BIG DATA & SOCIETY

Original Research Article

Big Data & Society
January–June: 1–14
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20539517231166886
journals.sagepub.com/home/bds
SAGE

Predictive privacy: Collective data protection in the context of artificial intelligence and big data

Rainer Mühlhoff

Abstract

Big data and artificial intelligence pose a new challenge for data protection as these techniques allow predictions to be made about third parties based on the anonymous data of many people. Examples of predicted information include purchasing power, gender, age, health, sexual orientation, ethnicity, etc. The basis for such applications of "predictive analytics" is the comparison between behavioral data (e.g. usage, tracking, or activity data) of the individual in question and the potentially anonymously processed data of many others using machine learning models or simpler statistical methods. The article starts by noting that predictive analytics has a significant potential to be abused, which manifests itself in the form of social inequality, discrimination, and exclusion. These potentials are not regulated by current data protection law in the EU; indeed, the use of anonymized mass data takes place in a largely unregulated space. Under the term "predictive privacy," a data protection approach is presented that counters the risks of abuse of predictive analytics. A person's predictive privacy is violated when personal information about them is *predicted* without their knowledge and against their will based on the data of many other people. Predictive privacy is then formulated as a protected good and improvements to data protection with regard to the regulation of predictive analytics are proposed. Finally, the article points out that the goal of data protection in the context of predictive analytics is the regulation of "prediction power," which is a new manifestation of informational power asymmetry between platform companies and society.

Keywords

Predictive analytics, data protection & privacy, data ethics, social inequality, profiling, anti-discrimination

Introduction

One of the today's most important applications of artificial intelligence (AI) technology is so-called predictive analytics. I use this term to describe data-based predictive models that make predictions about any individual based on available data. These predictions can relate to future behavior (e.g. what is someone likely to buy?), to unknown personal attributes (e.g. sexual identity, ethnicity, wealth, education level), to momentary vulnerabilities (vulnerable conditions such as frustration, depression, loneliness, financial difficulties, pregnancy, etc.), or to personal risk factors (e.g. mental or physical disease predispositions, addictive behavior, or credit risk). Predictive analytics is controversial because, although it has socially beneficial applications, the technology has an enormous potential for abuse and is currently scarcely regulated by law. Predictive analytics makes it possible to automate and, therefore, significantly scale the exploitation of individual vulnerabilities, as well as fostering unequal treatment of individuals in terms of access to economic and social resources such as employment, education, knowledge, healthcare, and law enforcement. Specifically, in the context of data protection and anti-discrimination, the application of predictive AI models needs to be analyzed as a new form of data power large IT companies wield and which relates to the stabilization and production of discriminatory structures, patterns of exploitation, and data-based societal inequalities.

Ethics of Artificial Intelligence, Institut für Kognitionswissenschaft, Universität Osnabrück, Osnabrück, Germany

Corresponding author:
Rainer Mühlhoff, Ethics of Artificial Intelligence, Institut für Kognitionswissenschaft, Universität Osnabrück, Wachsbleiche 27, 49090 Osnabrück, Germany.
Email: rainer.muehlhoff@uni-osnabrueck.de

Against the backdrop of the enormous societal impact of predictive analytics, I will argue (as others have argued before me, cf. Hildebrandt, 2009; Hildebrandt and Gutwirth, 2008; Mittelstadt, 2017; Taylor et al., 2016; Taylor, 2016; Vedder, 1999) that we need new approaches to data protection in the context of big data and AI. In my approach, I will use the concept of *predictive privacy* to normatively capture this novel form of privacy violation through *inferred* or *predicted* information. That is, applying predictive models to individuals in order to support decisions is a violation of privacy, yet it is one which does not come about either through "data theft" or a breach of anonymization. Predictive analytics proceeds according to the principle of "pattern matching" by learning algorithms that compare auxiliary data known about a target individual (e.g. usage data on social media, browsing history, geolocation data) against the data of many thousands of other users. This pattern matching is at the core of predictive privacy violations and is possible wherever there is a sufficiently large group of users disclosing their sensitive attributes alongside behavioral and auxiliary data—usually, because they are unaware that this data can be exploited using big data-based methods, or because they think they personally "have nothing to hide." As I will argue, the problem of predictive privacy denotes a limit to the liberalism inherent in contemporary views of data privacy as the individual's right to control what data is shared about them. The issue of predictive privacy thus strengthens the case for anchoring collectivist protective goods and collectivist defensive rights in data protection.

In the philosophical theories of privacy, collectivist perspectives have long taken into account that one's own data can potentially have negative effects on other people as well, and have therefore posited that individuals should not be free to decide in every respect what data they disclose about themselves to modern data companies (Hildebrandt, 2009; Hildebrandt and Gutwirth, 2008; Loi, Christen, 2020; Mantelero, 2016; Mittelstadt, 2017; cf. Regan, 2002; Taylor et al., 2016). I will also argue that large collections of anonymized data relating to many individuals should not be freely processable by data processors because predictive capacities can be extracted from anonymous data sets. This is in contrast to the current legal situation under the EU General Data Protection Regulation (GDPR), which does not restrict the processing and storage of anonymized data and the predictive models (or "profiles," to use the terminology of Hildebrandt, 2009) derived from them. Finally, I will call for the rights of data subjects as outlined by the GDPR (right of access, rectification, deletion, and so on) to be reformulated in a collectivist manner, so that affected groups and the community as a whole would be empowered, for the sake of the common good, to exercise such rights against data-processing organizations and thereby prevent the misuse of predictive capacities.

Predictive analytics

For the purpose of this article, precisely which algorithms and procedures a predictive system is based on is not pertinent. I will use predictive analytics as an umbrella term encompassing both machine learning methods and simpler statistical evaluations. While predictive analytics refers to the technological discipline, the "predictive model" refers to a specific manifestation of this technology. However, for an adequate understanding of the data protection problem, it is helpful to give a functional characterization of predictive models. Predictive models are data processing systems that receive as input a set of available data about an individual (or a "case") and output an estimate of some unknown piece of information, classification, or decision regarding the individual (hereafter referred to as the "target variable").

The input data are typically readily available auxiliary data, for example, tracking data, browser or location history, or social media data (e.g. "likes," posts, friends, group membership). The target variable is typically hard-to-access personal information on the individual or a decision about the individual relating to the business of the predictive model's operator (e.g. at what price someone is offered insurance or credit).

Hence, the goal of predictive analytics is to estimate hard-to-obtain information from easily accessible data about an individual. To do this, predictive models "pattern match" the case given by the input data against thousands or millions of other cases the model has previously analyzed, whether during a learning phase or by means of other, statistical methods. Such models are often trained by means of supervised learning methods. This requires a large amount of training data, that is, a data set in which both data fields— the auxiliary data and the target data— are recorded for a large cohort of individuals. For example, the subset of all Facebook users who explicitly state their sexual orientation in their profile produces a training data set for a model that predicts the sexual orientation of *any* Facebook user from their usage data, such as Facebook "likes" (see Figure 1).

If only a small percentage of the more than two billion Facebook users provide information about their sexual orientation, the resulting training data set will still comprise several million users. Predictive models that can be trained from this data set might then be used by the platform to estimate the sexual orientation of *all other* Facebook users, including users who have not consented to the processing of this information, have deliberately not provided it, or may be unaware that the company can estimate this information about them (Hildebrandt, 2009; Hildebrandt and Gutwirth, 2008; cf. also Skeba and Baumer, 2020).

Medical researchers at the University of Pennsylvania have shown that this approach can be used to predict whether a user suffers from conditions such as depression,

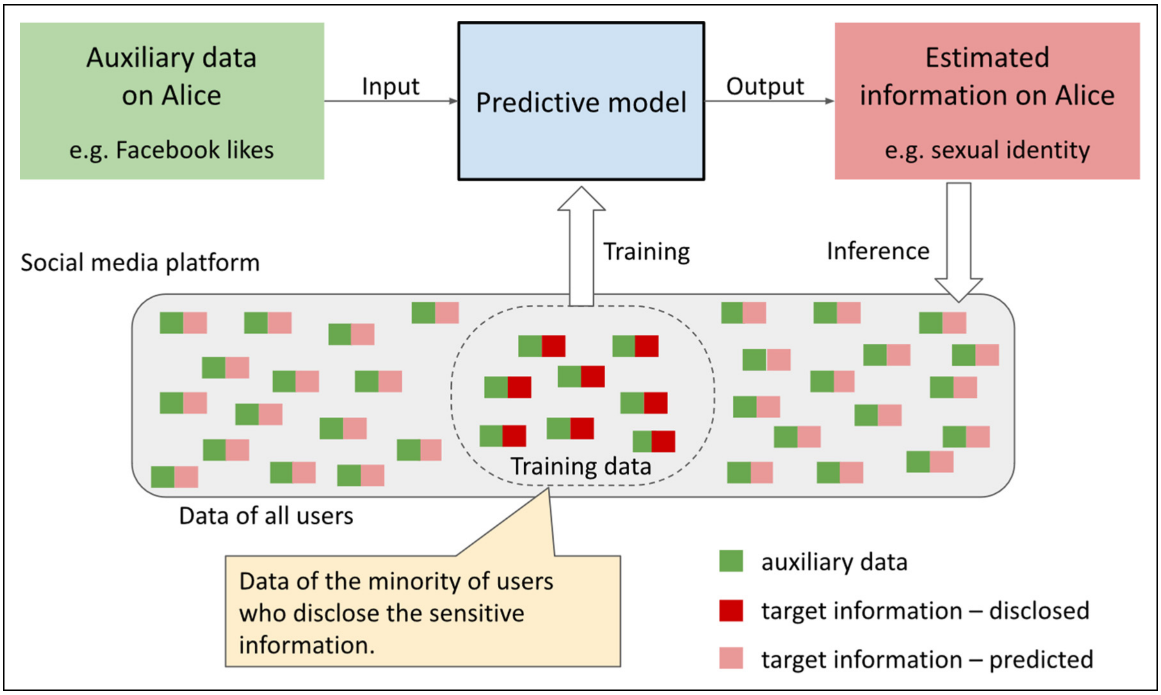

Figure 1. Schematic representation of the procedure of predictive analytics.

psychosis, diabetes, or high blood pressure (Merchant et al., 2019). Facebook has asserted that it can recognize potentially suicidal users by their postings (Goggin, 2019). A high-profile study by Kosinski et al. shows that data on Facebook "likes" can be used to predict "a range of highly sensitive personal attributes including sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender" (Kosinski et al., 2013: 5802).

Such predictive analyses are attracting growing interest from insurance and finance companies because they allow individual risk assessment beyond the classic credit-score approach (Lippert, 2014; O'Neil, 2016: ch. 8). Predictive models are also used in human resource management, for example, to carry out automated pre-selection of applicants in hiring processes, which is problematic as it allows recruitment decisions to be implicitly aligned with sensitive or protected attributes (such as race or pregnancy) through the phenomenon debated as proxy discrimination (O'Neil, 2016: 108, 148). One of the first and most common applications of predictive analytics is targeted advertising. In 2011, for example, a US supermarket chain was able to identify pregnant customers using purchase data collected through customer loyalty cards (Duhigg, 2012). In this context, predictive analytics can be used to exploit psychological, emotional, or socio-economic vulnerabilities of internet users, for example, when predictions of financial or health status are used to steer inappropriate or unfair offers to possibly vulnerable people, such as single mothers, people with low self-esteem or income, or those who have experienced the recent death of a loved one (O'Neil, 2016: 71). This harmful use of predictive analytics is related to so-called "hypernudges" (Yeung, 2017), strategies that use constantly updated and "highly personalised choice environment[s]" in online interactions to get users to click on an ad, reveal more of their data, or spend more time on a certain app. This combination of prediction and nudging brings with it a high risk of manipulation, as is also shown by the example of Facebook, which identifies emotionally vulnerable teenagers in order to target them with specific advertisements (Susser et al., 2019; cf. Zarsky, 2019).

Predictive privacy

Predictive analytics enables unknown and personal information to be estimated using readily available data about an individual or group. This is possible with modern machine learning techniques whenever many users of a digital platform provide the data basis to determine correlations between the auxiliary and target data. We thus face a situation in which the *data permissiveness* of a minority of users (e.g. Facebook users who provide information about their sexual orientation) sets the standard for the kind of information that will be predictable about *all* users of the same platform. The issue of predictive analytics has so far not been effectively addressed by data protection regulations in the EU context, with the GDPR failing to impose reasonable restrictions for the production and use of predictive models [see the “Current deficits in regulation” section; cf. Wachter (2019)].

In order to normatively anchor protection against the misuse of predicted information, I seek in this paper to construct a new protected good. In direct response to the danger posed by predictive analytics, I will make use of the concept of *predictive privacy* introduced in a previous work (Mühlhoff, 2021) to characterize this protected good. Predictive privacy, in an initial version of the idea, can be defined negatively by detailing cases when it is *violated*:

Definition 1: The predictive privacy of an individual or group is violated when personal information is predicted about them without their knowledge or against their will, in such a way that it could result in unequal treatment of an individual or group.

While the concept of predictive privacy was mainly introduced as a collective ethical value in the previous work, the present paper aims to discuss in more detail its legal and regulatory implications. As we will see, fully recognizing the value of predictive privacy requires us to break with the liberal assumptions at the root of most Western constructions of privacy[1] and data protection frameworks. This paper is in the tradition of philosophical works that have emphasized privacy as crucial not only to the individual but to society in general (Mantelero, 2016; cf. Regan, 2002; Taylor, 2016). For instance, Regan (2002) points out that privacy is not only a *common value* in that is it shared among individuals; it is also a *public value* in that it is vital to the democratic political system, and a *collective value* “in that technology and market forces are making it hard for any one person to have privacy without all persons having a similar minimum level of privacy” (Regan, 2002: 399).

The novel contribution of this paper is *not* to assert (as has already been established) that privacy is a collective value or good. In fact, my focus is not on privacy in general, but *on predictive* privacy. This is a specific—and relatively novel—aspect of privacy that has not yet been explicitly and sufficiently debated in the works mentioned above, as it has only recently become a pressing issue in the context of current technological innovations, particularly relating to the systematic and automated infringement of privacy through AI-based predictions. This new aspect of privacy, however, can only be seen from a collectivist perspective on privacy. The paper thus proposes a path by which we should extend the concept of privacy so as to include (or at least stress) a new and very contemporary battlefront in the fight against privacy infringements that remains unapparent and unprotected so long as we approach privacy from the standpoint of individual rights and damages.

It is vital to the ethical and privacy problem debated in this paper that, as I argued (in Mühlhoff, 2021), “prediction” refers to a specific epistemic operation in the real-life application of machine learning models which deviates in ethically doubtful ways from the epistemology of classical statistics. In predictive modeling, information about the data subject is derived from sophisticated, nonlinear comparisons made via machine learning models with data that many *other* data subjects have disclosed about themselves. Using the term “prediction gap,” I pointed out that a prediction translates population-based statistical inferences (which are probability distributions over a range of individual possibilities) into *single case*, or *point* predictions. The reasoning involved in this step is not warranted by the logic of traditional statistics, as it leaps from knowledge about a population to “betting” on a single case (Mühlhoff, 2021).[2] For instance, while a statistical inference might assert that at the population level there is a correlation between smoking and getting cancer, crossing the prediction gap means betting that a given individual Alice will be more likely to develop cancer based on the information that Alice is a smoker. As an important consequence, a breach of predictive privacy does *not* require the accuracy or correctness of the estimated information, but only the potential for unequal treatment of any individual or group based on that information. Under the ethical standard of predictive privacy, it would be no more legitimate to treat people differently based on predicted information merely because those predictions meet certain requirements of accuracy.

In the previous work, I also detailed why I prefer the term “predictive privacy” over various related notions. For instance, “inferential privacy” (see Loi and Christen, 2020) does not properly acknowledge the problem of the prediction gap. Loi and Christen acknowledge a privacy violation only when it arises through “logically valid inferences” and (unlike predictive privacy) do “not include the ‘non-logical’ *predictions* that result from applying statistical inferences to single cases, thus … *betting* on the possible outcome” (Mühlhoff, 2021: 679). The same objection applies to the notion of a “right to reasonable inferences” by Wachter and Mittelstadt (2019): according to predictive privacy, the use of “reasonably inferred” information also poses an ethical concern. Predictive privacy is thus a stronger demand than the right to reasonable inferences (Wachter and Mittelstadt, 2019). In a similar vein, predictive privacy differs from Hildebrandt’s case for a “paradigm shift from data to knowledge protection” in the face of privacy infringements through profiling (Hildebrandt, 2009: 247). In classical philosophy, knowledge is a true and justified belief. The violation of predictive privacy, however, does not presuppose that a prediction is genuinely valid, thus qualifying as knowledge. Finally, predictive privacy also differs from “group privacy” (cf. Floridi, 2014; Helm, 2016; Mittelstadt, 2017; Taylor et al., 2016), as it “does not tie the ethical concern to the precondition that mistreatment of an individual occurs at group scale" (Mühlhoff, 2021: 680). Predictive modeling is a novel privacy threat because it makes available a new domain of information–information that was never recorded but is only predicted about data subjects in the sense of betting on the most likely outcome. The threat of misuse of this kind of information is independent of whether the algorithms operate by virtually grouping individuals or by other means.

A new privacy problem: Differentiating three types of attacks

The specific privacy threat of predictive analytics arises under the conditions of insufficiently regulated AI and big data technology and has only been evident in the last 10 years. In order to work out the novel quality of this threat and the corresponding new need for protection, it is worthwhile to compare the new type of attack scenario with two older attack scenarios that have each played a prominent role in the discourses on data protection and privacy in past decades (see Table 1 for an overview).

Type A: Intrusion

The archetypal threat in data protection can be described as an intrusion. This type of attack is closely related to targeted surveillance focusing on specific individuals or groups. Since the proliferation of computerized data processing in the 1960s, the risk of data being stolen from more or less secure, or at least non-public, zones has been the mainstay of debates on data protection (today, protecting against this threat is known as data security). Although the main potential attacker is always the data-processing organization itself, this type of attack has in the popular imagination often been associated with hacking and cyber-attacks by criminals or intelligence agencies. The attack target of the intrusive privacy breach is sensitive data about individuals, cohorts, companies, and government processes that have not been designed to be accessible to attackers.

Type B: Re-identification

A second type of attack is known as re-identification. This type only became significant in the 1990s, after the digitalization of the healthcare system—for example, billing processes with insurance companies or patient administration in hospitals—made available extensive digital databases on healthcare processes, giving rise to the idea of using this data for statistical evaluations in the context of scientific research. This led to the issue of how one could anonymize the entries in such databases in order to be able to publish useful information without violating anyone's privacy.

In a now legendary case, the US state of Massachusetts at the end of the 1990s made the hospital treatment data of its approximately 135,000 state employees and their dependents available for research. For this purpose, the data was anonymized by deleting fields such as name, address, and social security number from the records. By means of a so-called linkage attack, Latanya Sweeney was able to identify in the anonymized data the record of then Massachusetts Governor William Weld and to reconstruct his medical records (Ohm, 2010; Sweeney, 2002). This case led to discussions in the academic and political worlds about the limits and feasibility of anonymization. The question of "secure" anonymization procedures is still being discussed today, and current proposals for anonymization procedures in computer science always seem to be cut short by far-reaching attacks.[3] It has thus become clear that "anonymity" is a multifaceted notion that cannot be defined absolutely, insofar as the degree of anonymity depends on assumptions about the background knowledge of the attacker as well as the statistical distribution of the data in the data set that is to be anonymized. Moreover, anonymization methods are required to anticipate all future attack techniques and to cover all possible configurations of background knowledge of any future attackers.

The danger of re-identification in anonymized data sets has become a second, much-discussed threat in data protection since the 1990s. This discussion has had a noticeable influence on data protection legislation in the context of

Table 1. Qualitative comparison of attack scenarios that represented a dominant threat in the public discourse on data protection at different times.

	Type A: Intrusion	Type B: Re-identification	Type C: Prediction
Most relevant since	1960	1990	2010
Means	Hacking, data leaks, breach of encryption, etc.	De-anonymization through statistical attacks or background knowledge	Prediction of unknown information by pattern matching in large data sets
Target	Sensitive/confidential data	Anonymity in large data sets that are purposefully published (e.g. for statistics)	Fairness; equality of treatment
Protection	Data security measures	Differential privacy; federated machine learning	Predictive privacy

medical data, for example on the 1996 Health Information Portability and Accountability Act (HIPPA) in the US. For the purposes of this article, it is important to point out the qualitative difference here to the attack type of intrusion (and prediction). Unlike data theft, the goal of re-identification attacks is to effect a breach of anonymity. Even though sensitive data on individuals or cohorts is obtained, this is different from intrusive data breaches, since the underlying data was deliberately published with the promise that it would not reveal individual, but only statistical, information.

Type C: Prediction

However, re-identification can no longer be viewed as the most important and dominant type of attack in data protection today. This is not to say that the principle of predicting unknown data by means of big data and AI technology makes the danger of re-identification disappear (no more than it does the danger of intrusion). However, the threat of unregulated predictive analytics far surpasses both classic attack scenarios in terms of reach and scalability (cf. Hildebrandt, 2009). Once a predictive model is created—and there are currently no effective legal restrictions on this—it can be applied to millions of users in an automated way with almost no marginal cost. The data permissiveness of some users determines what kinds of information can be estimated about almost anyone as long as predictive analytics technology remains an unregulated field.

This constitutes a qualitatively new threat in data protection because the means of violating predictive privacy is neither data theft nor a breach of anonymization. The risks originating from predictive analytics differ from the older attack scenarios in three respects. With regard to causation, the threat posed by predictive analytics relies on the availability of *collective* data sets; in terms of the perpetrator, the powers of predictive analytics are reserved for those actors who have access to aggregated collective data sets; and with regard to its effects, predictive analytics not only has an impact on the individual, it also affects society as a whole. As the most obvious consequence, the data power deriving from predictive analytics becomes commercially concentrated among a few large companies. Secondly, the potential harm of predictive privacy breaches lies in information being estimated not only about targeted individuals but about very large cohorts of users, automatically and synchronously, turning the predictive capacities of commercial actors into a structural factor in our societies. At the heart of predictive privacy violations, then, is not espionage directed at individuals, but automated and serial exploitation and unequal treatment of people throughout society, for example when we are offered different rates for insurance, or when we are shown a specific advertisement at perhaps an emotionally vulnerable time.

Predictive privacy as a protected good

The problem of predictive privacy represents the most significant new challenge for data protection at present. In order to recognize the protection of predictive privacy in its fullest extent as an issue of data protection, it is necessary to disentangle public discourse about data protection from its "roots in traditional liberal thinking" that constructs privacy as a "private good" (Regan, 2002: 397). In many Western legal traditions, the fixation on individual claims when it comes to data protection and privacy rights is enshrined by the fact that data protection serves the protection of fundamental rights, which are rights of the *individual* against the state. The protected good of predictive privacy should thus be viewed through a collectivist lens, which is premised upon a set of ethical values that prioritize the collective over the individual. It is of course still true that there remains a threat to the *single individual* that they may be treated adversely on the basis of predicted information—and we *do* need protection against this threat. But this danger alone is nothing new: long before the advent of AI-based predictive analytics, bank advisors were making decisions about creditworthiness based on stereotypes, doctors prioritized treatment programs based on personal assessments, and human resources staff were predicting the performance of job applicants during the hiring process.

Compared to these older scenarios, a broader scope to the risk arising from predictive analytics lies in the fact that these systems allow predictions to be made automatically about *any* person, thus *potentially* affecting all of us simultaneously and on a large scale. Predictive analytics technology is usually in the hands of powerful private or state actors that have an interest in algorithmically managing not isolated individuals but whole user cohorts or populations (Mühlhoff, 2020)—in other words, in categorizing and potentially exploiting the vulnerabilities of thousands of people simultaneously. The essence of predictive privacy breaches is thus not the invasion of a private "sphere," but in its capacity to open up a path for privacy to be structurally reconfigured in our digital societies. This reconfiguration concerns the technologically realistic expectations of privacy, the scalability of methods to subvert privacy, and the political values at stake when it comes to policies relating to privacy. In the context of AI and big data, threats to equality, fairness, and anti-discrimination increasingly originate from powerful private actors. Notably, anti-discrimination is at stake here in my argument not only with respect to the issues of possible biases in predictive systems but to the extent that such systems, even if "unbiased" (if that is ever possible), have significant societal consequences (cf. Eubanks, 2017; Noble, 2018).

The issue of predictive privacy thus fosters a collectivist interest in data protection which has become more relevant in the past two decades in the face of technological innovation. The aim of protecting predictive privacy is to address a growing power asymmetry between society and data-processing organizations that design, run and own predictive models. We should thus posit predictive privacy as a protected good in the name of the common good. Understood as such, the demand for protecting predictive privacy is a response to the *potential* for abuse of the predictive capacities of certain actors. Hence the purpose of predictive privacy is to bring the growing prediction power of private and sometimes public actors under democratic accountability and control.

A positive definition of the protected good of "predictive privacy" thus goes beyond the negative, harm-focused, and still implicitly individualist definition of the predictive violation of individual privacy that was introduced in Definition 1 above, following Mühlhoff (2021). I am therefore supplementing this first definition with a positively framed and more collectivist version of the same concept:

Definition 2: Predictive privacy as a protected good designates a (legal and ethical) level of protection of the community against the predictive capacities of large data processors; this is a demand for protection against a specific, contemporary, technological manifestation of informational power asymmetry.

From this more nuanced viewpoint, predictive privacy is about regulating a technology that can structurally harm many of us, and thus our society in general, when it comes to values of equality, fairness, and human dignity. A violation of predictive privacy—and there is a subtle semantic difference between "violation of predictive privacy" and "predictive violation of (individual) privacy"—refers to a political, economic, and technological constellation that furthers social inequality, automated exploitation of individual vulnerabilities and data-based socio-economic selection on a structural level through the use of predictive models. Adopting predictive privacy as a protected good shifts the focus in data protection from the defensive rights of the individual to a preventive protection of the community against a new form of technologically enabled social and political power asymmetry—*prediction power*.

This shift from the first to the second definition also helps us emphasize the preventive aspect. Protecting society against the power asymmetries enabled by predictive modeling requires legal interventions already during the stage of *potential* (or looming) violations of privacy. We thus need a legal and ethical framework to govern the emergence and use of predictive powers. In demanding mechanisms to protect predictive privacy that are already in place before an actual violation of an individual's privacy, I refer here to the legal concept of data protection as a type of "protection in advance."[4] The accumulation of prediction power in the hands of particular companies is what predictive privacy measures should address; protective mechanisms should not kick in only after predictive power has actually been exercised in specific cases. Predictive privacy as a protected good is, therefore, necessary to guarantee that the fundamental values of free, egalitarian, and democratic societies are maintained in the face of a technological situation that otherwise impels us to increasingly entrust the maintenance of these self-same values to the putatively benign intentions of global economic actors.

Predictive privacy as a collective duty

In addition to the collectivist construction of the protected good of predictive privacy, the violation of predictive privacy is also characterized by a collective "perpetration" or causation. This is because predictive analyses are only possible when two conditions are met: first, a sufficiently large group of users provides their sensitive data in connection with auxiliary data when using digital services. Secondly, platform companies and other economic actors are legally permitted to aggregate this data (potentially also in anonymized form) and use it to train predictive models. Given these preconditions, protecting predictive privacy requires nothing less than a departure from the deeply entrenched, liberalist way that many people think about the ethics of our everyday use of networked digital services, whereby we also believe it is everyone's own business to decide whether platform companies are allowed to siphon off their personal data. Protecting predictive privacy demands a broadly shared awareness that one's own data potentially harm others—and that modern data protection means more than simply giving control to each single user over what personal data is collected from them. To fully internalize this point, a useful approach might be to reverse this argument: that the data which many others more or less knowingly and voluntarily disclose about themselves (and which is collected by platform companies perfectly legally) can be used to estimate sensitive information about oneself.[5]

These elementary observations about the social externalities of one's data practices,[6] externalities that arise from the basic technical structure of predictive analytics, reveal a significant limit to the legal basis of consent which in the context of social media is one of the most widely used tools provided by the EU GDPR (EU, 2016) for platform companies to lawfully aggregate large sets of user data. Here it becomes clear that when a user is asked for consent, they are making a decision on behalf of many other people who can be discriminated against on the basis of this data—provided that a number of other users also disclose such data about themselves, of course, but this is usually the case, as the numerous examples from social media show. In our current legal and regulatory landscape, in which the construction and the use of predictive models is broadlyunregulated, *individual* consent decisions are of *supra-individual* scope, not limited to the data subject itself.

In this context, it should be noted that anonymous data is sufficient for the training of predictive models. One only needs the correspondence of auxiliary data and target information—for example, Facebook “likes” and information about health conditions; the training data for predictive analytics does not need to contain identifying data fields. Promises of anonymization are therefore routinely leveraged to minimize users’ reluctance about consenting to the processing of their sensitive data; anonymization is an innocuous requirement for big data business models that are based on predictive analytics.[7] In situations where users are not using a digital service anonymously, it is likely that platform companies can still avoid specifying the training of predictive analytics as the purpose of data processing, because they can anonymize the data directly after collection and then make further use of it. The reason for this is that anonymized data does not fall within the scope of the GDPR and can be freely used—especially in aggregated form.[8] It can also be stored indefinitely and only later be used for predictive analytics. Finally, it should be borne in mind that the trained predictive models themselves represent derived, highly aggregated, anonymized data,[9] which thus do not fall within the scope of the GDPR and, in particular, can be sold and circulated without effective data protection constraints.

Current deficits in regulation

Predictive analytics and AI technology have significantly increased the potential for misuse of anonymized mass data over the past 15 years (see Table 1). However, in the current legal situation, the generating (and to a lesser extent the use) of predictive models is largely unregulated, so the possibility of misuse is a potentially serious societal issue that can stabilize help to produce and reinforce socioeconomic inequalities and patterns of discrimination.

Producing predictive models

First, the question arises as to why the EU’s GDPR framework does not effectively regulate the production of predictive models. One reason lies in the individual-oriented normative conception of the GDPR, which is ultimately based on the conception of fundamental rights as individual rights. It is, moreover, a characteristic of the public discourse, case law, and business practices developing around the GDPR that both the protected goods and the defensive rights of data protection are always focused on someone’s relationship to their own data. The interpretation is usually that the sovereignty of individuals in relation to the use of their (personal) data must be preserved; everyone must be asked for consent in relation to their own data or another legal basis must be declared. Acts of infringement under the GDPR, therefore, refer to an individual who claims that *their* personal data were processed in a way that was not covered by the claimed legal basis. In particular, the defensive rights of data subjects, such as the right of access (Art. 15), rectification (Art. 16), erasure (Art. 17), restriction of processing (Art. 18), and portability (Art. 20), are framed in the GDPR as individual rights that can only be exercised by the individual in relation to their own data.

Another, related reason why the GDPR weakly regulates the production of predictive models is that it refers to “personal data” (Art. 4(1)) and does not concern itself with anonymous data (Wachter, 2019). The distinction between personal and anonymous data is outdated in the context of AI and big data. This is not merely because anonymization can be broken, e.g. leveraging background knowledge,[10] but because predictive analytics can use the anonymized data of *many* individuals to estimate sensitive and “personal” data about *other* people whose data was never recorded and thus never anonymized. In the reality of data businesses, the distinction between personal and anonymous data is often only made at the “input stage” of data processing, for instance, to decide whether certain data has been legally recorded by the system (Wachter and Mittelstadt, 2019: 125f.) even though, according to Art 4(1) GDPR, all stages of data processing should be taken into account. Furthermore, the evaluation of data as personal versus anonymous in practice only considers the relation of the data in question to the *one* data subject from whom the data is collected.[11] The fact that the anonymized data of *many* data subjects enable a new kind of privacy violation against arbitrary *others* remains unrecognized in this scheme. Information derived in the course of data processing can thus undermine the initial distinction between anonymous versus personal data, not only insofar as supposedly anonymous data could be linked back to the data subject to whom they referred before anonymization, but rather because new insights into *any* third person can be gained by combining the anonymized data of many. The notion of “personal data” in this case would have to refer to variable individuals—in particular, third persons—and is therefore obsolete as a concept.

The legal and theoretical judgment of the danger posed by derived data is controversial and inconsistent. The German Federal Constitutional Court already argued in the 1983 census ruling that there is no such thing as “irrelevant data” (Bundesverfassungsgericht, 1983: 34; Wachter and Mittelstadt, 2019: 125)—but the focus here was not on the mass data scenario, which did not exist at the time, but on the derivation of sensitive information about a particular individual from seemingly less sensitive or anonymized data about the same individual (attack type B). The former *Article 29 Working Party* has recommended in various opinions to include derived information under personal data according to Art. 4 GDPR (Article 29 Data Protection Working Party, 2018). However, what has been insufficiently addressed in its guidelines and opinions is the phenomenon of anonymous mass data as opposed to the danger of re-identification. With regard to the categorization of data (as discussed above, e.g., as anonymous vs. personal), the *Article 29 Working Party* progressively advocates looking at processing purposes and consequences rather than at reference to individuals at the input stage (Article 29 Data Protection Working Party, 2007; Wachter and Mittelstadt, 2019: 126). Wachter and Mittelstadt argue that the European Court of Justice has clarified in several rulings that the scope of the GDPR is limited to the "input stage" of data processing (Wachter and Mittelstadt, 2019: 6) and that the defense against the consequences of data processing, also with regard to automated decisions, must be based on sector-specific regulations [Wachter and Mittelstadt, 2019: 7]. With the instrument of the data protection impact assessment, the GDPR provides a mechanism that can explicitly include the consequences of data processing even beyond the "input stage" and thus in particular also with regard to the effects of anonymized mass data. However, even this comparatively unwieldy instrument is likely to be limited by the distinction between anonymized and personal data. In particular, according to the current interpretation of the right to erasure, this right can also be satisfied by anonymizing data records.[12] This opens a loophole for the unlimited and unregulated processing of formerly personal data beyond the purpose limitation, for example for the training of predictive models, to the extent that anonymized data provides sufficient material for this.

Using predictive models

The second question is why the GDPR does not effectively regulate the *use* of predictive models—that is, the application of already existing and trained models to individuals. This relates to the question of whether the GDPR sufficiently protects against the production and use of personal predictions. According to Sandra Wachter and Brent Mittelstadt,

> Compared to other types of personal data, inferences are effectively "economy class" personal data in the General Data Protection Regulation ("GDPR"). Data subjects' rights to know about (Art. 13–15), rectify (Art. 16), delete (Art. 17), object to (Art. 21), or port (Art. 20) personal data are significantly curtailed when it comes to inferences, often requiring a greater balance with the controller's interests (e.g., trade secrets or intellectual property) than would otherwise be the case. Similarly, the GDPR provides insufficient protection against sensitive inferences (Art. 9) or remedies to challenge inferences or important decisions based on them (Art. 22(3)). (Wachter and Mittelstadt, 2019: 6)

Wachter and Mittelstadt see evidence, in the case law of the European Court of Justice of recent years, that derived information about individuals does not have to be fully treated as "personal data" under the GDPR with respect to the legal consequences of data processing [Wachter and Mittelstadt, 2019: 5ff., 105ff.]. In this point the *California Consumer Privacy Act* (CA, 2018), which was adopted in 2018 and came into force in 2020, makes a clearer distinction (cf. Blanke, 2020: 99). The CCPA offers a definition of "personal information" which explicitly includes, in addition to various directly personal kinds of data,

> Inferences drawn […] to create a profile about a consumer reflecting the consumer's preferences, characteristics, psychological trends, predispositions, behavior, attitudes, intelligence, abilities, and aptitudes. (CCPA § 1798.140 (o))

In this context, it should also be mentioned that the regulation of profiling and automated decisions by the GDPR (see Art. 22) is too weak because it is limited to fully automated processing. Procedures that treat people differently by means of predictive models can be implemented comparatively easily as semi-automated routines by integrating human supervision and intervention possibilities (e.g. by what are known as click workers) into the processing cycle in order to circumvent the provisions of Art. 22.

A third reason for the effectively weak regulation of the use of predictive models is that the hurdle of consent is psychologically low for the gathering of auxiliary data—that is, the data on the target individual needed as input for the inferential use of a predictive model. Most users consent without hesitation to the processing of such data because behavioral data such as Facebook "likes" seem to them to be less sensitive. Moreover, this data is often collected routinely and without specific consent during people's everyday social media use.

Proposals for regulation

Alongside the previous discussion of deficits in regulation based on the EU GDPR, this section offers some conceptual proposals to inform future discussions on how to improve the regulation of predictive analytics in the context of data protection legislation. According to the principle of data protection as a type of "protection in advance,"[13] its purpose is to establish a preventative safeguard for equality and fairness in how we are treated by data-processing entities. The aim is to balance a power asymmetry between society and organizations; this asymmetry already exists in the potential and looming violation of predictive privacy, as well as in the unequally distributed vulnerability of different groups and people with regard to the potential for misuse of anonymized mass data and predictive models.

The protectiveness of a data protection regulation that effectively limits the risks of abuse of predictive analytics cannot be placed solely on the shoulders of the defensive rights of affected individuals. Such instruments always lag behind actual incidents of infringement and their effectiveness is further weakened in the present context by the fact that the violations are often difficult to identify and prove from the individual's perspective. While individual legal recourse, therefore, holds little promise of success, due to a dispersal effect resulting when predictive analytics technology is automatically applied to thousands of people in parallel, the damage to society as a whole can be considerable (cf. Ruschemeier, 2021).

Derived information

First, and analogous to California's CCPA, any future regulation should avoid, with respect to the legal consequences of data processing, drawing an effective difference between personal information that is being *recorded* (e.g. explicitly stated demographic information such as gender) versus *inferred* (e.g. estimations about purchasing power, predisposition to illness, substance abuse, etc.). In this regard, it is vital that protective mechanisms do *not* depend on whether the information is factually correct (e.g. being denied a job because of a prediction that the applicant might be predisposed to depression violates the applicant's privacy, regardless of whether that prediction turns out to be accurate or inaccurate). In contrast to the proposal of a "right to reasonable inferences" by Wachter and Mittelstadt (2019), equipping data protection with instruments to protect individuals from false or inaccurate predictions alone would not solve the present problem, which is the need to balance the power asymmetries between global data aggregators and societies. As argued previously, even accurate predictions can be used and misused in ways that are harmful to society and individuals (e.g. denying insurance to people whom the provider predicts might be predisposed to certain health conditions). Providing protection against this should be within the remit of the legal implementation of predictive privacy.

Anonymous data and purpose limitation for trained models

To safeguard society from companies' capacities to accumulate prediction power in an uncontrolled way, future regulation should be designed to apply to cover anonymized mass data by way of data protection principles similar to those enshrined in the GDPR. This does *not* mean, as the proposal is often misunderstood to imply, categorically prohibiting the processing of anonymized data but, in the same way as with personal data, to legally subject this processing to reasonable obligations like purpose limitation. The legal basis of consent is questionable here if the consequences of data processing potentially affect third parties (see below). A list of valid legal bases should ideally be drawn up as a result of a political, societal and ethical debate regarding which uses of anonymized mass data are considered socially beneficial as opposed to harmful. For instance, legal regulation could enable the processing of anonymized data specifically for health research, spam filtering, content moderation, and other measures. Importantly, compiling such a list of lawful purposes should be connected to a principle of purpose limitation for the processing of anonymized data that curtails the possibility of abuse of data or the resulting models for secondary purposes.

Given the immense potential for misuse of anonymized mass data and the ensuing predictive models, processing of anonymized data ought not to be allowed without restriction, nor occur in an unregulated field of business outside the reach of the GDPR's stipulations. My proposal is similar to the seemingly radical step of the GDPR to generally forbid the processing of *personal* information—with reasonable exceptions that enable research and business to continue to take place. Furthermore, the reasoning behind this proposal is to avoid focusing data protection regulation solely on the danger of re-identification of individuals in anonymized data sets (the type B attack scenario). Rather, to mitigate the risks of attack type C, regulation must start from the potential for abuse of *large collections* of anonymized data and of data sets in which various, more or less sensitive, data fields can be examined for correlations. A growing societal acknowledgement of the rich information content of anonymized mass data is needed for this regulatory concern, so that the problem is not reduced in public and political debate solely to the risks of re-identification.

The regulation of the processing of anonymized data must not be limited to the input stage of data processing. It should also be kept in mind that trained predictive models themselves represent aggregated, anonymized data.[14] Future regulation ought therefore to cover the dissemination and use of trained machine learning models and instil a principle of purpose limitation also with respect to trained models. Predictive models generated from customer data sets can at the moment circulate or be sold relatively unrestrictedly. Suitable mechanisms of purpose limitation and supervision should be devised in political, ethical, and legal discussions that ensure that the use of trained models is really limited to purposes beneficial to society.[15] This could include registering and preauthorizing the creation of predictive models with an appropriate supervisory body.

Restricting consent

A third direction concerns consent as a legal basis. Since in the context of big data and AI technology the processing of one's own data generally has an impact on others, the validity of individual consent is questionable. It might be appropriate to restrict the availability of consent as a legal basis only to situations when the consequences of the consent decision affect exclusively the person giving their consent.

Consent is today one of the most practically relevant legal bases. Arguably, the current practices around consent, largely based on pop-up dialogues and cookie banners, are shaping our perceptions in that each new consent dialogue affirms the liberal story that reduces data protection to individual choices about the sharing of personal data. The current practice of consent could therefore even be harmful with regard to the data protection challenges set out in this paper as it distracts people's attention from the collective dangers of data accumulation and predictive analytics (Kröger et al., 2021; Sloan and Warner, 2014). Furthermore, it has been much discussed that consent dialogues tend not to inform users properly, but tend to dupe or coerce them into giving their consent by means of design tricks, nudges and interminable small print as well as often popping up at the most inconvenient moments (cf. Baruh and Popescu, 2017; Mühlhoff, 2018).

An updated legislation should therefore render the legal basis of consent insufficient to data processing that links a person's data with the data of other users or which feeds into the production of predictive models. Other legal frameworks should be put in place to ensure that big data applications are beneficial rather than potentially harmful (see above).

Collective defense rights

Establishing collectivist counterparts to the data subjects' rights under the GDPR is another key proposal. This means that the rights of access, rectification, erasure, portability, etc. should be collectivistically extended so that, for example, groups affected by discrimination, as well as society in general, are empowered to demand information from platform operators about predictive models and the processing of anonymized data.[16] So the idea, to be fleshed out in future research, is to create a defensive right that can already be applied to the trained predictive model, rather than when the model is applied to a specific person to derive a prediction. Such an extension of existing data protection regulation should afford democratic institutions more control over what information commercial organizations can derive about any individuals from auxiliary data and what predictive models an organization trains on the basis of the data of many users.[17] This collective right of access should help to uncover which patterns of discrimination are inscribed in the predictive models. A collective right to correction or deletion of such models should be exercised once patterns of exclusion and discrimination, or stabilizing and reinforcing effects in relation to social inequality, can be observed. For the exercise of these collective rights of defense, supervisory bodies as well as appropriate instruments of collective redress such as class actions should be provided for (cf. in detail Ruschemeier, 2021).

Conclusion

I have argued that the novel potentials of predictive analytics also come with novel risks to society that result from the accumulation of prediction power, which is a new manifestation of informational power asymmetry between a few economic players, individuals, and society at large. If unregulated, prediction power will likely lead to a further increase in social and economic disparities, discrimination, and algorithmic social sorting across the globe. To make the risks of predictive analytics debatable ethically and politically, I have adopted the ethical concept of "predictive privacy" (Mühlhoff, 2021), which describes predicted personal information as a potential vehicle of privacy violations. Extending this ethical interpretation, I argued that to socially control prediction power, predictive privacy should be considered as a protected good. Current data protection regulations do not suffice to reasonably constrain prediction power as they do not regulate the production and circulation (only the application) of predictive models. Given the specifically collective risks of big data and AI technology, which manifest themselves more in the cementing of social inequality than in attacks on isolated individuals, effective data protection in the current decade will need to be assessed regarding the degree to which it forms a sustainable alliance with anti-discrimination measures. The field of predictive modeling based on anonymized mass data that all of us generate for big data companies every day (for free) is at the moment largely unregulated. In order to recognize the need for regulation, data protection (and especially liberal discourses surrounding privacy) must move away from its preferred focus of reference, the protection of an individual's informational sphere, and focus on the effects of contemporary data processing as they increasingly structure our societies.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Rainer Mühlhoff https://orcid.org/0000-0002-3936-9919

Notes

1. For established definitions of privacy in philosophical discourse, see Tavani (2007). Two of the main traditions in privacy appear in the anglophone world as "non-intrusion theory" and as "control theory" (on the latter, see Westin, 1967). Since the 2000s, "privacy as contextual integrity" by Nissenbaum (2011) has provided a refined conceptualization of privacy which has been influential in the US-American discourse. Why traditional frameworks are unsuitable for addressing the novel privacy challenge of predictive analytics has been discussed by in Skeba and Baumer (2020).
2. This clearly relates to a fundamental paradigm shift in statistics currently driven by the dominance of Machine Learning, in which Bayesian statistics replaces classical statistics, see Joque (2022).
3. Cf. Ohm (2010) and as examples, see the spectacular re-identification of Netflix users in a pseudonymously published database of film ratings by Narayanan and Shmatikov (2008), or the reconstruction of the family names from anonymously available genome data by Gymrek et al. (2013).
4. German *Vorfeldschutz*, cf. Britz (2010) and Von Lewinski (2009, 2014).
5. In this proposal for a rhetoric which would publicly communicate the concern for predictive privacy, the collectivist concern for protection against *violations of predictive privacy* is pragmatically retranslated in terms highlighting the threat of *predictive violation of* (*individual*) *privacy*. I see this oscillation between the terminology of the common good (protecting predictive privacy) and the individual interest (avoiding disadvantages from predictive violations of one's privacy) as quite pragmatic in terms of the persuasiveness of the argument, even among those who are less collectivist in their political sensibilities.
6. In this sense see also the concept of "data pollution," Ben-Shahar (2019).
7. See in particular the research on differential privacy in machine learning, cf. Abadi et al. (2016) and Dwork (2006).
8. The right to erasure in the context of the GDPR can also be fulfilled by anonymising the data, cf. the "Current deficits in regulation" section.
9. This presupposes that established anonymization procedures are used, which have been developed for this purpose for fifteen years under keywords such as differential privacy and differentially private machine learning.
10. This is of course also a problem, which would correspond to type B of the attack scenarios; however, this issue is not my focus here, as I argue that there is now a new data protection threat (type C).
11. For example, when a social media app accesses the phone book of a smartphone, only the smartphone owner consents to the processing of these data, not all the people listed in the phone book.
12. See the decision of the Austrian Data Protection Board (Österreichische Datenschutzbehörde, 2018). See also the website of the European Commission: https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/dealing-citizens/do-we-always-have-delete-personal-data-if-person-asks_en (last visit: 2022-03-10).
13. See above, note 4.
14. Such models are represented by millions of entries in a large matrix calibrated in the training procedure of simulated neural networks. These parameters are themselves derived data and, as long as the training procedure meets certain technically well-defined requirements, no individual entries of the training data can be reconstructed from them, so they are formally anonymous data. See the discourse on differential privacy in machine learning, note 7 above.
15. Ideas in this direction recently became visible in the European Commission's plans for a European Health Data Space: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_2711 (accessed 2022-12-20).
16. See also similar proposals by Mantelero (2016) and Pohle (2016).
17. This proposal deliberately remains in the realm of data protection. This is in contrast to more ambitious debates in political philosophy that propose alternative democratic "regimes" for collectivist governance of data use and infrastructure under terms such as "data-owning democracy" and "digital socialism" (Fischli, 2022; cf. Muldoon, 2022). These approaches explicitly go beyond data protection (Muldoon, 2022: 3) to fundamentally reorganise and communise "data ownership," data infrastructure, digital participation, etc. While I believe that effective protection of predictive privacy requires the creation of appropriate democratic institutions (such as the supervisory authority mentioned in this article), it is far from automatic that communised data processing would be aligned with this goal. The protection of predictive privacy is neither positively nor negatively enshrined in the concept of communitarian data infrastructures. It is therefore quite conceivable that community-driven data cooperatives could train predictive models, which could subsequently be used to discriminate against minorities.

References

Abadi M, Chu A, Goodfellow I, et al. (2016) Deep learning with differential privacy. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security - CCS'16, pp.308–318. DOI:10.1145/2976749.2978318.

Article 29 Data Protection Working Party (2007) *Opinion 4/2007 on the concept of personal data*. 01248/07/EN WP 136. Available at: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp140_en.pdf.

Article 29 Data Protection Working Party (2018) *Guidelines on automated individual decision-making and Profiling for the purposes of Regulation 2016/679*. 17/EN WP251rev.01. Available at: https://ec.europa.eu/newsroom/article29/items/612053/en.

Baruh L and Popescu M (2017) Big data analytics and the limits of privacy self-management. *New Media & Society* 19(4): 579–596.

Ben-Shahar O (2019) Data pollution. *Journal of Legal Analysis* 11: 104–159.

Blanke JM (2020) Protection for "Inferences Drawn": A comparison between the general data protection regulation and the California Consumer Privacy Act. *Global Privacy Law Review* 1(2): 81–92.

Britz G (2010) Informationelle Selbstbestimmung zwischen rechts wissenschaftlicher Grundsatzkritik und Beharren des Bundes verfassungsgerichts. In: Edmund B, Martin E, Bernd H, et al.

(eds) *Offene Rechtswissenschaft: ausgewählte Schriften von Wolfgang Hoffmann-Riem mit Begleitenden Analysen*. Tübingen: Mohr Siebeck, pp. 561–596.

Bundesverfassungsgericht (1983) *BVerfG, Urteil des Ersten Senats vom 15. Dezember 1983 – Zur Verfassungsmäßigkeit des Volkszählungsgesetzes* 1983. *1 BvR* Bundesverfassungsgericht. https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/DE/1983/12/rs19831215_1bvr020983.html.

CA (2018) California consumer privacy act. *Cal. Legis. Serv. Ch. 55 (A.B. 375 west)*. https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5.

Duhigg C (2012) How companies learn your secrets. *The New York Times*. https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.

Dwork C (2006) Differential privacy. In: Bugliesi M, Preneel B, Sassone V, et al. (eds) *Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10–14, 2006, Proceedings, Part II Lecture Notes in Computer Science 4052*. Berlin and Heidelberg: Springer, pp. 1–12.

EU (2016) Regulation 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). *OJ 2016 L 119/1*. http://data.europa.eu/eli/reg/2016/679/2016-05-04/eng.

Eubanks V (2017) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, First Edition New York, NY: St. Martin's Press.

Fischli R (2022) Data-owning democracy: Citizen empowerment through data ownership. *European Journal of Political Theory*: 14748851221110316. DOI:10.1177/14748851221110316.

Floridi L (2014) Open data, data protection, and group privacy. *Philosophy & Technology* 27(1): 1–3.

Goggin B (2019) Inside Facebook's suicide algorithm: Here's how the company uses artificial intelligence to predict your mental state from your posts. *Business Insider*. https://www.businessinsider.com/facebook-is-using-ai-to-try-to-predict-if-youre-suicidal-2018-12.

Gymrek M, McGuire AL, Golan D, et al. (2013) Identifying personal genomes by surname inference. *Science (New York, N.Y.)* 339(6117): 321–324.

Helm P (2016) Group privacy in times of big data. A literature review. *Digital Culture & Society* 2(2): 137–152.

Hildebrandt M (2009) Who is profiling who? Invisible visibility. In: Gutwirth S, Poullet Y, Hert De P, et al. (eds.) *Reinventing Data Protection?* Dordrecht: Springer Netherlands, 239–252. DOI:10.1007/978-1-4020-9498-9_14.

Hildebrandt M and Gutwirth S (eds) (2008) *Profiling the European Citizen: Cross-Disciplinary Perspectives*. New York: Springer.

Joque J (2022) *Revolutionary Mathematics: Artificial Intelligence, Statistics and the Logic of Capitalism*. London New York: Verso.

Kosinski M, Stillwell D and Graepel T (2013) Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences* 110(15): 5802–5805.

Kröger JL, Lutz OH-M and Ullrich S (2021) The myth of individual control: Mapping the limitations of privacy self-management. *SSRN Electronic Journal*. DOI:10.2139/ssrn.3881776.

Lippert J (2014) ZestFinance issues small, high-rate loans, uses big data to weed out deadbeats. *Washington Post*. https://www.washingtonpost.com/business/zestfinance-issues-small-high-rate-loans-uses-big-data-to-weed-out-deadbeats/2014/10/10/e34986b6-4d71-11e4-aa5e-7153e466a02d_story.html.

Loi M and Christen M (2020) Two concepts of group privacy. *Philosophy & Technology* 33: 207–224.

Mantelero A (2016) Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection. *Computer Law & Security Review* 32(2): 238–255.

Merchant RM, Asch DA, Crutchley P, et al. (2019) Evaluating the predictability of medical conditions from social media posts. *PLOS ONE* 14(6): e0215476.

Mittelstadt B (2017) From individual to group privacy in big data analytics. *Philosophy & Technology* 30(4): 475–494.

Mühlhoff R (2018) Digitale Entmündigung und User Experience Design: Wie digitale Geräte uns nudgen, tracken und zur Unwissenheit erziehen. *Leviathan – Journal of Social Sciences* 46(4): 551–574.

Mühlhoff R (2020) Automatisierte Ungleichheit: Ethik der Künstlichen Intelligenz in der biopolitischen Wende des Digitalen Kapitalismus. *Deutsche Zeitschrift für Philosophie* 68(6): 867–890.

Mühlhoff R (2021) Predictive privacy: Towards an applied ethics of data analytics. *Ethics and Information Technology* 23: 675–690.

Muldoon J (2022) Data-owning democracy or digital socialism? *Critical Review of International Social and Political Philosophy* 0(0): 1–22. DOI:10.1080/13698230.2022.2120737.

Narayanan A and Shmatikov V (2008) Robust de-anonymization of large sparse datasets. In: 2008 IEEE Symposium on Security and Privacy (sp 2008), Oakland, CA, USA, 2008, pp. 111–125. IEEE. DOI:10.1109/SP.2008.33.

Nissenbaum H (2011) A contextual approach to privacy online. *Daedalus* 140(4): 32–48.

Noble SU (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.

Ohm P (2010) Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review* 57: 1701–1777.

O'Neil C (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, First edition New York: Crown.

Österreichische Datenschutzbehörde (2018) *Datenschutzbeschwerde von Dr. Xaver X*. DSB-D123.270/0009-DSB/2018. https://www.ris.bka.gv.at/Dokumente/Dsk/DSBT_20181205_DSB_D123_270_0009_DSB_2018_00/DSBT_20181205_DSB_D123_270_0009_DSB_2018_00.html.

Pohle J (2016) Personal data not found: Personenbezogene Entscheidungen als überfällige Neuausrichtung im Datenschutz. *Datenschutz Nachrichten*. https://www.datenschutzverein.de/wp-content/uploads/2016/04/DANA_16_1_Heft.pdf.

Regan PM (2002) Privacy as a common good in the digital world. *Information, Communication & Society* 5(3): 382–405.

Ruschemeier H (2021) Kollektiver rechtsschutz und strategische prozessführung gegen digitalkonzerne. *MMR* 24(12): 942–946.

Skeba P and Baumer EP (2020) Informational friction as a lens for studying algorithmic aspects of privacy. In: *Proceedings of the*

ACM on Human-Computer Interaction. New York: ACM, pp. 1–22.

Sloan RH and Warner R (2014) Beyond notice and choice: Privacy, norms, and consent. *Journal of High Technology Law* 14(2): 370–414. https://heinonline.org/HOL/P?h=hein.journals/jhtl14&i=371.

Susser D, Roessler B and Nissenbaum H (2019) Online manipulation: Hidden influences in a digital world. *Georgetown Law Technology Review* 4(1): 1–52.

Sweeney L (2002) k-Anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10(05): 557–570.

Tavani HT (2007) Philosophical theories of privacy: Implications for an adequate online privacy policy. *Metaphilosophy* 38(1): 1–22.

Taylor L (2016) The ethics of big data as a public good: which public? Whose good? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374(2083). DOI:10.1098/rsta.2016.0126.

Taylor L, Floridi L and van der Sloot B (2016) *Group Privacy: New Challenges of Data Technologies*. New York: Springer Berlin Heidelberg.

Vedder A (1999) KDD: The challenge to individualism. *Ethics and Information Technology* 1(4): 275–281.

von Lewinski K (2009) Geschichte des Datenschutzrechts von 1600 bis 1977. In: *Freiheit – Sicherheit – Öffentlichkeit: 48. Assistententagung Öffentliches Recht, Heidelberg 2008*. Baden-Baden: Nomos, pp. 196–220. DOI: 10.5771/9783845215532-196.

von Lewinski K (2014) *Die Matrix des Datenschutzes Besichtigung und Ordnung eines Begriffsfeldes*. Tübingen: Mohr Siebeck. http://public.eblib.com/choice/PublicFullRecord.aspx?p=6624481.

Wachter S (2019) Data protection in the age of big data. *Nature Electronics* 2(1): 6–7.

Wachter S and Mittelstadt B (2019) A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. *Columbia Business Law Review* 2019(1): 1–130.

Westin AF (1967) *Privacy and Freedom*. New York: Atheneum Press.

Yeung K (2017) “Hypernudge”: Big Data as a mode of regulation by design. *Information, Communication & Society* 20(1): 118–136.

Zarsky TZ (2019) Privacy and manipulation in the digital age. *Theoretical Inquiries in Law* 20(1): 157–188.