Predictive Privacy: Collective Data Protection in the Context of AI and Big Data

Rainer Mühlhoff <rainer.muehlhoff@uni-osnabrueck.de>

Pre-Print manuscript, currently under review.

Manuscript version: 2022-03-28

Big data and artificial intelligence (AI) pose a new challenge for data protection. This is because these techniques are used to make predictions about third parties based on the anonymous data of many people, for example about purchasing power, gender, age, sexual orientation, ethnicity, the course of an illness, etc. The basis for such applications of "predictive analytics" is a comparison of behavioural data (e.g. usage, tracking or activity data) of the individual in question with the potentially anonymously processed data of many others using machine learning models or simpler statistical methods.

The article first points out that there is considerable potential for abuse associated with predictive analytics, which manifests itself as social inequality, discrimination and exclusion. These potentials for abuse are not regulated by current data protection law (EU GDPR); in fact, the use of anonymised 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. The predictive privacy of a person or group is violated when sensitive information about them is predicted based on the data of many other individuals without their knowledge and against their will. Predictive privacy is then formulated as a collectivist protected good of data protection and various improvements of the GDPR with regard to the regulation of predictive analytics are proposed.

1. Introduction

One of the currently most important applications of 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 behaviour (e.g. what is someone likely to buy?), to unknown personal attributes (e.g. sexual identity, ethnicity, wealth, education level) or to personal risk factors (e.g. mental or physical disease predispositions, addictive behaviour or credit risk). Predictive analytics is controversial because, though it has socially beneficial applications, the technology has an enormous potential for abuse and is currently barely regulated by law. Predictive analytics makes it possible to automate and therefore significantly scale 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 analysed as a new form of data power which large IT companies yield and which relates to the

stabilisation and production of discriminatory structures, social stratification and data-based social inequality.

Against the backdrop of the enormous societal impact of predictive analytics, I will argue in this article that we need new approaches to data protection in the context of Big Data and AI. I will use the term *predictive privacy* to normatively capture the 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 through either "data theft" or a breach of anonymisation. Predictive analytics proceeds according to the principle of “pattern matching”, comparing auxiliary data known about a target individual (e.g. usage data on social media, browsing history, geo-location 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 behavioural and auxiliary data – usually because they are not aware this data can be exploited using Big Data-based methods, or think they personally “have nothing to hide”. As such, the problem of predictive privacy denotes a limit to the liberalism widespread in contemporary understandings of data privacy as the individual right to control what data is shared about oneself and helps anchor collectivist protective goods and collectivist defensive rights in data protection.

Such a collectivist perspective in data protection firstly takes into account that individuals should not be free to decide in every respect what data they disclose about themselves to modern data companies, because one's own data can potentially have negative effects on other individuals as well. Secondly, this collectivist perspective suggests that large collections of anonymised data on many individuals should not be freely processable by data processors due to the sensitive data fields which may be correlated with less sensitive ones thanks to such data sets. This is in contrast to the current legal situation under the GDPR, which does not restrict processing and storage of anonymised data. Thirdly, and finally, in the collectivist perspective I will call for the rights of data subjects as outlined by the GDPR (right of access, rectification, deletion, ...) to be reformulated in a collectivist manner, so that affected collectives and the community as a whole would be empowered to, in the interest of the common good, exercise such rights against data-processing organisations.

2. Predictive Analytics

For the purpose of this article, it is irrelevant on which algorithms and procedures concretely a predictive model is based. 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, "predictive model" refers to a con-

crete manifestation of this technology. However, for an adequate understanding of the data protection problem, it is helpful to give a functional characterisation 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 (likes, posts, friends, group memberships, …). The target variable is typically hard-to-access or particularly sensitive information on the individual, or a decision about the individual relating to the business of the predictive model's operator (for example, at what price the individual is offered insurance or credit).

Hence, the goal in predictive analyticsis to estimate information about individuals which is difficult to access using easily accessible data. 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 analysed, whether during a learning phase or by means of other, statistical methods. Often, such models are trained with supervised learning methods. This requires a large amount of training data, i.e. 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 dataset for predictive models to estimate the sexual orientation of *any* Facebook user by pattern matching Facebook usage data such as Facebook likes (see Figure 1).

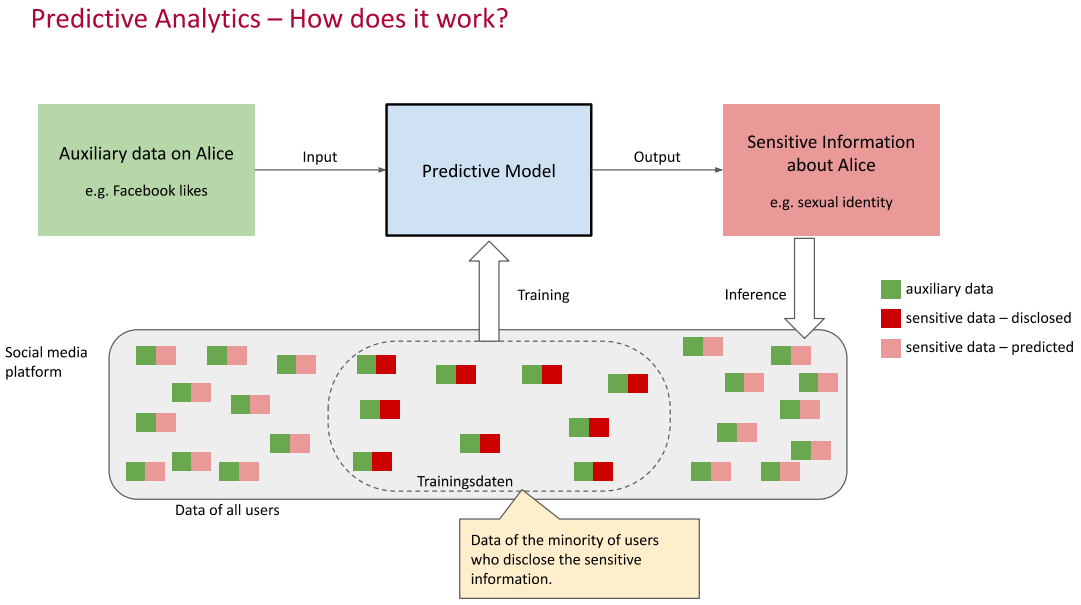

Figure 1: Schematic representation of the procedure of predictive analytics.

If only a few percent of the more than 2 billion Facebook users provide information about their sexual orientation, the resulting training data set still comprises a few 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 would not consent to the processing of this information, have deliberately not provided it, or may be unaware that the company can estimate it about them [cf. also @Skeba-Baumer2020].

Medical researchers at the University of Pennsylvania have shown that this approach can be used to predict whether a user suffers from diseases such as depression, psychosis, diabetes or high blood pressure [@Merchant-et-al2019]. Facebook itself has announced that it can recognise suicidal users by their postings [@Goggin2019]. 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-al2013: 5802].

Such predictive analyses are attracting great interest from insurance and finance companies because they allow individual risk assessment beyond the classic credit scores.[1] Such predictive models are also used in human resource management, for example to carry out automated pre-selection of applicants in hiring processes [@ONeil2016: 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 [@Duhigg2012].

3. Predictive Privacy

Predictive analytics allows unknown or potentially sensitive information about individuals or groups to be estimated using supposedly less sensitive and readily available data (auxiliary data). This is possible with modern machine learning techniques given that many other members of society have provided a 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 (for example, the few Facebook users who provide information about their sexual orientation[2]) sets the standard of information that can be inferred about all members of society. The economic and legal practice of predictive analytics does not provide for the individuals affected by predictions to be informed or asked for consent. Furthermore, there is currently no legal regulation in the EU that prevents or responsibly restricts the production or use of predictive models in general.

1 See [@Lippert2014] on the example of the company ZestFinance as well as [@ONeil2016: Chp. 8] on so-called "e-scores" as alternative credit scoring methods.

2 Ben-Shahar [-@Benshahar2019] offers the helpful conceptualisation of “data pollution” to describe the externalities of ones own data permissiveness.

In addition to potentially beneficial applications, predicted information about individuals or groups can be used in numerous harmful and abusive ways, which can lead to discrimination, unequal treatment and further encroachments on the fundamental rights of those affected. In order to normatively anchor protection against the misuse of estimated information – first ethically, then politically and legally – I therefore seek to construct a new protected good. In direct response to the danger posed by predictive analytics, I propose the concept of *predictive privacy* [cf. @Mü2020:BBAW, @Mü2021:ETIN]. Predictive privacy, in a first approach to the concept, can be defined negatively by pinning down when it is *violated*:

> The predictive privacy of an individual or group is violated when sensitive information is predicted about them without their knowledge or against their will, in such a way that unequal treatment of an individual or group could result.[3]

This very general conceptualisation seeks to adapt and expand the socially and culturally transmitted concept of privacy given the changes in the technological situation brought about by AI. In classic approaches to privacy, breaches of (informational) privacy have mostly been associated with unauthorised access to the private "information sphere" or encroachments on the informational self-determination of the individual,[4] through which information is "stolen" from the data subject that they did not want to disclose about themselves.[5] While a breach of predictive privacy also extracts information which the concerned subject presumably does not want to disclose, this does not happen by way of "theft" or intrusion into a private sphere (it can be doubted whether this metaphor is still adequate in the current technological situation). Rather, in violations of predictive privacy, the information about the data subject is assessed by means of comparison with the data that many *other* data subjects disclosed about themselves. It is important to note that 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. In other words, under the ethical standard of predictive privacy, it

3 Cf. [@Mü2021:ETIN], where it is also explained why the term *predictive* privacy is preferred over “inferential privacy” [@Loi-Christen2020]. Moreover, the source deals in more detail with delineating predictive privacy from related approaches such as “group privacy” [cf. @Floridi2014; @Taylor-EtAl2016; @Mittelstadt2017], “right to reasonable inferences” [@Wachter-Mittelstadt2018] or the older but very pointed proposal of “categorical privacy” [@Vedder1999].

4 Two of the main traditions in privacy appear in the Anglophone world as "non-intrusion theory" and as "control theory" of privacy (cf. [@Tavani2007], who distinguishes a total of four categories). The understanding of privacy as non-intrusion emphasises one (or even several nested) private sphere(s) of each individual, which should be protected from intrusion. Control theories, on the other hand, focus less on seclusion per se, but rather on the individual's ability to effectively and potentially differentiate who has what kind of "access" to one's personal information; see as one of the origins, [@Westin1967].

5 Since the 2000s, Nissenbaum's "privacy as contextual integrity" has provided a refined framework for conceptualising privacy which has been influential in the US-American discourse [@Nissenbaum2011]. Violations of privacy are understood here as violations of context- and culture-specific norms regarding the acceptable flow of information. Why this framework is also unsuitable for addressing the novel privacy challenge of predictive analytics has been discussed in detail in [@Skeba-Baumer2020].

would not be any more legitimate to treat people differently based on predicted information simply because the predictions meet certain requirements of accuracy.[6]

4. A new privacy problem: Three types of attacks

The estimation of potentially sensitive information about individuals based on mass data represents a new dominant attack scenario in data protection. This privacy threat arises under the conditions of insufficiently regulated AI and Big Data technology and has only been evident for about ten years. In order to work out the new 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 recent decades (see Table 1 for an overview).

Dominant Attack Scenarios in Data Protection since 1960

	Type 1: Intrusion	**Type 2: Re-Identification**	**Type 3: Prediction**
Relevant since	1960	1990	2010
Means	Hacking, data leaks, breach of encryption etc.	De-anonymisation throug 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	Fairness; equality of treatment
Protection	Data security	Differential privacy, Federated machine learning	Predictive Privacy

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

a) Intrusion

The archetypal threat in data protection can be described as intrusion. This attack type is closely related to targeted surveillance focusing on specific individuals or groups. Since the proliferation of computerised data processing in the 1960s, the danger 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, protection against this threat is known as data security). Although the main potential attacker is always the data-processing organisation itself, this

6 On this point, the ethical and data protection norm of predictive privacy goes further than Sandra Wachter's and Brent Mittelstadt's [cf. -@Wachter-Mittelstadt2018] demand for a "right to reasonable inferences".

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, government processes, ... that is not meant to be accessible to the attackers.

b) Re-identification

A second type of attack is called re-identification. This type only became significant in the 1990s, after the digitalisation of the healthcare system – for example, billing processes with insurance companies or patient administration in hospitals – made available extensive digital databases on healthcare processes, inspiring the idea to use this data for statistical evaluations in the context of scientific research. This raised the question of how one could anonymise the entries in such databases in order to be able to publish the 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 to research. For this purpose, the data was anonymised by deleting from the records fields such as name, address, and social security number. Latanya Sweeney, then a computer science student at MIT, was able to through a linkage attack identify the record of then Massachusetts Governor William Weld in the anonymised data and reconstruct his medical records [@Sweeney2002; @Ohm2010]. This case triggered an intense discussion in academia and politics about the limits and feasibility of anonymisation. The question of "secure" anonymisation procedures is still being discussed today; current proposals for anonymisation procedures in computer science are always broken a short time later by spectacular attacks;[7] it has thus become clear that "anonymity" is a complex concept which cannot be defined absolutely, insofar as it depends on assumptions about the background knowledge of the attacker as well as the statistical distribution of the data in the data set which is to be anonymised. Moreover, anonymisation methods are required to anticipate all future attack techniques and to cover all possible configurations of background knowledge of future attackers.

The danger of re-identification in anonymised data sets has become a second, much-discussed threat in data protection since the 1990s. This discussion had a noticeable influence on data protection legislation in the context of 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 a breach of anonymity. Even though sensitive data on individuals or cohorts is obtained,

7 Cf. [@Ohm2010] and as examples, see the spectacular re-identification of Netflix users in a pseudonymously published database of film ratings [@Narayanan-Shmatikov2008] or the reconstruction of the family names from anonymously available genome data [@Gymrek-et-al2013].

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.

c) Prediction

My point is that even re-identification can no longer be considered the most important and dominant type of attack in data protection today. The principle of predicting unknown data by means of big data and AI technology does not make the danger of re-identification disappear (just as little as the danger of intrusion). However, the threat of unregulated predictive analytics far surpasses both classic attack scenarios in terms of reach and scalability. 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 the often privileged users who provide the training data for predictive analytics (e.g. the group of Facebook users who provide explicit information about a sensitive attribute, see above) set the standard of knowledge that can be obtained about almost everyone, as long as predictive analytics technology remains unregulated.

This represents a qualitatively new threat in data protection, because the means of violating predictive privacy is neither data theft nor the breach of anonymisation. Predictive analytics of the type significant today hinges on the availability of collective data sets, is possible precisely for those actors who have access to aggregated collective data sets, and has an impact on society as a whole. As a first 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 not only in information being estimated about targeted individuals, but about very large cohorts of users, automatically and synchronously, affecting a broad majority of our societies. At the heart of predictive privacy violations, then, is not espionage directed at individuals, but automated and serialised unequal treatment of people. This unequal treatment is a structural factor insofar as it is directed at all of us in our interaction with automated systems, for example, when we are offered different prices for insurance, when automated decisions are made about who is invited for a job interview, and so on. What is at stake in the violation of predictive privacy is thus the equality and fairness of social treatment. Fairness and equality are, compared to the other types of attacks, a new kind of protected good being violated here: namely, a collectivist good.

5. Predictive privacy as a collectivist protected good

The problem complex of predictive privacy represents a new challenge for data protection, and probably its most significant one at present. In order to recognise the protection of pre-

dictive privacy in the full sense as a problem of data protection, it is necessary to free the mindset of data protection from its fixation on individual claims for protection and to reconstruct a *collectivist protected good* that reaches beyond a sum of individual protective rights. It is true that it is a danger for the single individual to be treated adversely on the basis of predicted information. But this danger alone is nothing new: long before the advent of AI-based predictive analytics, bank advisors made decisions about creditworthiness based on gut feelings, experience and prejudices, doctors prioritised treatment programmes based on personal assessments, and human resource managers predicted the performance of job applicants during the hiring process.

The new quality to the risk arising from predictive analytics lies less in the fact that information about a concrete person X is predicted against their will or without their knowledge, but rather in the fact that the placeholder "X" can represent *any person* at the same time. The technologies used for predictive analytics can make predictions about *any* person X simultaneously and on a large scale, provided that auxiliary data is known about them. The development of predictive analytics technologies usually takes place where there is an interest in algorithmically managing user cohorts and populations [@Mü2020:DZPhil], i.e. in sorting large crowds of people. The essence of predictive privacy breaches is thus not the invasion of a private "sphere", but opening the way 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 with privacy: In the context of AI and Big Data, we are increasingly dealing with issues of equality, fairness and anti-discrimination.

The potential harms resulting from misuse of predictive analytics are thus not fully recognised if one only looks at the consequences for one individual. One has to look at the structural asymmetry of power between individuals and data-processing organisations. A positive definition of the protected good of "predictive privacy" thus goes beyond the negative definition of the "violation of predictive privacy of an individual or a group" that was initially introduced above. Predictive privacy is about regulating a technology that can harm many of us at the same time in our predictive privacy, and thus our society at large in its values of equality, fairness and human dignity. The focus in data protection is hereby shifted from the defensive claims of the individual to a positive assertion of the value of equal treatment. The protection of the community thus takes centre stage. Predictive privacy is about protecting the common good by balancing an asymmetry of power which results from technology’s novel contributions to stabilising and producing social inequalities and unfair discrimination.

The data protection concern of predictive privacy thus relates in a special way to a collectivist dimension of data protection. This dimension can be strengthened by articulating predictive privacy as a collectivist protected good, in a direct response to the potential for abuse of predictive analytics insofar as it affects collectives and not just individuals. The po-

tential for abuse is of a structural quality, it affects everyone synchronously and potentially. It is true that individual damage from *predictive infringements* of (individual) privacy can be palpable from time to time. However, a violation of *predictive privacy* (understood as a collectivist good) means, from the perspective of society as a whole, a cementing or reproduction of social inequality and data-based socio-economic selection through predictive models, which represents damage to the community. In this respect, predictive privacy designates a claim for protection of the community; the protected good of predictive privacy supports the fundamental values of free, egalitarian and democratic societies.

Predictive privacy as a collective duty

In addition to the collectivist nature of the protected good, the violation of predictive privacy is also characterised by a collective "perpetration" or causation. This is because predictive analyses are only possible where two conditions are met: First, a sufficiently large group of users (which sometimes represents the more privileged part of a society) provides, without hesitation, their sensitive data in connection with auxiliary data when using digital services. Secondly, platform companies and other economic actors are technically and legally able to aggregate this data (potentially also in anonymised form) and use it to train predictive models. Secondly, platform companies and other economic actors are technically and legally able to aggregate this data (potentially also in anonymised form) and use it to train predictive models. Given these two conditions, the protection of predictive privacy requires nothing less than a departure from the deeply entrenched liberalist thinking of Western populations regarding data protection. It calls for a common sense of data protection beyond the widespread reduction to the demand that individuals retain control over the use of their personal data.

The starting point for such an upheaval of the common sense in data protection could be the realisation 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 me through predictive analytics, even if I am someone who does not consent to the disclosure of that information.[8] Conversely, this means that one's own data potentially has an impact on other people. These elementary observations about the social externalities of one's data practices,[9] which latter arise from the basic technical structure of predictive analytics, reveal a significant limit to the legal basis of consent at the heart of the liberalist data protection regulation as exemplified by

8 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.

9 In this sense see also the concept of "data pollution", [@Ben-Shahar2019].

the EU GDPR [@EU-GDPR2016]. 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 etc. clearly show. In our current legal and regulatory situation, in which the construction and use of predictive models is not regulated, *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 analyses. One only needs the correspondence of auxiliary data and target information, for example, Facebook Likes and information about diseases; the training data for predictive analytics does not need to contain identifying data fields. Promises of anonymisation are therefore routinely leveraged for promoting users' willingness to consent to the processing of their sensitive data; this is innocuous for Big Data business models based on predictive analytics.[10] In situations where users do not use a digital service anonymously, it is likely that platform companies can still avoid specifying the training of predictive analytics as a data processing purpose, because they can anonymise the data directly after collection and then make further use of it. The reason for this is that anonymised data does not fall within the scope of the GDPR and can be freely used – especially in aggregated form.[11] It can also be stored indefinitely and only later used for predictive analytics. Finally, it should be borne in mind that the trained predictive models themselves represent derived, highly aggregated, anonymised data,[12] which thus do not fall within the scope of the GDPR and, in particular, can be sold and circulated without effective data protection hurdles.

6. Current Deficits in Regulation

Predictive analytics and AI technology have significantly increased the potential for misuse of anonymised mass data over the past 15 years (see also Table 1). However, in the current legal situation, the production and use of predictive models is largely unregulated, so the possibility of misuse is a potentially serious societal force that can stabilise and produce socio-economic inequality and patterns of discrimination.

10 See in particular the research on differential privacy in machine learning, cf. [@Abadi-Chu-EtAl2016; @Dwork2006].

11 The right to erasure in the context of the GDPR can also be fulfilled by anonymising the data, cf. section 6 below.

12 This presupposes that established anonymisation procedures are used, which have been developed for this purpose for fifteen years under keywords such as differential privacy and differentially private machine learning.

Production of predictive models

First, the question arises as to why the EU GDPR [@EU-GDPR2016] does not effectively regulate the production of predictive models. One reason lies in the widespread individualistic thinking in the practice of the GDPR. Although it is controversial whether the GDPR can be dogmatically reduced to data individualism, it is 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 the relation of the individual to their own data. The interpretation is usually: "The sovereignty of individuals in relation to the use of their (personal) data must be preserved; everyone is asked for consent in relation to their own data or another legal basis is declared"; acts of infringement, as well, refer to an individual who claims that *his/her* personal data were processed in a way that was not covered by the claimed legal basis. In particular, the 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 ever be exercised by the individual in relation to their own data.

Another reason, related to individualism, why the GDPR weakly regulates predictive analytics is that it refers to "personal data" (Art. 4 (1)) and does not concern anonymous data [@Wachter2019]. The distinction between personal and anonymous data is outdated in the context of AI and big data. This is not merely because anonymisation can be broken and depends on background knowledge,[13] but because predictive analytics can use the anonymised data of *many* individuals to estimate sensitive and "personal" data about *other* individuals whose data was never recorded and thus never anonymised. The distinction between personal and anonymous data only refers to the "input stage" of data processing [@Wachter-Mittelstadt2018: 125f.] and only considers the relation of the data in question to the concrete data subject from whom the data is collected.[14] The fact that the anonymised data of *many* data subjects enable a new kind of privacy violation against arbitrary *others* remains unrecognised in this scheme. Information derived in the course of data processing can thus undermine the initial distinction between anonymous vs. personal data, not only insofar as supposedly anonymous data could be linked back to the data subject to whom they referred before anonymisation, but rather because new insights into *any* third person can be gained by combining anonymous data of many. The notion of "personal data" in this case would have to refer to variable individuals X and in particular third persons, and is therefore obsolete as a concept.

The legal and theoretical judgement of the danger posed by derived data is controversial and inconsistent. The German Federal Constitutional Court already argued in the 1983

13 This is of course *also* a problem, it would correspond to type 2 of the attack scenarios; however, this is not my focus here, as I argue that there is a new data protection threat (type 3).

14 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.

census ruling that there is no such thing as “irrelevant data” [@Bundesverfassungsgericht1983: 34; @Wachter-Mittelstadt2018: 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 an individual X from seemingly less sensitive or anonymised data about the same individual X (attack type 2). The former *Article 29 Working Party* has recommended in various opinions to include derived information under personal data according to Art. 4 GDPR [@Art29WP251_EN]; however, remaining 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 categorisation of data (as discussed above, for example, 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 [@Art29WP136_EN; @Wachter-Mittelstadt2018: 126]. The European Court of Justice, on the other hand, has clarified in several rulings that the scope of the GDPR is limited to the "input stage" of data processing [@Wachter-Mittelstadt2018: 6] and that the defence against the consequences of data processing, also with regard to automated decisions, must be based on sector-specific regulations [@Wachter-Mittelstadt2018: 7]. With the instrument of the data protection impact assessment, the GDPR provides for a mechanism which can explicitly include the consequences of data processing even beyond the "input stage" and thus in particular also with regard to the effects of anonymised mass data. However, even this comparatively unwieldy instrument is likely to be limited by the distinction between anonymised and personal data. In particular, according to the current interpretation of the right to erasure, this right can also be satisfied by anonymising data records.[15] 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, insofar as anonymised data is sufficient 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. Again, the central reason is that predictive information is not considered "personal data" in the GDPR, as, among others, the standing jurisprudence of the EU Court of Justice has variously confirmed [@Wachter-Mittelstadt2018: 5ff., 105ff.]. This is a point that the *California Consumer Privacy Act* (CCPA [@CCPA2018]), which was adopted in 2018 and came into force in 2020, has ahead of the GDPR: Compared to the GDPR, the CCPA offers a broader definition of "personal information" which also includes, in addition to various directly personal kinds of data:

15 See the decision of the Austrian Data Protection Board [@DSB2018]. See also directly 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).

> "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), quoted after [@Blanke2020: 90])

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 explicitly limited to fully automated processing. Procedures which treat people differently by means of predictive models can comparatively easily be implemented as semi-automated routines by integrating human supervision and intervention possibilities (e.g. by 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 collection 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 behavioural data such as Facebook Likes seem to them to be less sensitive. Moreover, this data is often collected routinely and without specific consent when using social media in everyday life.

7. Proposals for Regulation

Alongside the previous discussion of deficits in regulation, this sections outlines proposals on how to improve the regulation of predictive analytics in the context of the EU GDPR. According to the principle of data protection as a "protection in advance"[16], it is important here to consider the protective effect as a preventative safeguard of equality and fairness in how one is treated by private and public data-processing organisations. The aim is to balance an asymmetry of power between society and organisations; 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 actors with regard to the potential for misuse of anonymised mass data and predictive models.

The protectiveness of a data protection regulation that effectively limits the risks of abuse of predictive analytics can thus not be placed solely on the shoulders of the defensive rights of affected individuals. This is because such an approach always lags behind the actual incidents of infringement. The effectiveness of such instruments is further weakened in the present context by the fact that the violations are often difficult to prove from the individual perspective. Moreover, from the perspective of the affected individual, the demonstrable damage caused by predictive violations of privacy is often marginal, so that individual legal recourse holds little promise of success; however, due to a dispersal effect caused by the automated application of the corresponding techniques to thousands of individuals in parallel, the damage to society as a whole can be considerable [cf. @Ruschemeier2021].

16 German *Vorfeldschutz*, cf. [@Britz2010; @Lewinski2009; @Lewinski2014].

Instead of emphasising individual protection rights, structures must therefore be created at the level of (a) the protected good, (b) the rights of defence and (c) procedural law, respectively, which enable collective action against data companies – both by groups and the community as a whole.

1. Derived Information

The first proposal involves defining derived information as personal information – analogous to the California CCPA. In particular, this would mean that the legitimacy of a data processing operation should not be determined solely at the moment of data collection, but in relation to the purposes and effects of the processing of any data, including, for example, anonymous data and data of other individuals. This means that if, at any point in the course of data processing, information is obtained or processed that relates to any person, this processing would have to fall within the scope of the GDPR. This is because, as described above, the aim is to regulate the extraction of information about *any* individual from the potentially even anonymously processed data of *many other* individuals.

2. Anonymous Data

In order to strengthen this normative intention, anonymised data should also be covered by the GDPR principles.[17] Given the potential for misuse of anonymised mass data, it should not be taken for granted that the processing of anonymised data is allowed without restrictions and takes place in a largely unregulated field of business outside the reach of the GDPR. Moreover, anonymisation of data sets should no longer be equated with deletion.[18] An improved regulation should not focus solely on the danger of re-identification of individuals in anonymised data sets (type 2 attack scenario). Rather, to mitigate the risks of attack type 3, regulation must start from the potential for abuse of *large collections* of anonymised data and of data sets in which various, more or less sensitive data fields can be examined for correlations. A growing social awareness for the richness in information of anonymised mass data would be beneficial for this regulatory concern, so that it is not reduced to the danger of re-identification in the public and political debate. There is also a need to raise awareness of how the information wealth of anonymised mass data is being commercially exploited, with potentially large societal impacts that also affect people whose anonymised data was not used to train predictive models. The GDPR has so far

17 This does not mean, as the proposal is often misunderstood, to categorically prohibit the processing of anonymised data, but, analogous to personal data, to place it under a general prohibition of processing, the exceptions to which must be regulated by legal bases. The legal basis of consent is not applicable here if the consequences of the data processing potentially affect third parties, see below. A political debate must then be held on which uses of anonymised mass data are considered socially beneficial vs. harmful.

18 See above, note 15.

been toothless against this, as Big Data business models capitalise precisely on the uses of data which are possible despite anonymisation and GDPR regulations.

The restriction of the processing of anonymised data must also not be limited to the input stage of data processing. In particular, it must be kept in mind that trained predictive models themselves represent aggregated, anonymised data.[19] Regulation of the processing of anonymised data must therefore cover the circulation and use of trained machine learning models. Predictive models generated from customer datasets can currently circulate or be sold freely and, in particular, without purpose limitation, because they do not fall within the scope of the GDPR. In the context of a new regulation, an expanded form of the purpose limitation and supervision by independent bodies would have to be included. When authorising the production of a predictive model, the purpose for which this model will be used by designated actors would have to be specified and approved in advance, so that repurposing or distribution of the trained model would be prohibited.

3. Restricting Consent

A third pillar of the modernisation of GDPR-style data protection 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 the legal basis of consent is fundamentally questionable. Consent in that sense should only be asked if the consequences of the consent decision solely affect the consenting individual.

Given the average use of the internet and smartphone apps, consent is today one of the most dominant manifestations of data protection regulation. The mind-shaping function of consent should not be underestimated – it confirms the liberalist misunderstanding of data protection which distracts from the dangers of predictive analytics [@Solan-Warner2014; @KrögerLutzUllrich2021]. Each new consent dialogue the user is confronted with affirms the socially damaging understanding that data protection is about one’s individual choice regarding the disclosure of personal data. Furthermore, it is well known and has been much discussed that consent dialogues do not inform users properly, but often trick or coerce them into giving their consent by means of design tricks, nudges, lengthy small print and because they are shown at the most inappropriate moments [cf. @Baruh-Popescu2017; @Mü2018:Leviathan].

The instrument of consent could remain important in the application of predictive models to individuals. An individual affected by predictive knowledge production should have to consent to the acquisition of information or decisions about them before the auxiliary data

19 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 if 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; footnote 10 above.

used for this purpose, which usually seem less sensitive, are collected. In this context, it is worth referring back to the first requirement that inferred data should be treated as personal data and covered by the GDPR principles. In which areas of application consent should be made available as a legal basis for the application of predictive models would have to be assessed in more detail [cf. @Mü2021:ETIN].

4. Collective Rights of the Data Subjects

The establishment of collectivist counterparts to the data subject rights of 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, but also the community as a whole, are enabled to demand information from platform operators about predictive models and the processing of anonymised data.[20] Such a regulation should afford interest groups and democratic society as a whole more control over what information commercial organisations can derive about any individuals from auxiliary data and what predictive models an organisation trains on the basis of the data of many users. This collective right of access should serve to reveal 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 stabilising and reinforcing effects in relation to social inequality, can be observed. For the exercise of these collective rights of defence, supervisory bodies as well as appropriate instruments of collective redress such as class actions should be provided for [cf. in detail @Ruschemeier2021].

Anti-discrimination

Given the potential for abuse of Big Data and AI, effective data protection in the current decade will have to be measured by the degree to which it enters into a sustainable alliance with anti-discrimination legislation. In the context of these technologies, the main issue of data protection is not to prevent spying on individuals, but mass assessment operations which affect us all and can lead to individualised – and that means different – treatment of individuals and groups, and thus to social inequality, discrimination and exclusion. The field of predictive knowledge extraction based on anonymised mass data, which we all produce every day free of charge for big data companies, is currently largely unregulated. In order to recognise the need for regulation, data protection (and especially the Anglophone discourse on privacy) must move away from its favourite point of reference, the protection of the informational sphere and self-control of the individual, and focus on the social structuring effects of modern data processing.

20 See also similar proposals by [@Mantelero2016; @Pohle2016PersonalDataNotFound].

References

[Abadi-Chu-EtAl2016] Abadi, M., Chu, A., Goodfellow, I., McMahan, H.B., Mironov, I., Talwar, K., Zhang, L., 2016. Deep Learning with Differential Privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security - CCS'16 308–318. https://doi.org/10.1145/2976749.2978318

[Art29WP251_EN] Article 29 Data Protection Working Party, 2018. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (No. 17/EN WP251rev.01).

[Art29WP136_EN] Article 29 Data Protection Working Party, 2007. Opinion 4/2007 on the concept of personal data (No. 01248/07/EN WP 136).

[Baruh-Popescu2017] Baruh, L., Popescu, M., 2017. Big data analytics and the limits of privacy self-management. New Media & Society 19, 579–596. https://doi.org/10.1177/1461444815614001

[Ben-Shahar2019] Ben-Shahar, O., 2019. Data Pollution. Journal of Legal Analysis 11, 104–159. https://doi.org/10.1093/jla/laz005

[Blanke2020a] Blanke, J.M., 2020. Protection for 'Inferences Drawn': A Comparison Between the General Data Protection Regulation and the California Consumer Privacy Act. Global Privacy Law Review 1.

[Britz2010] Britz, G., 2010. Informationelle Selbstbestimmung zwischen rechtswissenschaftlicher Grundsatzkritik und Beharren des Bundesverfassungsgerichts, in: Offene Rechtswissenschaft: ausgewählte Schriften von Wolfgang Hoffmann-Riem mit begleitenden Analysen. Mohr Siebeck, Tübingen, pp. 561–596.

[Bundesverfassungsgericht1983] Bundesverfassungsgericht, 1983. BVerfG, Urteil des Ersten Senats vom 15. Dezember 1983 – Zur Verfassungsmäßigkeit des Volkszählungsgesetzes 1983, 1 BvR. Bundesverfassungsgericht.

[CCPA2018] CA, 2018. California Consumer Privacy Act, Cal. Legis. Serv. Ch. 55 (A.B. 375) (west).

[Duhigg2012] Duhigg, C., 2012. How Companies Learn Your Secrets. The New York Times.

[Dwork2006] Dwork, C., 2006. Differential Privacy, in: Bugliesi, M., Preneel, B., Sassone, V., Wegener, I. (Eds.), Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10–14, 2006, Proceedings, Part II, Lecture Notes in Computer Science. Springer, Berlin and Heidelberg, pp. 1–12.

[EU-GDPR2016] EU, 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), OJ 2016 L 119/1.

[Floridi2014] Floridi, L., 2014. Open Data, Data Protection, and Group Privacy. Philos. Technol. 27, 1–3. https://doi.org/10.1007/s13347-014-0157-8

[Goggin2019] 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.

[Gymrek-et-al2013] Gymrek, M., McGuire, A.L., Golan, D., Halperin, E., Erlich, Y., 2013. Identifying Personal Genomes by Surname Inference. Science 339, 321–324. https://doi.org/10.1126/science.1229566

[Kosinski-et-al2013] Kosinski, M., Stillwell, D., Graepel, T., 2013. Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences 110, 5802–5805. https://doi.org/10.1073/pnas.1218772110

[KrögerLutzUllrich2021] Kröger, J.L., Lutz, O.H.-M., Ullrich, S., 2021. The myth of individual control: Mapping the limitations of privacy self-management, in: SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3881776

[Lewinski2014] Lewinski, K. von, 2014. Die Matrix des Datenschutzes Besichtigung und Ordnung eines Begriffsfeldes. Mohr Siebeck, Tübingen.

[Lewinski2009] Lewinski, K. von, 2009. Geschichte des Datenschutzrechts von 1600 bis 1977, in: Freiheit – Sicherheit – Öffentlichkeit: 48. Assistententagung Öffentliches Recht, Heidelberg

2008. Nomos, pp. 196–220. https://doi.org/10.5771/9783845215532-196

[Lippert2014] Lippert, J., 2014. ZestFinance issues small, high-rate loans, uses big data to weed out deadbeats. Washington Post.

[Loi-Christen2020] Loi, M., Christen, M., 2020. Two Concepts of Group Privacy. Philos. Technol. 33, 207–224. https://doi.org/10.1007/s13347-019-00351-0

[Mantelero2016] 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, 238–255. https://doi.org/10.1016/j.clsr.2016.01.014

[Merchant-et-al2019] Merchant, R.M., Asch, D.A., Crutchley, P., Ungar, L.H., Guntuku, S.C., Eichstaedt, J.C., Hill, S., Padrez, K., Smith, R.J., Schwartz, H.A., 2019. Evaluating the predictability of medical conditions from social media posts. PLOS ONE 14, e0215476. https://doi.org/10.1371/journal.pone.0215476

[Mittelstadt2017] Mittelstadt, B., 2017. From Individual to Group Privacy in Big Data Analytics. Philos. Technol. 30, 475–494. https://doi.org/10.1007/s13347-017-0253-7

[Mü2021:ETIN] Mühlhoff, R., 2021. Predictive privacy: towards an applied ethics of data analytics. Ethics Inf Technol. https://doi.org/10.1007/s10676-021-09606-x

[Mü2020:Blätter] Mühlhoff, R., 2020a. Die Illusion der Anonymität: Big Data im Gesundheitssystem. Blätter für Deutsche und Internationale Politik 8, 13–16.

[Mü2020:BBAW] Mühlhoff, R., 2020b. Prädiktive Privatheit: Warum wir alle »etwas zu verbergen haben«, in: Markschies, C., Hermann, I. (Eds.), #VerantwortungKI – Künstliche Intelligenz Und Gesellschaftliche Folgen. Berlin-Brandenburgische Akademie der Wissenschaften.

[Mü2020:DZPhil] Mühlhoff, R., 2020c. Automatisierte Ungleichheit: Ethik der Künstlichen Intelligenz in der biopolitischen Wende des Digitalen Kapitalismus. Deutsche Zeitschrift für Philosophie 68, 867–890. https://doi.org/10.1515/dzph-2020-0059

[Mü2018:Leviathan] 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, 551–574. https://doi.org/10.5771/0340-0425-2018-4-551

[Narayanan-Shmatikov2008] Narayanan, A., Shmatikov, V., 2008. Robust De-anonymization of Large Sparse Datasets, in: 2008 IEEE Symposium on Security and Privacy (Sp 2008). Presented at the 2008 IEEE Symposium on Security and Privacy (sp 2008), IEEE, Oakland, CA, USA, pp. 111–125. https://doi.org/10.1109/SP.2008.33

[Nissenbaum2011] Nissenbaum, H., 2011. A contextual approach to privacy online. Daedalus 140, 32–48. https://doi.org/10.1162/DAED_a_00113

[Ohm2010] Ohm, P., 2010. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. UCLA Law Review 77.

[ONeil2016] O'Neil, C., 2016. Weapons of math destruction: how big data increases inequality and threatens democracy, First edition. ed. Crown, New York.

[DSB2018] Österreichische Datenschutzbehörde, 2018. Datenschutzbeschwerde von Dr. Xaver X.

[Pohle2016PersonalDataNotFound] Pohle, J., 2016. PERSONAL DATA NOT FOUND: Personenbezogene Entscheidungen als überfällige Neuausrichtung im Datenschutz. Datenschutz Nachrichten 39, 14–19.

[Ruschemeier2021] Ruschemeier, H., 2021. Kollektiver Rechtsschutz und strategische Prozessführung gegen Digitalkonzerne. MMR 24, 942–946.

[Skeba-Baumer2020] Skeba, P., Baumer, E.P., 2020. Informational Friction as a Lens for Studying Algorithmic Aspects of Privacy. Proceedings of the ACM on Human-Computer Interaction 4, 1–22.

[Sloan-Warner2014] Sloan, R.H., Warner, R., 2014. Beyond Notice and Choice: Privacy, Norms, and Consent. J. High Tech. L. 14, 370–414.

[Sweeney2002] Sweeney, L., 2002. k-Anonymity: A Model for Protecting Privacy. Int. J. Unc. Fuzz. Knowl. Based Syst. 10, 557–570. https://doi.org/10.1142/S0218488502001648

[Tavani2007] Tavani, H.T., 2007. Philosophical Theories of Privacy: Implications for an Adequate Online Privacy Policy. Metaphilosophy 38, 1–22.

https://doi.org/10.1111/j.1467-9973.2006.00474.x

[Taylor-EtAl2016] Taylor, L., Floridi, L., van der Sloot, B., 2016. Group privacy: new challenges of data technologies. Springer Berlin Heidelberg, New York.

[Vedder1999] Vedder, A., 1999. KDD: The challenge to individualism. Ethics and Information Technology 1, 275–281.

[Wachter2019] Wachter, S., 2019. Data protection in the age of big data. Nat Electron 2, 6–7. https://doi.org/10.1038/s41928-018-0193-y

[Wachter-Mittelstadt2018] Wachter, S., Mittelstadt, B., 2018. A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI (preprint). LawArXiv. https://doi.org/10.31228/osf.io/mu2kf

[Westin1967] Westin, A.F., 1967. Privacy and Freedom. Athenum Press, New York.