to the metrics but does not clarify who sets acceptable risk levels, error rates, or limits. In addition, terms such as "residual risk," "acceptable," and "as far as technically feasible" (Article *9*(5) of the AI Act) or "appropriate level of accuracy" (Article *15*(1) of the AI Act) allow for balancing risk avoidance with cost-effectiveness considerations. This should be seen against the background that costs are incurred for the testing of AI systems and risk avoidance. The AI Act leaves open whether the European Commission, standardisation organisations, developers, providers, users, or actors that certify compliance with the regulation will make the normative decisions about discrimination levels and thus about the social realisation of justice. This can lead to residual risks of discrimination at levels undetermined by society. Individual ADMs and AI systems can have a wide reach, for example, due to market concentration, if one system is used in many companies or administrations, or if used as a component (as general-purpose AI component, foundation model, or "AI as a service") in many other systems. Despite seemingly low error rates, this can lead to systematic discrimination affecting a large proportion of the population.

Residual discrimination risks are confronted with anti-discrimination laws. Bias problems lead to legally defined discrimination if biased algorithms are used in differentiations that result in unjustified unequal treatment when using legally protected characteristics (e.g., gender, ethnicity, religion, disability, age, or sexual identity) or when using seemingly neutral characteristics, procedures, rules, or practices that have a connection to the protected characteristics and then actually make groups with protected characteristics worse off (Hacker, 2018; von Ungern-Sternberg, 2022).

The anti-discrimination law, however, has weaknesses with regard to algorithmic discrimination, because in view of algorithmic, often personalised or individualised differentiations, it can be difficult for affected individuals to perceive unequal treatment compared to others and to provide the legally necessary initial evidence of being in a worse position than comparable other persons. However, these are prerequisites for legal proceedings to be initiated, even if the person or entity accused of discrimination has the burden of proof that they are not discriminating (Orwat, 2020, pp. 72–73; von Ungern-Sternberg, 2022). The already high hurdles for affected individuals, which often prevent a legal discrimination case from being filed, are raised even further.

3. The Understanding of Human Dignity in Constitutional Law

Human dignity is often seen as an abstract concept that can be interpreted in different ways (e.g., Mahlmann, 2008). This is a frequent criticism of its use. However, human dignity is included in many human and fundamental rights documents and has been concretised through implementation in legal systems and jurisprudence (e.g., Mahlmann, 2012; McCrudden, 2008). In the following, reference is made to the decisions of the German Federal Constitutional Court and to the discussions and developments of the constitutional law because human dignity is considered to be largely concretised there.

According to this, human dignity is a fundamental claim to value and respect to which every human being is entitled. Human dignity is inherent in the human being. Everyone possesses it, regardless of their

characteristics, achievements, or social status (BVerfGE [Decision] 87, 209; see Federal Constitutional Court, 1992, para. 107). Above all, it includes the protection of personal individuality, identity, and integrity as well as elementary legal equality (BVerfGE 144, 20; Federal Constitutional Court, 2017, headnote 3a, para. 539). The understanding "is based on a conception of human beings as persons who can make free and self-determined decisions and shape their destiny independently" (BVerfGE 144, 20; Federal Constitutional Court, 2017, para. 539). Here, human dignity is primarily concretised by self-determination in shaping one's own life (von der Pfordten, 2023, pp. 48–50).

To further substantiate human dignity, the so-called "object formula" was developed. According to this formula, it is incompatible with human dignity to make the human being a mere object of state action (BVerfGE 27, 1; Federal Constitutional Court, 1969, para. 33). According to the object formula, the human being may not be treated like a thing, reified, or degraded to a mere object. For further concretisation, the Federal Constitutional Court has developed the "subject formula," according to which it is prohibited to treat individual persons in a way that fundamentally calls into question their subject quality by lacking respect for the value they have for their own sake (Hong, 2019, pp. 418–428, 672–690). Höfling (2021, para. 16) suggests that to determine a violation of human dignity in concrete decision-making situations, it is necessary to consider whether the subject status of a human being is still secured by compensation mechanisms despite the objectification in relationships of subordination and dependency.

Certain forms of discrimination directly constitute a violation of human dignity. This is seen, among other things, in the case of direct discrimination by encroachment on the fundamental rights of freedom everyone is entitled to as set out in Article *3*(3) of the German Basic Law (Herdegen, 2022, para. 120). Höfling (2021) sees an unacceptable violation of human dignity in racial discrimination and similar humiliating unequal treatment (Höfling, 2021, para. 35; see also, in particular, BVerfGE 144, 20; Federal Constitutional Court, 2017, para. 541). Hillgruber (2023, para. 17) emphasises that a violation of human dignity occurs not only when people of a certain "race," skin colour, religion, or gender are regarded as "inferior," but also when people are discriminated against based on a physical or mental disability, especially when there is a threat of exclusion because of their disability. The characteristics mentioned are particularly relevant because, on the one hand, they are immutable and personality-constituting characteristics. On the other hand, they are historically justified as a demarcation from the atrocities of the National Socialist injustice regime (Hong, 2019, p. 407; Lehner, 2013, pp. 226–248).

Human dignity is further concretised through its development into the constitutional general right of personality. This includes, among others, the right to informational self-determination (developed to realise the protection of human dignity and free development of personality), the prohibition of discrimination, but also the right to self-expression (Britz, 2007), which are particularly relevant for the following considerations. Thus, anti-discrimination law not only serves to realise the right to equal treatment and socio-political goals but also the right to free development of personality and the protection of human dignity. In this way, Baer (2009) sees dignity as the promise of recognition of different perceptions of self, all of which deserve equal respect. The right to informational self-determination and the prohibition of discrimination serve, among other things, to prevent inappropriate external images of one's personality. These rights should enable individuals to co-decide what they regard as belonging to and constituting their personality (Britz, 2007, p. 16). The core guarantee of the right of personality is to provide mechanisms that involve individuals in the processes of constituting personality in such a way that they can understand their personality as freely chosen (Britz, 2008, p. 191).

In a series of decisions, the Federal Constitutional Court has specified the right to human dignity and personality and related them to the risks posed by information and communication technologies. In the "microcensus" decision (BVerfGE 27, 1; Federal Constitutional Court, 1969), the court made it clear that it is contrary to human dignity to make human beings mere objects in the state. It is incompatible with human dignity to compulsorily register and catalogue human beings in their entire personality and thus treat them as a thing that is available to an inventory in every respect (para. 33).

The census decision (BVerfGE 65, 1; Federal Constitutional Court, 1983) established the fundamental right to informational self-determination, which serves the right to free development of personality in conjunction with the right to human dignity. It guarantees individuals the authority to decide for themselves about the disclosure and use of their personal data (headnote 1). The right to informational self-determination not only serves to secure the external and internal dimensions of freedom (freedom of action and identity formation) but also to avoid chilling effects that may arise for data subjects due to uncertainties about data processing (Britz, 2010).

In its decision on acoustic surveillance (BVerfGE 109, 279; Federal Constitutional Court, 2004), the court recognises a core area of private life that enjoys absolute protection concerning the inviolability of human dignity. What belongs to the core of private life depends on whether the facts of the case are of a highly personal nature (para. 123). Similarly, sweeping surveillance, in terms of time and location, violates human dignity if it is carried out over an extended period of time and all movements and expressions of the life of the person concerned are recorded and can become the basis of a personality profile (para. 150).

In the decision on the right to be forgotten I (BVerfGE, 152, 152; Federal Constitutional Court, 2019), the court interprets the right to informational self-determination for relationships between private actors in such a way that the right ensures the individual "substantial influence in deciding what information is attributed to their person" (para. 87). The court found that in many life situations, private companies provide the basic services that play a crucial role in shaping public opinion, allocating or denying opportunities, or enabling participation in social or daily life. In many cases, this is done based on extensive data collection and processing, often by companies with market power, where large-scale disclosure of personal data can hardly be avoided in order not to be excluded from the services or opportunities (para. 85). In cases of extensive dependencies or imposition of unacceptable contractual conditions (para. 85) or "where private companies take on a position that is so dominant as to be similar to the state's position…the binding effect of the fundamental right on private actors can ultimately be close, or even equal to, its binding effect on the state" (para. 88).

In the court's ruling on automated data analysis in police work (BVerfG 1 BvR 1547/19; Federal Constitutional Court, 2023), the court emphasises, among other things, that automated data analysis can be used to generate new otherwise inaccessible information, that come close to full profiles of persons affected, by linking existing datasets (para. 69). The court also refers to the discrimination risks of automated data analyses (para. 100), which become less tolerable the more the analyses are capable to produce disadvantages prohibited under Article *3*(3) of the German Basic Law (para. 77). Furthermore, it stresses the importance of the ability to scrutinise algorithms for individual legal protection and administrative oversight in order to identify and correct errors (para. 90). A risk of loss of state oversight is seen particularly in the use of self-learning systems or AI, as these can become detached from the original human programming in the course of the machine learning

process and results become increasingly difficult to scrutinise (para. 100, with reference to the judgement of the European Court of Justice, 2022).

4. Factors of Violation of Human Dignity With Algorithmic Differentiation

4.1. Severe and Structural Discrimination

Algorithmic differentiation can lead to systematic or structural discrimination due to the possibility of residual discrimination risks and the wide reach of the system, which can cover entire populations. Some factors may indicate a violation of human dignity when discrimination is based on "race" or ethnicity, gender, physical and mental disabilities, and the other protected characteristics of Article *3*(3) of the German Basic Law, and when algorithmic differentiation is used in application contexts where there is a strong dependency on the benefits, products, services, or where applications affect particularly vulnerable persons. These groups are considered particularly worthy of protection under constitutional law. Severe discrimination would violate their dignity, as they are systematically degraded. This applies in particular to decisions concerning a dignified life (e.g., in the cases of welfare recipients, refugees, or migrants). Here, discrimination can involve forms of humiliation or degradation, as a person may not be regarded and treated with equal moral worth (e.g., the SyRi and Robodebt scandals; see Teo, 2023, p. 17).

In this respect, algorithmic differentiations that consolidate or expand structural inequalities through negative feedback loops are also problematic. In addition to exacerbating the problem of not being treated as persons of equal moral worth, negative feedback loops can impair the self-determination of those affected by making it difficult or even impossible for them to escape inappropriate classifications or stereotypes on their own. Such negative feedback loops can arise when results from algorithmic differentiation systems that predict the behaviour of persons are recaptured and the systems use this uncorrected data as the basis for further data analysis, inference, or the further development or learning of the algorithms. Examples can be found in predictive policing systems (FRA, 2022; Lum & Isaac, 2016).

4.2. Generalisation and Lack of Individual Justice

Algorithmic differentiation often takes on and alters forms of so-called statistical discrimination (Barocas & Selbst, 2016; Binns et al., 2018). Statistical discrimination is a form of proxy discrimination. Instead of using an elaborate case-by-case examination to determine the actual personality traits or the differentiation target (e.g., the social construct "actual ability to pilot an aircraft"), comparatively easy-to-obtain proxy information (e.g., age in years) is used. This form of differentiation is intended to efficiently overcome an information deficit. Discrimination can occur if the proxies are legally protected characteristics or contain characteristics that correlate with protected characteristics (e.g., Britz, 2008; Hellman, 2008; Schauer, 2018). The proxy information can be derived from empirical studies or, in the case of machine learning, be present in models trained on data.

However, statistical discrimination and generalisation (either by human decision-makers or through the use of algorithms) are already ethically problematic in themselves because group information is applied to individuals and thus acts as quasi-stereotypes and prejudices in decision-making (Gandy, 2010, p. 34). In principle, case-by-case justice is not guaranteed as there is no case-by-case examination and the

individual subject characteristics and the individual situations and contexts are not considered (Britz, 2008). What is often referred to as "prediction" in AI research and practice is not a prediction of an individual's potential behaviour derived from an individual examination. Rather, it is an assignment of individuals to categories, scores, or rankings that are formed statistically or by machine learning and are expected to produce certain outcomes for the sorted individuals in the future.

Moreover, in many cases of algorithmic differentiation, the categories to which individuals are assigned are constructed using data from groups that do not contain the individuals who are actually decided upon (Eckhouse et al., 2019, pp. 198–199). In addition, the categories constructed with AI are usually not comprehensible to those affected or to third parties. In contrast to the use of clearly communicated criteria (e.g., age limits in public administration), such decision criteria and rules evade scientific and public scrutiny and discussion, in particular, whether there is a causal relationship between the criteria and the differentiation goal at all, whether this can be substantiated with evidence or whether the use of certain criteria is socio-politically or morally controversial or undesirable. The main aim of scientific and public scrutiny is to avoid unfounded or spurious correlations (Schauer, 2018), which carry the risk of unjustified degradation.

4.3. Not Treating Persons as Individuals and With Respect

In contrast to the individualisation of decisions about people, statistical discrimination and generalisation treat people as information objects rather than as individuals. The question of when it is morally problematic to treat people not as individuals but only as members of a group (stereotyping) is controversial (e.g., Beeghly, 2018; Lippert-Rasmussen, 2011).

This question often brings to mind the moral considerations of Kant, who demands respect for human beings and the prohibition of instrumentalisation (according to Dillon, 2022, Chapter 2.2; see also Hill, 2014, pp. 315–318). Thus, every human being is required to "acknowledge, in a practical way, the dignity of humanity in every other human being. Hence there rests on him a duty regarding the respect that must be shown to every other human being" (Kant, 1797/2017, p. 225). The respect that people owe each other and that they can demand from other people is respect for their dignity (Kant, 1797/2017, p. 225; according to Schaber, 2016, p. 256; Ulgen, 2017, 2022, pp. 14–15). Respecting the dignity of another person (and of oneself) means treating others "always at the same time as an end, never merely as a means" (Kant, 1785/2012, p. 41).

According to Schaber (2016), who interprets Kant's explanations of the false promise for this purpose, Kant also means that one treats another person merely as a means if one treats the other person in a way the person cannot possibly consent to. This is the case if they have no reason to do so and would not behave rationally if they consented. Respecting the dignity of another person therefore means treating them in such a way that they can reasonably consent (Schaber, 2006, p. 256).

Drawing on Kant, Korsgaard (1996) elaborates the idea that the test for treating another person as a mere means is whether the other person can consent to the way they are treated. In cases of coercion or deception, the other person cannot do so, since in both forms of treatment the other person has been given no chance to choose the end. Hence, treatment is morally wrong if other persons are not able to choose.

She therefore concludes that coercion and deception are, according to Kant's formula of humanity, the most fundamental forms of wrongdoing towards others, the root of all evil (Korsgaard, 1996, pp. 137–140). Schaber (2013, pp. 134–136) adds that deception can be problematic in situations where it impairs the rights of the affected persons to determine essential aspects of their own lives.

Also referring to Kant, Ulgen (2022) develops requirements for treating persons with respect for their inherent dignity with regard to AI. Dignity arises from the autonomy and rational capacity of humans to exercise reasoning, judgements, and choices. Human autonomy is protected if humans "are able to act influenced by reason; if they can identify the motivations prompting their action; or they can change their motivations if they cannot identify with them" (Ulgen, 2022, p. 19). AI systems that diminish the human agency to exercise reasoning, judgement, and choice undermine human dignity (Ulgen, 2022, p. 27). The argument is also relevant to technologically implemented social rules (see Section 4.6).

In summary, these arguments show the importance of having requirements for how persons are treated as well as functioning compensation mechanisms to prevent individuals from being treated as mere objects. This includes, in particular, the opportunity to consent to and influence treatments. The protection of human dignity also includes the protection of choice of ends and the requirement to be informed about choices in such a way that individuals can act in a self-determined manner.

Philosophical approaches to explaining when differentiation is morally wrong come to similar conclusions. These include approaches based on discrimination theory oriented towards human dignity and disregard (Khaitan, 2015, pp. 6–8), even if the term "dignity" is not always used. They consider a differentiation to be wrong if the discriminating person assesses the moral value of the discriminated person wrongly, in particular as lower, or if the discriminating person expresses a wrong assessment, i.e., acts as if the discriminated person has a lower moral value (Thomsen, 2017).

For Hellman (2008), the moral wrong of discrimination is that it degrades a person. Degrading rules or practices express a disregard for the moral equality of those being discriminated against. According to Hellman (2016) it is important to first clarify the meaning of the term "dignity" before using it. Discrimination is wrong from the perspective of human dignity because it results either from the fact that people are not treated with equal worth, i.e., people do not receive the same level of recognition and respect, or from the fact that people are denied rights to which they are entitled (Hellman, 2016, pp. 943–946). Such rights can include the right to self-determination over essential aspects of life and the right to receive a justification.

Eidelson (2015) conceives the wrong of discrimination as a failure to treat a person correctly as an individual. The error lies in not seeing the person as the (partial) result of their past efforts at self-creation and as an autonomous agent whose future decisions they can make for themselves. About the problem that statistical discrimination and generalisation do not treat persons as individuals, he therefore calls for an understanding that persons are treated as individuals if—and only if—(a) the differentiating person X gives appropriate weight to the evidence of how the affected person Y has exercised autonomy in shaping his or her life, provided that this evidence is reasonably available and relevant to the decision at hand; and (b) in addition, X's judgements, when they relate to the choices of person Y, must not be made in such a way as to disparage Y's capacity to make those choice decisions as an autonomous agent (Eidelson, 2015, pp. 144, 227).

Although questions remain about the scope and types of adequate and relevant evidence and about obligations to provide it, it can be deduced that the object of information and decision-making should be the self-determined personality development of the persons concerned, in particular their possibilities of self-perception, self-determination, and self-expression. However, a dilemma must be avoided. If personal data on self-determination is to be collected to solve the problem of generalisation and to better respect persons as individuals, this can only be done by having the data and profiles controlled by the persons affected themselves to avoid a violation of the right to informational self-determination. In addition, it may be necessary not to rely solely on automated data collection and analysis but to involve human decision-makers to collect sufficient additional information and to make situation- and person-specific judgements that require a high degree of situational balancing when evaluating the information provided.

4.4. Automated Decision-Making

According to Citron (2008) and Kaminski (2019), automated decision-making based on generalisations are ethically problematic because no other information about the persons concerned is processed apart from the generalisation information. If persons are merely assigned to algorithmically formed categories, scores, or rankings, persons are no longer treated as individuals. If automated decisions no longer allow persons to express their individuality, this violates their dignity and turns people into objects based on a few characteristics instead of treating them as whole persons. Both the exercise of human discretion and individual procedural rights (of appeal, correction, etc.) are necessary not only to avoid error but also to properly recognise and respect individuality (Citron, 2008, p. 1304; Kaminski, 2019, pp. 1541–1545). Furthermore, human discretion is necessary when human decision-makers must also be able to consider mitigating circumstances that the algorithm cannot, as well as when there are indeterminate terms in the decision rules that require the human decision-maker to make trade-offs between conflicting interests (Citron, 2008, p. 1304).

One of the justifications for regulating automated decision-making is the protection of human dignity. This refers to Article 22 of the General Data Protection Regulation (GDPR; European Parliament and Council of the European Union, 2016) and its predecessor, Article 15 of the Data Protection Directive 95/46/EC (European Parliament and Council of the European Union, 1995). According to Dammann and Simitis (1997), the prohibition of automated decision-making of Article 15 of the Data Protection Directive was intended to prevent data subjects from being treated in personality assessments only by computer and based on stored data. This would ignore the individuality of the person and degrade the person to a mere object of computer operations (Dammann & Simitis, 1997, pp. 218–219; see also Jones, 2017; Kaminski, 2019; Martini, 2021, para. 8; Scholz, 2019, para. 3).

According to Martini and Nink (2017), the subject quality of a person is not necessarily disregarded by the fact that personal data alone are the object of an algorithmic analysis. In the case of automated (administrative) decisions, the quality of the subject is only affected when algorithmic procedures impose adverse consequences on the person concerned without allowing them to defend themselves against the decision in an appropriate way. To protect informational self-determination, the legal practice relies on the procedure of (a) informing about the automated decision, (b) communicating and explaining the essential reasons for the decision upon request, (c) allowing the data subject to assert their own point of view in order to obtain a review and re-evaluation if necessary (Martini & Nink, 2017, p. 7).

However, some deficits in the actual regulation of automated decision-making raise doubts as to whether it can still serve to protect human dignity. The so-called "prohibition" is provided with extensive exceptions, particularly if the automated decision-making serves to conclude or fulfil a contract, is required by law to be permissible, or explicit consent is given. Although the regulation provides that the operator of an automated decision must inform about the existence of an automated decision and the so-called logic involved, it is still unclear what the content of this information obligation is, in particular, whether and how information about decision criteria or possible discrimination risks must be provided (Orwat, 2020, pp. 77–82).

Moreover, the "prohibition" is often interpreted only as a right of intervention for the persons affected in justified individual cases (Martini & Nink, 2017, p. 4). In this context, the persons affected must first have knowledge of the automated decision-making and its effects, and justify their desire for human intervention and an explanation of the logic involved. As this can be very burdensome, it may have a chilling effect if individuals perceive it as an unreasonably high hurdle to avail themselves of the regulation. Persons affected may be deterred from exercising the rights to which they are entitled and which were established to protect their dignity.

4.5. Emergence of New Knowledge and Comprehensive and Meaningful Profiles

The possibilities of data aggregation, data reuse, data combination and inference, de-anonymisation and re-identification of individuals, categorisation, ranking, assessment, and individual or group profiling of individuals have greatly increased with AI (e.g., FRA, 2020; Smuha, 2021; Yeung, 2019). Some AI systems have been developed to make automated inferences about identity, personality-constituting traits, and other sensitive information such as emotions, character traits, mental states, or political orientations (e.g., Kosinski, 2021; Matz et al., 2023). AI-based biometric and psychometric evaluations (e.g., emotional AI) can be used for targeting (e.g., in marketing), risk assessment (e.g., in applicant selection or calculating the probability of dropping out of university or defaulting on a loan), and behavioural control (Valcke et al., 2021). With other AI methods, researchers strive to predict future events in the lives of persons, including personality nuances and the time of death (Savcisens et al., 2023).

The systems are often based on a reduction of personality to quantifiable measures and classes that attempt to map the personality traits relevant to a differentiation goal. Further critical issues are the standardisation of personality (Köchling et al., 2021) or the pseudo-scientific approach (Sloane et al., 2022). Even if the scope and types of application of such systems are still little known in practice, this illustrates that AI can generate and use sweeping and meaningful profiles that are suitable for imposing an (almost) complete external image on a person, with personality-constituting characteristics and even without the valid consent of the person concerned (see Section 4.7).

All in all, this leads to a further detachment of the data representation by the operators from the possibilities of controlling the self-representation by the persons concerned (cf. Teo, 2023). The options of agency and the identity of those affected are then solely determined externally, including how they (have to) see themselves (as normal, healthy, conforming to rules, etc.). The abilities to develop and shape their life in a self-determined way are then eliminated. According to the standards outlined above, this can be a violation of (informational) self-determination and human dignity. Such AI applications can also have chilling effects and thus lead to self-restriction of the free development of personality as a form of violation of dignity.

4.6. Structural Dominance

In the relationship between the state and persons affected (e.g., citizens or migrants), the structural dominance of the state must be assumed as a matter of principle, because there are usually situations with a monopoly on the use of force, a lack of options for evasion, subjection, non-negotiability, and the complete, unilaterally determined binding nature of the rules. A number of factors can also increase the structural dominance of users of AI systems (e.g., providers, employers, banks) over those affected in private relationships (e.g., customers, applicants, credit seekers).

Firstly, social rules are increasingly being implemented in software or algorithms in both the public and private sectors. For technical implementation, the rules must be written in programming languages or generated in the form of algorithms or models through machine learning. With these forms of specification, however, the scope for interpretation and discretion of social rules for human decision-makers is also lost, which is often required so that social rules can be applied to many, sometimes unpredictable, situations. If rules are enforced fully automatically, such as in fully automated decision-making, deviation from the rules is normally technically prevented. Also, in algorithmic choice architectures (nudging), the space of choices is technically predetermined and often limited. The rules are then often established unilaterally by those developing and applying them, and the affected persons' possibilities for negotiation, contesting, influencing, and correction are reduced or eliminated, which can increase structural dominance. If the scope for interpretation, discretion, and choice is reduced, the possibilities for action, autonomy, and the opportunities for self-enforcement of autonomy by those affected are also reduced (Deutscher Ethikrat, 2023, pp. 120–137; Teo, 2023, pp. 27–31; Ulgen, 2022). Secondly, if the private users of AI systems are also those who operate platforms (and in some cases also those who develop AI systems), strong network effects of the platforms can reduce the opportunities for evasion and increase users' dependency on the applications. Thirdly, as just shown, some AI systems can determine sensitive personality traits such as mental states, character traits, and emotions even from seemingly "trivial" data such as communication in social networks. In this way, reliance on a product, service, or position can be better determined and exploited (e.g., Härtel, 2019), as can human weaknesses, especially when the systems are used for "dark patterns" or other forms of manipulation (e.g., Ulgen, 2022, pp. 22–24).

Developments that increase the structural dominance of the users of AI systems over those affected can tend to restrict human dignity and the free development of personality because those affected are restricted in their ability to shape their own lives. This can happen in private relationships by way of disrupting contractual parity and thus reducing the possibilities for those affected to assert their self-determination themselves. For even if freedom of contract applies, i.e., everyone has the freedom to determine with whom and under what conditions contracts are entered into, it must be ensured that also the conditions of free self-determination are actually given (BVerfGE 81, 242; Federal Constitutional Court, 1990, para. 47).

4.7. Absence of the Possibility of Valid Consent

The instrument of consent as a compensation mechanism for treating persons as mere objects can transform morally impermissible treatments into permissible ones. However, it can only achieve this moral transformation if certain preconditions are met. These include that the affected persons consent voluntarily and have choices to do so, that they are sufficiently informed, i.e., understand the data processing,

decision-making, and the consequences thereof, and that they have the necessary decision-making skills (Bullock, 2018). On the other hand, from a philosophical point of view, it is also doubted that consent can transform treatment as a mere object into morally permissible treatment if the treatment already violates the duty to treat other people with respect (Fahmy, 2023). This is likely to be the case, for example, with serious algorithmic discrimination or differentiations based on profiles with (almost) complete recording and determination of personality without control by the persons affected.

In the field of data protection, the problems of informed consent have long been recognised. The effectiveness and meaningfulness are increasingly limited by non-negotiable, long, incomprehensible data protection declarations or privacy policies formulated in legalese, the increasing collection of data that is based on so-called legitimate interests without the need for consent (Article *6*(1)f GDPR; European Parliament and Council of the European Union, 2016), strong network effects and thus the tying of customers and users to systems or platforms, which reduces voluntariness as well as interface designs that entice consent to data collection. Those affected often lack knowledge about the necessity and legal possibilities of informed consent. They can hardly assess the actual consequences of consent regarding potentially detrimental, sometimes long-term treatments, which can also arise from data accumulation that is difficult to trace or further data processing and transfer that can no longer be assessed. Furthermore, the decision criteria of complex AI algorithms may be incomprehensible or unknown, particularly in the case of self-learning or adaptive systems. Similarly, AI-based reasoning can generate new knowledge from existing personal data, even from anonymised data, and even for individuals or groups not involved in the original consent. It can then be assumed that the data subjects can no longer sufficiently recognise what they are consenting to (e.g., Orwat, 2020, pp. 71–72). Due to these factors, established types of informed consent becomes increasingly useless as a legitimisation of the treatment of people as mere objects and as an instrument of self-determination.

Zarsky (2013) explains in more detail that in order to protect human dignity, there must be an understanding of the inner workings of automated data analyses, as without this understanding the results can still appear arbitrary and wrong. The problem is that existing (legal) transparency requirements are at best sufficient to provide information about the correlations or classifications a person might fall into. Instead, the automated prediction process must also be interpretable, i.e., the selection process must be explainable. The protection of human dignity therefore requires that causations and not merely correlations must be ascertainable for those affected before conclusions and measures are taken (Zarsky, 2013, p. 1548).

5. Conclusions

The human dignity perspective on algorithmic differentiation and discrimination can help determine how humans or machines should treat other humans as individuals and with respect. This perspective can complement work on the (relative) fairness of algorithmic differentiation and the technical mitigation of discrimination risks. The perspective of human dignity goes beyond efforts to de-bias datasets and algorithms. It leads to the question of how algorithm-based decision-making applications should be shaped and what information bases and forms of communication should be used. It also helps differentiate more clearly between different application situations, for example, for which applications and purposes it should be able or allowed to use the new capabilities of AI. Moreover, the perspective provides justifications for when AI and ADMs should not be used due to the possible restriction or violation of human dignity.

For example, the applicability of AI and ADMs is questionable if decision-making situations are unavoidable for those affected and they have de facto no possibility to influence decisions, if self-determination is severely restricted, or if serious degradation may occur. It also points out that even the ideal case of "accurate" profiles or categories as the basis of decisions is problematic if overpowering external profiles suppress the informational self-determination of those concerned.

The use of AI and algorithmic differentiation can lead to a violation or restriction of human dignity. Problematic factors include: (a) the use of generalisations and disregard of personality in decisions on unequal treatment using variables (proxies) that have no comprehensible connection to the differentiation goal, use morally dubious connections, lack rational justification, and cannot be contested by the persons concerned; (b) the reach of systems with residual risks of systematic and structural discrimination, which, among other things, expose some persons to higher risks of discrimination, do not guarantee equal protection for all, and thus treat them as persons with inferior moral worth; (c) the increasingly inadequate established types of informed consent, which has a particularly drastic effect with AI systems whose decision-making criteria and far-reaching consequences are no longer comprehensible to those affected; (d) the inadequate regulation of (fully) automated decisions, the role of the human decision-makers involved in them, and informed consent to it; (e) the loss of control over the generation and use of sweeping and meaningful personal and group profiles by those affected, leaving no scope for self-determination over the external image created; and (f) the increasing structural dominance of the state and private companies through increasing technical enforcement of social rules, market concentration, specific capabilities of AI to generate new knowledge about persons from existent data and to detect and exploit dependencies or other human weaknesses and thus situations with distorted contractual parity, strong dependencies, limited options for action, contestation, influence, avoidance, and the resulting inevitability. As a result, in situations where these factors have an impact, either alone or in combination, the protection of human dignity can no longer be guaranteed. These impacts are all the more severe the more essential the products and services are for the self-determined shaping of life and identity, or for a dignified existence of people with special needs and vulnerabilities (e.g., access to education, employment, health and welfare services, finance, housing, or asylum).

It is important to further clarify when human dignity and the development of personality are specifically restricted or violated and how they can be protected. A context-specific approach is necessary as different products, services, or treatments have varying degrees of relevance to the preservation of dignity. The prohibitions of certain AI applications in the European Union AI Act is justified, among other things, by the aim of protecting human dignity (e.g., Recital 28 and 31 AI Act; European Parliament 2024). The prohibitions include certain social scoring systems, subliminal, manipulative, or deceptive techniques, systems that exploit vulnerabilities (due to age, disability, social or economic situation), systems that infer emotions in the workplace or in education, biometric categorisation systems that use or infer sensitive characteristics (e.g., political opinions, religious or philosophical beliefs, sexual orientation, race), with many exceptions, certain remote biometric identification systems for law enforcement, and other applications. However, the following considerations may help interpret and develop the AI Act, the regulatory framework in general, and in designing AI and ADM applications. The issues include:

1. Which personal and group profiles are so comprehensive or constitutive of one's personality that the image of the personality must be described as externally determined, as no longer freely chosen, and

the self-determination of essential aspects of the shaping of life as undermined (e.g., if the boundary to social scoring is exceeded, if an external profile unduly restricts access to services and products essential for personality development or if it is completely beyond the control of those affected)?

2. Under what conditions does serious, systematic, or structural discrimination actually exist or is to be expected, and in what areas can residual discrimination risks not be tolerated as this would lead to a violation of dignity (e.g., in areas where there are no or limited possibilities to influence or evade decisions that belong to the core of a self-determined shaping of life, for persons with special vulnerabilities or large dependencies, or where decision-making could result in a massive, unjustified, serious degradation of persons)?
3. To what extent and in what form should the personality of those affected be respected in algorithm-based decisions, and what form of justification for decisions must those affected receive (e.g., by explaining and justifying the use of causal relationships between the criteria used and the differentiation goals, by providing all information necessary for self-determination, by preventing trade secret interests from taking precedence over information claims based on human dignity, or by involving human decision-makers who ensure consideration of specific personal information and situational balancing)?
4. What options and capabilities must those affected have to influence decisions and their personality profile, and what should communicative processes look like in this respect (e.g., requirements for the explainability or comprehensibility of AI and ADMs to provide an adequate information basis for contesting and influencing decisions by those affected)? Where are the capabilities of those affected inappropriate or limited and other institutional actors should become active (e.g., through collective redress)?
5. How is it possible not only to (re)strengthen the human dignity and personality development of those directly affected but also to ensure the protection of groups or third parties who are unaware that they are affected?

Acknowledgments

The author would like to thank the three anonymous reviewers as well as his colleagues Harald König, Philipp Frey, Reinhard Heil, Michael Schmidt, Sylke Wintzer, and the editor of the thematic issue for their valuable feedback.

Conflict of Interests

The author declares no conflict of interest.

References

Baer, S. (2009). Dignity, liberty, equality: A fundamental rights triangle of constitutionalism. *University of Toronto Law Journal*, *59*(4), 417–468.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, *104*(3), 671–732.

Beeghly, E. (2018). Failing to treat persons as individuals. *Ergo: An Open Access Journal of Philosophy*, *5*(26), 687–711.

Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., & Shadbolt, N. (2018). *'It's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions* [Paper presentation]. 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada.

Britz, G. (2007). *Freie Entfaltung durch Selbstdarstellung. Eine Rekonstruktion des allgemeinen Persönlichkeitsrechts aus Art. 2 I GG*. Mohr Siebeck.

Britz, G. (2008). *Einzelfallgerechtigkeit versus Generalisierung. Verfassungsrechtliche Grenzen statistischer Diskriminierung*. Mohr Siebeck.

Britz, G. (2010). Informationelle Selbstbestimmung zwischen rechtswissenschaftlicher Grundsatzkritik und Beharren des Bundesverfassungsgerichts. In W. Hoffmann-Riem (Ed.), *Offene Rechtswissenschaft* (pp. 561–596). Mohr Siebeck.

Bullock, E. C. (2018). Valid consent. In P. Schaber & A. Müller (Eds.), *The Routledge handbook of the ethics of consent* (pp. 85–94). Routledge.

Federal Constitutional Court. (1969). *Mikrozensus: Beschluß des Ersten Senats vom 16. Juli 1969* (1 BvL 19/63). http://www.servat.unibe.ch/dfr/bv027001.html

Federal Constitutional Court. (1983). *Volkszählung: Judgment of the First Senate of 15 December 1983* (1 BvR 209/83). https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/1983/12/rs19831215_1bvr020983en.html

Federal Constitutional Court. (1990). *Handelsvertreter: Beschluß des Ersten Senats vom 7. Februar 1990* (1 BvR 26/84). https://www.servat.unibe.ch/dfr/bv081242.html

Federal Constitutional Court. (1992). *Tanz der Teufel: Beschluß des Ersten Senats vom 20. Oktober 1992* (1 BvR 698/89). https://www.servat.unibe.ch/dfr/bv087209.html

Federal Constitutional Court. (2004). *Großer Lauschangriff: Order of the First Senate of 3 March 2004* (1 BvR 2378/98). https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/2004/03/rs20040303_1bvr237898en.html

Federal Constitutional Court. (2017). *NPD-Verbotsverfahren: Judgment of the Second Senate of 17 January 2017* (2 BvB 1/13). https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/2017/01/bs20170117_2bvb000113en.html

Federal Constitutional Court. (2019). *Right to be forgotten I: Order of the First Senate of 6 November 2019* (1 BvR 16/13). https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/2019/11/rs20191106_1bvr001613en.html

Federal Constitutional Court. (2023). *Automated data analysis: Judgment of the First Senate of 16 February 2023* (1 BvR 1547/19). https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/2023/02/rs20230216_1bvr154719en.html

Citron, D. K. (2008). Technological due process. *Washington University Law Review*, *85*(6), 1249–1313.

European Court of Justice. (2022). *Ligue des droits humains ASBL v Conseil des ministres* (Case C-817/19). https://curia.europa.eu/juris/liste.jsf?lgrec=fr&td=%3BALL&language=en&num=C-817/19&jur=C

European Parliament and Council of the European Union. (1995). Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (Data Protection Directive). *Official Journal of the European Communities*, L 281/31. http://data.europa.eu/eli/dir/1995/46/oj

European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, L 119/1. http://data.europa.eu/eli/reg/2016/679/2016-05-04

European Parliament (2024). *European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts* (COM(2021)0206—C9-0146/2021—2021/0106(COD)). https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html

Dammann, U., & Simitis, S. (1997). *EG-Datenschutzrichtlinie: Kommentar*. Nomos.

Deutscher Ethikrat. (2023). *Mensch und Maschine—Herausforderungen durch Künstliche Intelligenz*.

Dillon, R. S. (2022). Respect. In E. N. Zalta, & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy*. https://plato.stanford.edu/archives/fall2022/entries/respect

Eckhouse, L., Lum, K., Conti-Cook, C., & Ciccolini, J. (2019). Layers of bias: A unified approach for understanding problems with risk assessment. *Criminal Justice and Behavior*, *46*(2), 185–209.

Eidelson, B. (2015). *Discrimination and disrespect*. Oxford University Press.

Fahmy, M. S. (2023). Never merely as a means: Rethinking the role and relevance of consent. *Kantian Review*, *28*(1), 41–62.

FRA. (2020). *Getting the future right—Artificial intelligence and fundamental rights*. European Union Agency for Fundamental Rights.

FRA. (2022). *Bias in algorithms—Artificial intelligence and discrimination*. European Union Agency for Fundamental Rights.

Gandy, O. H., Jr. (2010). Engaging rational discrimination: Exploring reasons for placing regulatory constraints on decision support systems. *Ethics and Information Technology*, *12*(1), 1–14.

Hacker, P. (2018). Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law. *Common Market Law Review*, *55*(4), 1143–1185.

Härtel, I. (2019). Digitalisierung im Lichte des Verfassungsrechts—Algorithmen, Predictive Policing, autonomes Fahren. *Landes- und Kommunalverwaltung*, *29*(2), 49–60.

Hellman, D. (2008). *When is discrimination wrong?* Harvard University Press.

Hellman, D. (2016). Two concepts of discrimination. *Virginia Law Review*, *102*(4), 895–952.

Herdegen, M. (2022). Art. 1 Abs. GG (Schutz der Menschenwürde). In T. Maunz & G. Dürig (Eds.), *Grundgesetz-Kommentar*. Beck.

Hill, T. E., Jr. (2014). In defence of human dignity: Comments on Kant and Rosen. In C. McCrudden (Ed.), *Understanding human dignity* (pp. 313–325). Oxford University Press.

Hillgruber, C. (2023). GG Art. 1 (Schutz der Menschenwürde). In V. Epping & C. Hillgruber (Eds.), *BeckOK (Online-Kommentar) Grundgesetz*.

Höfling, W. (2021). Art. 1 GG Schutz der Menschenwürde, Menschenrechte, Grundrechtsbindung. In M. Sachs (Ed.), *Grundgesetz: Kommentar* (pp. 70–102). Beck.

Hong, M. (2019). *Der Menschenwürdegehalt der Grundrechte. Grundfragen, Entstehung und Rechtsprechung*. Mohr Siebeck.

Jones, M. L. (2017). The right to a human in the loop: Political constructions of computer automation and personhood. *Social Studies of Science*, *47*(2), 216–239.

Kaminski, M. E. (2019). Binary governance: Lessons from the GDPR's approach to algorithmic accountability. *Southern California Law Review*, *92*(6), 1529–1616.

Kant, I. (2012). *Groundwork of the metaphysics of morals—Revised edition*. Cambridge University Press. (Original work published 1785)

Kant, I. (2017). *The metaphysics of morals*. Cambridge University Press. (Original work published 1797)

Khaitan, T. (2015). *A theory of discrimination law*. Oxford University Press.

Köchling, A., Riazy, S., Wehner, M. C., & Simbeck, K. (2021). Highly accurate, but still discriminatory. *Business & Information Systems Engineering*, *63*(1), 39–54.

Korsgaard, C. M. (1996). *Creating the kingdom of ends*. Cambridge University Press.

Kosinski, M. (2021). Facial recognition technology can expose political orientation from naturalistic facial images. *Scientific Reports*, *11*(1), Article 100. https://doi.org/10.1038/s41598-020-79310-1

Lehner, R. (2013). *Zivilrechtlicher Diskriminierungsschutz und Grundrechte. Auch eine grundrechtliche Betrachtung des 3. und 4. Abschnittes des Allgemeinen Gleichbehandlungsgesetzes (§§19-23 AGG)*. Mohr Siebeck.

Lippert-Rasmussen, K. (2011). "We are all different": Statistical discrimination and the right to be treated as an individual. *The Journal of Ethics*, *15*(1), 47–59.

Lum, K., & Isaac, W. (2016). To predict and serve? *Significance*, *13*(5), 14–19.

Mahlmann, M. (2008). *Elemente einer ethischen Grundrechtstheorie*. Nomos.

Mahlmann, M. (2012). Human dignity and autonomy in modern constitutional orders. In M. Rosenfeld & A. Sajó (Eds.), *The Oxford handbook of comparative constitutional law* (pp. 1–26). Oxford University Press.

Martini, M. (2021). DS-GVO Art. 22 Automatisierte Entscheidungen im Einzelfall einschließlich Profiling. In B. P. Paal & D. A. Pauly (Eds.), *Beck'sche Kompakt-Kommentare. Datenschutz-Grundverordnung, Bundesdatenschutzgesetz* (3rd ed.). Beck.

Martini, M., & Nink, D. (2017). Wenn Maschinen entscheiden…—Vollautomatisierte Verwaltungsverfahren und der Persönlichkeitsschutz. *Neue Zeitschrift für Verwaltungsrecht*, *36*(10), 1–14.

Matz, S. C., Bukow, C. S., Peters, H., Deacons, C., & Stachl, C. (2023). Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics. *Scientific Reports*, *13*(1), Article 5705. https://doi.org/10.1038/s41598-023-32484-w

McCrudden, C. (2008). Human dignity and judicial interpretation of human rights. *European Journal of International Law*, *19*(4), 655–724.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, *54*(6), 1–35.

Orwat, C. (2020). *Risks of discrimination through the use of algorithms*. Federal Anti-Discrimination Agency.

Pessach, D., & Shmueli, E. (2022). A review on fairness in machine learning. *ACM Computing Surveys (CSUR)*, *55*(3), 1–44.

Savcisens, G., Eliassi-Rad, T., Hansen, L. K., Mortensen, L. H., Lilleholt, L., Rogers, A., Zettler, I., & Lehmann, S. (2023). Using sequences of life-events to predict human lives. *Nature Computational Science*, *4*(1), 43–56.

Schaber, P. (2013). *Instrumentalisierung und Menschenwürde* (2nd ed.). Mentis.

Schaber, P. (2016). Menschenwürde. In A. Goppel, C. Mieth, & C. Neuhäuser (Eds.), *Handbuch Gerechtigkeit* (pp. 256–262). J. B. Metzler.

Schauer, F. (2018). Statistical (and non-statistical) discrimination. In K. Lippert-Rasmussen (Ed.), *The Routledge handbook of the ethics of discrimination* (pp. 42–53). Routlegde.

Scholz, P. (2019). DSGVO Art. 22 Automatisierte Entscheidungen im Einzelfall einschließlich Profiling. In S. Simitis, G. Hornung, & I. Spiecker genannt Döhmann (Eds.), *Datenschutzrecht. DSGVO mit BDSG*. Nomos.

Sloane, M., Moss, E., & Chowdhury, R. (2022). A Silicon Valley love triangle: Hiring algorithms, pseudo-science, and the quest for auditability. *Patterns*, *3*(2), Article 100425. https://doi.org/10.1016/j.patter.2021.100425

Smuha, N. A. (2021). Beyond the individual: governing AI's societal harm. *Internet Policy Review*, *10*(3), 1–32.

Teo, S. A. (2023). Human dignity and AI: Mapping the contours and utility of human dignity in addressing challenges presented by AI. *Law, Innovation and Technology*, *15*(1), 1–39.

Thomsen, F. K. (2017). Discrimination. In W. R. Thompson (Ed.), *Oxford research encyclopedia of politics*. Oxford University Press.

Ulgen, O. (2017). Kantian ethics in the age of artificial intelligence and robotics. *Questions of International Law*, *43*, 59–83.

Ulgen, O. (2022). AI and the crisis of the self: Protecting human dignity as status and respectful treatment. In A. J. Hampton & J. A. DeFalco (Eds.), *The frontlines of artificial intelligence ethics: Human-centric perspectives on technology's advance* (pp. 9–33). Routledge.

Valcke, P., Clifford, D., & Dessers, V. K. (2021). Constitutional challenges in the emotional AI era. In H.-W. Micklitz, O. Pollicino, A. Reichman, A. Simoncini, G. Sartor, & G. De Gregorio (Eds.), *Constitutional challenges in the algorithmic society* (pp. 57–77). Cambridge University Press.

von der Pfordten, D. (2023). *Menschenwürde* (2nd ed.). Beck.

von Ungern-Sternberg, A. (2022). Discriminatory AI and the law—Legal standards for algorithmic profiling. In S. Voeneky, P. Kellmeyer, O. Mueller, & W. Burgard (Eds.), *The Cambridge handbook of responsible artificial intelligence: Interdisciplinary perspectives* (pp. 252–277). Cambridge University Press.

Yeung, K. (2019). *Responsibility and AI. A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework*. Council of Europe.

Zarsky, T. (2013). Transparent predictions. *University of Illinois Law Review*, *2013*(4), 1503–1569.

About the Author

Carsten Orwat (PhD) is a senior researcher at the Institute for Technology Assessment and Systems Analysis, Karlsruhe Institute of Technology. Since 2000, he has worked on numerous technology assessment projects on information and communication technologies. His research also focuses on the governance and regulation of technology. Currently, he is researching the social consequences of artificial intelligence, algorithmic discrimination, and systemic risks.