The Right Not to Be Subjected to AI Profiling Based on Publicly Available Data—Privacy and the Exceptionalism of AI Profiling

Ploug, Thomas

Published in:
Philosophy and Technology

DOI (link to publication from Publisher):
10.1007/s13347-023-00616-9

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Philosophy & Technology (2023) 36:14
https://doi.org/10.1007/s13347-023-00616-9

RESEARCH ARTICLE

The Right Not to Be Subjected to AI Profiling Based on Publicly Available Data—Privacy and the Exceptionalism of AI Profiling

Thomas Ploug[1]

Received: 10 October 2022 / Accepted: 22 February 2023 / Published online: 7 March 2023

Abstract
Social media data hold considerable potential for predicting health-related conditions. Recent studies suggest that machine-learning models may accurately predict depression and other mental health-related conditions based on Instagram photos and Tweets. In this article, it is argued that individuals should have a sui generis right not to be subjected to AI profiling based on publicly available data without their explicit informed consent. The article (1) develops three basic arguments for a right to protection of personal data trading on the notions of social control and stigmatization, (2) argues that a number of features of AI profiling make individuals more exposed to social control and stigmatization than other types of data processing (the exceptionalism of AI profiling), (3) considers a series of other reasons for and against protecting individuals against AI profiling based on publicly available data, and finally (4) argues that the EU General Data Protection Regulation does not ensure that individuals have a right not to be AI profiled based on publicly available data.

Keywords Artificial intelligence · Privacy · Right not to be profiled · Social control · Stigmatization

1 Introduction

The analysis of social media data with artificial intelligence (AI) models holds considerable potential for predicting health-related conditions. In a study of Instagram photos, a machine-learning model was able to identify depressed users with greater precision than unassisted general practitioners (Reece & Danforth, 2017). Among others, the model used photo brightness and colours, number of

✉ Thomas Ploug
ploug@ikp.aau.dk

[1] Centre for Applied Ethics and Philosophy of Science, Department of Communication and Psychology, Aalborg University, A C Meyers Vænge, 2450 Copenhagen, SV, Denmark

comments, and likes as predictors and found that depression was associated with postings of bluer, darker, and greyer photos that received more comments, but fewer likes. In a related study of Twitter data, the developed model also achieved greater accuracy than unassisted physicians in identifying people with depression (Reece et al., 2017). The study showed that the dominant contributors to the difference between depressed and healthy individuals were an increase in the use of negative words and a decrease in the use of positive words among depressed people. The study also found that increases in word count were positively associated with depression. Other studies have shown equally promising results for models using social media data for the prediction of mental disorders, including anxiety, bipolar, borderline personality disorder, schizophrenia, autism (Gkotsis et al., 2017; Kim et al., 2020; Kumar et al., 2019), and the risk of suicide and anorexia (Amini et al., 2022; Coppersmith et al., 2018; Zirikly et al., 2019).

The increased predicting potential and widespread use of AI models across sectors has sparked an intense debate on how to regulate AI. A host of actors have issued a significant number of AI guidelines (Jobin et al., 2019), and the EU Commission has recently issued a proposal for the regulation of AI. The promises and perils of mental health profiling based on social media data add fuel to this debate. Thus, predicting an identifiable individual's health-related conditions (output data) based on social media data (input data) may benefit the individual in various ways, but it also raises significant issues of privacy. This article concerns an individual's privacy rights in situations where the data used to make sensitive predictions has been made publicly available on, for instance, social media platforms. In particular, it is argued that individuals should be granted a sui generis legal right not to be subjected to AI profiling based on publicly available data without their explicit informed consent. As such, the legal right proposed is (1) a negative right that entitles individuals to non-interference with respect to AI profiling producing sensitive personal data (output data) based on online and publicly available personal data (input data), (2) it is a right that can be waived through informed consent, and (3) it is a pro tanto right, i.e. it is not an absolute or unconditional right, but a right that may be infringed under certain exceptional conditions (Frederick, 2014).

The need for a sui generis legal right is substantiated through a four-step analysis. In the course of this analysis, it is argued:

1) that there are strong reasons for protecting personal data as such data may drive different types of social control and may lead to stigmatization
2) that a number of features of AI profiling make individuals more exposed to social control and stigmatization than other types of data processing and that AI profiling thus poses a unique threat to individuals (the exceptionalism of AI profiling)
3) that there are strong reasons for protecting public discourse and interaction on social media and that there are no obvious trumping concerns to the contrary
4) that existing EU legislation—i.e. the General Data Protection Regulation (GDPR)—does not ensure individuals a right not to be AI profiled and that in any case, there are reasons for making it an explicit sui generis right

The article is ended by a few reflections on some of the key questions raised by the proposed arguments, and an agenda for future research is suggested.

All throughout the notion of 'data' will be used to denote both the data being produced by AI profiling, i.e. the output data, and the data on which the AI profiling is based, i.e. the input data. The context will determine the type of data being referred to.

2 Three Cases of AI Profiling Based on Social Media Data

The four-step analysis provided in this article will involve three stylised cases of AI-driven mental health profiling. These cases are included in order to illustrate the various different contexts in which AI-driven mental health profiling may be conducted. The cases are as follows:

The Friends Case
Two close friends, A and B, are socializing on a regular basis. At their recent gatherings, B has been absentminded and shown some signs of a general lack of motivation. Being concerned about the friend but unwilling to confront B with weakly evidenced suspicions, A decides to do an AI-generated mental health profile of B. Therefore, A collects a series of photos that B has made publicly available on Instagram and feeds them a locally stored copy of the highly accurate mental health prediction model *Deepmood*. Much to the surprise of A, the model predicts that B is bipolar. A reports the findings to B and B's family.

The Public Servant Case
A public servant, A, in charge of the unemployment benefit scheme at the local municipalities comes to suspect that a client, B, may be suffering from a chronic mental illness likely to keep B in long-term unemployment. Wanting to ensure that B is adequately handled by 'the system' based on objective information, but unwilling to cause unnecessary unrest, A decides to do an AI-generated mental health profiling of B based on a series of public tweets from B. A locally stored copy of the highly accurate mental health prediction model *Deepmood* suggests that B suffers from anxiety. A reports the findings to B.

The Prime Minister Case
A concerned citizen, A, fears that the upcoming general election may pave the way for a prime minister candidate, B, whom A suspects may be mentally unstable and thus unfit for the job. Believing that it is in the public interest to have reliable information about the mental health of B, A decides to do an AI-generated mental health profile of B. A collects a sample of public tweets from B and feeds them to a locally stored copy of the highly accurate mental health prediction model *Deepmood*. The model suggests that the candidate suffers from severe depression. A reports the findings to the general public.

3 Reasons for Protecting Personal Data

The philosophical literature on the value and definition of privacy is vast (Leino-Kilpi et al., 2001). This article develops and considers three arguments in favour of a right to privacy specifically in relation to the use of personal data. As will become evident, the three arguments are particularly relevant for an analysis of the previously introduced cases involving AI profiling.

3.1 The Social Pressure Argument

The argument from social pressure is an autonomy-based argument for the right to privacy regarding personal data. Essentially, the argument contends that access to personal data about individuals may be used to exert communicative pressure on the choices and actions of these individuals in ways and to an extent incompatible with their interests and preferences. Communicative social pressure can be defined the following way: A communicative act—verbal or nonverbal—performed by a sender, say A, amounts to social pressure on a recipient, say B, if and only if it succeeds in making B believe that certain choices and actions are associated with costs to an extent, making it less likely that B will make the relevant choice or perform the relevant action. In the friends case, A may use knowledge of B's diagnosis to pressurize B into treatment by communicating in one way or another that if B remains untreated, it will have negative consequences not only for B, but also for family and friends. In the case of the prime minister, knowledge of B's diagnosis can similarly be used to pressure B to abandon the candidacy. The influence of unwanted social pressure based on personal data may be limited through a right to privacy. Insofar as individuals have a right to exercise personal autonomy, where this is taken to include a right to protect themselves against unwanted social pressure, they should have a right to privacy limiting the access of others to data about them. Thus defined, a right to privacy empowers individuals to influence the ways and extent to which they are subjected to social pressure. It does so by letting individuals shape the social pressure by deciding the level, character, and with whom they share data about themselves.

Here, three observations must be made. First, the argument does not claim that access to accurate data is a necessary condition of social pressure, but only that it may be sufficient for others to be able to exercise such pressure. An individual may put pressure on another individual in various ways without having any knowledge of the latter. Second, it does not make any distinctions between different types of data about individuals, e.g. personal and non-personal data. One may claim, however, that personal data for various reasons is more effective in the attempt to put pressure on another individual and that individuals would therefore withhold such data to a larger extent than nonpersonal data. Third, the argument does not operate on a distinction between different methods of exerting social pressure on choices and actions. Social pressure methods are many and versatile. They cover a spectrum of verbal and non-verbal, explicit, and covert communicative acts that include persuasion, coercion, and manipulation. Persuasion can be defined as the act of pointing

out verbally the good and bad consequences of a certain course of action (Powers, 2007), and coercion as the verbal act of threatening to make someone worse off than he or she would otherwise be or ought to be (Nozick, 1969; Wertheimer, 1990). Manipulation covers a range of covert attempts at influencing other people through the tailoring of information, e.g. by lying or withholding information, by making exaggerations, and by framing information that is likely to lead others to believe what is false (Beauchamp & Childress, 2001). While the ethical gravity of performing acts of persuasion, coercion, and manipulation differs significantly, the argument from social pressure is indifferent to such distinctions. It implies a right to be left alone in the broader sense of being able to limit all kinds of data-facilitated *social pressure*.

3.2 The 'open future' Argument

The 'open future' argument is also an autonomy-based argument for the right to privacy. At the heart of the open future argument lies the observation that personal data shared by individuals at a certain point in their life may come to shape the future opportunities afforded to them by others in ways that run counter to their interests. In short, the 'open future' argument contends that access to the personal data of individuals can lead to unwanted interventions in future choices. A choice-set intervention can be defined as follows: An agent, say A, makes an intervention in the set of choices of another agent, say B, if and only if A, on the basis of having access to certain data about B, shapes the choices of B differently than they would otherwise have been. In the public servant case, the public servant's knowledge of the diagnosis of client B can change the opportunities offered to B within social services. Similarly, the general public's knowledge of the diagnosis of candidate B in the prime minister case may have consequences for the career opportunities offered by existing political parties. In both cases, the effects on the choice set of B can go against the interests and preferences of B. If people have a right to exercise personal autonomy, where this includes a right to protect themselves against unwanted influences on the options available to them in the future, they must also have a right to privacy limiting the access of others to the personal data that may be used for shaping their future choice set.

Several observations must be made here. First, the notion of *shaping future options* should be understood both quantitatively and qualitatively. An open future is both a matter of the number of available options, but also a matter of the availability of certain vital options (Garrett et al., 2019). While the prime minister candidate may not necessarily be robbed of a political career in the highest office, it may be a career under constant accusations of being unduly influenced by depression. Second, the argument from an 'open future' differs from the argument from social pressure. Although the exercise of social pressure may be effective in associating a certain choice with costs to an extent making this choice unlikely to be made, it may not limit or otherwise significantly alter the choice set prior to the choice situation obtaining. Moreover, social pressure requires a communicative act of some kind. In the following, the exercise of social pressure and choice-set interventions shall both

be referred to as forms of *social control*. Third, the 'open future' argument presented here differs from the standard 'open choice' argument in relation to children. Feinberg famously argued that a child's right to autonomy in adult life can be violated in advance through parental choices that limit a child's opportunities in later life (Feinberg, 1980). The 'open future' argument advocated does not posit a future autonomy right that can be violated in advance. It is an individual's right to personal autonomy in the present that grounds a right to act in the interest of protecting the future autonomy of the individual.

3.3 The Stigmatization Argument

The stigmatization argument grounds the right to privacy in the potential harms of suffering stigmatization. According to Goffman, stigmatization occurs when a person is attributed with a discreditable trait that makes the person inferior, dangerous, and perhaps even inhuman, and when this act of discrediting leads to discrimination (Goffman, 1963; Ploug, 2020). In a later reinterpretation of Goffman, stigmatization is taken to be a complex phenomenon at the societal level constituted by a process in which a difference between people is identified, labelled, and associated with negative stereotypes, followed by acts of segregation ('us' and 'them') and discrimination made possible by a greater social, economic, and political power of the labelling party (Link & Phelan, 2001; Ploug et al., 2015). So defined, stigmatization clearly presupposes access to data about people. The identification and labelling of a difference between people—intentionally or unintentionally—is an act that requires access to data. In the three cases, access to AI-generated diagnoses would allow the friend, client, and candidate to be identified as different and labelled, e.g. 'mentally ill', and thus be included in a stigmatized group in society (more on this below). A right to privacy that limits access to certain data could protect people from stigmatization.

One may perhaps be tempted to object here that it seems as if the real moral concern with stigmatization is discrimination, i.e. morally unjustified differential treatment. However, research shows that stigmatization is associated with other harms to stigmatized individuals. The case of mental illness is highly illustrative. Studies show not only that mental illnesses are associated with negative stereotypes (Bhugra, 1989) but also indicate discrimination against mentally ill people in the labour market and in health care (Druss et al., 2011; Stuart, 2006). Therefore, the stigmatization elements introduced above—that is, labelling, negative stereotyping, segregation, discrimination, and power asymmetry—are present in relation to this group (Link & Phelan, 2001). But there are other harms to this group than discrimination. Evidence on labelling of mentally ill people in healthcare suggests that it can lead to the self-application of labels and negative stereotypes, where this is associated with lower self-esteem, demoralization, income loss, unemployment, and self-discrimination (Corrigan et al., 2009; Link, 1987). Studies of barriers to mental health treatment find that negative attitudes toward mental illnesses keep patients from seeking health care (Mojtabai et al., 2011; Voorhees et al., 2005). In turn, this can drive inequality in health care (Stuber et al., 2008). Studies on the effects of stigmatization of

people with AIDS show that the feeling of stigmatization due to one's health status is associated with anxiety, depression, distrust, and the disruption of normal social relationships (Crandall & Coleman, 1992; Herek, 1999). Similar findings of stigmatization and its harmful effects have been reported for smokers and obese people (Goldstein, 1991; Hilbert et al., 2008; Myers & Rosen, 1999; Peretti-Watel et al., 2014; Stuber et al., 2009).

A comment must be made. In this section, it has been argued that the threat of stigmatization and the associated harms could and should ground a right to privacy, but perhaps less than stigmatization is needed to make this argument. Research suggests that negative stereotyping promulgated, for instance, in the media, is associated with a number of harms to members of the stereotyped groups. It may elicit situational responses of negative emotions and stress, induce self-discrimination in relation to a variety of activities, and impair cognitive and educational achievements (Appel & Weber, 2021). A right to privacy may protect against such weaker forms of singling out individuals on the basis of personal data.

4 Reasons for Protecting an Individual Against AI Profiling

The three arguments developed above have foremost provided reasons for a *general* right to data privacy. Some of the considerations along the way have indicated that personal data may make an individual particularly vulnerable to attempts at social control and the harm of stigmatization. However, the arguments do not distinguish AI profiling from any other processing of personal data, and hence, they cannot in and of themselves sustain a sui generis right not to be subjected to automated profiling based on publicly available data. In this section, it is therefore argued that a number of features of AI profiling make individuals more exposed to attempts of social control and stigmatization and that—in doing so—AI profiling presents a particularly invasive kind of data processing.

4.1 The Exceptionalism of AI Profiling 1: Predictive Accuracy and Social Control

As evidenced by the studies cited in the introduction, AI profiling differs from other types of data processing by potentially making highly accurate predictions about the future behaviour and dispositions of individuals. Mental health profiling as in the three cases is not only a matter of categorizing health problems in the past and present, but it also involves predicting mental health dispositions that will shape an individual's future. In the debate on the exceptionalism of genetic data, it has been claimed that such data are exceptional in that they can be used as a predictor of future disease (McGuire et al., 2008). The counterargument has been that the predictive power of genetic data is limited, because genetic risk factors are only—with the exception of a few rare, highly penetrant genotypes—among many contributors to disease (Evans & Burke, 2008). Notably, AI profiling does not have such inherent limitations. AI profiling, e.g. in relation to diagnostics, can incorporate all social, genetic, and other data types relevant for making predictions about future disease. In

doing so, AI profiling may make highly accurate predictions about the future behaviour and dispositions of individuals.

Predictive profiling of behaviour and dispositions exposes individuals more to attempts of social control than they would otherwise have been. Here is why. Attempts at socially controlling the future behaviour of individuals may—at least in some cases—rest on data-driven beliefs about their future behaviour and dispositions. Individuals for whom there are no (accurate) personal data indicating that they are disposed for smoking seem unlikely to be subjected to attempts of social control (with regard to smoking behaviour) to the same degree as individuals for whom such data exist. What matters here is first and foremost the existence and accessibility of personal data indicating that an individual may come to smoke in the future. However, secondly, it seems equally important that the data are accurate and therefore reliable. Although access to more or less reliable historic data evidencing that an individual once tried smoking may trigger attempts at socially controlling his or her future smoking behaviour, it seems that a reliable prediction of future smoking is more likely to do so. Third, it also matters whether the prediction concerns future behaviour or more deeply rooted dispositions, where a disposition may for present purposes be defined as a trait leading to repeated occurrences of a particular type of behaviour in particular circumstances. A prediction of a smoking disposition may trigger a more intensive effort at social control than a prediction of an isolated occurrence of smoking behaviour.

AI profiling may bring to the fore highly accurate predictions of future behaviour and dispositions that may lead to beliefs triggering various attempts of social control that would not otherwise have been attempted. Such a scenario is uniquely dependent on predictive profiling. To the extent that this scenario is considered possible—perhaps even likely—the predictive character of AI profiling makes an individual more exposed to attempts of social control.

4.2 The Exceptionalism of AI Profiling 2: Predictive Accuracy, Determinism, and Stigmatization

The potential of AI profiling to make highly accurate predictions may also increase the risk of stigmatization. Although predictability and determinism are considered logically independent notions (Bishop, 2003; Rummens & Cuypers, 2010), they seem to play a similar role in our daily and practical lives. If someone's behaviour and dispositions are fully predictable, then the common-sense interpretation of this is that the behaviour and dispositions could not have been otherwise—they are necessary and inescapable. The same holds for determinism. Fully determined behaviour or dispositions entail the inevitability of this behaviour and these dispositions. Why does it matter?

Studies on public beliefs in genetic and social determinism, i.e. that genetic and social factors determine human behaviour and personality, show that such beliefs are associated with social cognitive effects such as negative stereotyping, the formation of prejudice, and tendencies to discriminate by confining the civil rights of

others (Keller, 2005; Rangel & Keller, 2011). Furthermore, these studies also show that such beliefs are related to correlates of essentialist thinking, where the latter may be defined as the tendency of individuals to attribute to others an unalterable essence that makes them what they are and that explains their actions (Medin, 1989; Yzerbyt et al., 1997). In combination, these findings are taken to suggest a mechanism whereby individuals attribute to others an essential nature—a set of dispositions—to the extent they believe their behaviour and dispositions to be genetically and socially determined, and this attribution of essence results in negative stereotyping, the formation of prejudice, and tendencies to discriminate.

The argument to be made here is simply this. If a belief in the predictability of behaviour and dispositions plays the same role as or may cause a belief in the determinism of behaviour and dispositions, then it may lead to essentialist thinking producing negative stereotyping and discrimination, which are defining components of stigmatization. AI profiling may produce highly accurate predictions and thus substantiate beliefs in the predictability of behaviour and dispositions. Hence, AI profiling may lead to stigmatization.

The argument is somewhat speculative, and more conceptual and empirical work is certainly needed. However, note that it has some intuitive appeal. If we were to believe that the client in the public servant case or the prime minister candidate in the prime minister case was *essentially* and *unalterably* mentally ill, then it does not seem wholly implausible that this would lead to negative stereotyping and discrimination. If we believe others to be machines, then we will treat them as machines. If the argument holds, it goes to show specifically that the predictive accuracy of AI profiling makes individuals more exposed to stigmatization.

4.3 The Exceptionalism of AI Profiling 3: A Purely Feature-Driven Approach

The arguments in the previous two sections have a similar structure. It has been argued that a special feature of AI models, i.e. predictive accuracy, in combination with certain assumptions about human belief formation makes profiled individuals more exposed to social control and stigmatization. However, there are other features of AI profiling that in and of themselves make profiled individuals more exposed to social control and stigmatization.

AI modelling is versatile. Different predictive AI models can be trained on some data and subsequently applied to the same kind of data. In other words, it is possible to develop AI models that make different predictions based on the same input data. In the three cases, AI profiling based on publicly available photos and tweets serves the purpose of predicting mental health dispositions. The same publicly available data could potentially be used to predict other personal dispositions, including creditworthiness, risk of criminal behaviour and drug abuse, political leanings, and parental competencies. Any data made publicly available at any time could turn out to be a relevant input for an AI model that makes predictive profiling. The versatility of AI modelling implies that a vast amount of personal data may potentially be generated from limited publicly available data. The potential increase in the availability

of personal data that may be used for the purpose of social control and stigmatization makes an individual more exposed to such interventions.

Predictive AI profiling is human-unpredictable. AI profiling may reveal personal data that would otherwise be hidden from a human being. In the three cases, the mental dispositions are not available to any human seeing the photos or reading the tweets, including the individual making them publicly available. There are no human interpretable signs that suggest that photos and tweets contain reliable data on mental dispositions. But the mental dispositions are also human unpredictable in a deeper sense. Thus, any other nonautomated method for making the same predictions based on the available data, e.g. statistical modelling, would be bounded by the availability of human expertise and other humanly limitations and may not deliver the same accuracy. The human unpredictability of AI profiling makes an individual more exposed to social control and stigmatization for two reasons. Firstly, because the unpredictability impedes an individual's ability to exercise self-protection by withholding data. If it is impossible to foresee what predictions can be made from a set of public data, individuals cannot withhold data that they believe may generate unwelcome attempts of social control and that they believe would place them in groups already stigmatized or likely to become stigmatized. Avoiding entirely to make any data public to anyone seems practically impossible in a modern-day society with the growing digitization of human relations. Second, because AI makes predictable what is humanly unpredictable, it removes the protection against social control and stigmatization that is tied to human unpredictability. That no human being may predict mental dispositions just by looking at photos and tweets and that other nonautomated methods are bounded by the availability of human expertise constitutes a protection of individuals against social control and stigmatization. It means that the predictions will not be generated and therefore cannot feed attempts of social control and stigmatization.

AI technology and embedded models are highly transferable. Although the development of a high-performing AI model requires access to very large datasets, expert knowledge, hardware, and various other resources, the resulting model may be widely used by a host of different stakeholders. As imagined in the three cases, highly accurate predictive AI models may well be made available for public use. AI models differ in this respect from other technologies. Whole genome/exome sequencing is a comparable technology in that it can be used to map the complete DNA sequence of an individual from a sample of human tissue. This could, for instance, be human tissue left in public as fingerprints on a glass. The resulting genetic data can to some extent be used to make predictions about human dispositions and, in particular, disease risks. Contrary to AI technology, however, the whole genome/exome sequencing technology along with required lab facilities, biomedical expertise, etc. cannot be made publicly available. The inaccessibility of this technology appears to reduce the risk of being genetically profiled inadvertently, while the transferability of AI technology arguably increases the risk of being profiled by AI and thus increases the exposure of an individual to social control and stigmatization.

Finally, AI predictive profiling goes beyond an individual. AI predictions of dispositions will in certain cases reveal personal data not only about individuals, but also about their relatives. An AI model that predicts that an individual is likely to

develop diabetes 2—known to be partly heritable and partly due to lifestyle—will also, as a matter of fact, have revealed an increased risk of diabetes 2 among close relatives. Such inferences will be warranted to a very different degree. However, insofar as AI predictive profiling may reveal personal data about relatives, it makes those relatives more exposed to social control and stigmatization than they would otherwise have been.

5 Reasons for and Against Protecting Publicly Available Online Data

The previous section provided ground for granting individuals a right not to be subjected to AI profiling. This article seeks, however, to sustain the right in relation to personal data made publicly available online, e.g. on social media platforms such as Twitter and Instagram. Notably, the availability of personal data on such platforms is to a large extent influenced by the data subjects themselves. If an individual refrains from any use of online services and social media platforms, little personal data would ceteris paribus be publicly available. In so doing individuals would therefore—presumably to a large extent—protect themselves against AI profiling based on online available, personal data. But if individuals may protect themselves against AI profiling by disengaging from online data sharing, then the right not to be AI profiled based on publicly available data seems to have been rendered somewhat superfluous. This section therefore explores, firstly, the reasons for protecting and promoting online data sharing and, secondly, the reasons for allowing AI profiling based on such data.

5.1 Social and Democratic Potentials of Online Data Sharing

The online sharing of data may serve important social and democratic purposes.

Socially, it can be used to build and maintain relationships. Research finds that socializing is a primary motivation for information sharing on social networking sites (SNS) (Kümpel et al., 2015). Studies also suggest that the use of SNS is positively related not only to the quality of friendships defined as satisfaction with friends, but also to social capital defined as the resources that accrue to an individual by virtue of his or her membership of a social network, e.g. the feeling of belonging to a community, the forming of new relationships, the feedback and support of other people in relation to various issues, and loneliness (Ahn, 2012; Antheunis et al., 2016; Brandtzæg, 2012; Ellison et al., 2007). Other studies indicate that social media use with close friends is positively associated with the experience of feeling close to those friends (Kahlow et al., 2020; Pouwels et al., 2021).

Democratically, it may be used to form and express opinions, to share information, and to engage in collective reasoning involving decision-makers at all levels of society, and thus, it may ultimately increase participation in democratic processes. A meta-analysis found that social media use generally is positively—but modestly—related to various forms of both online and offline political participation, e.g. voting, supporting campaigns, and protest activities (Skoric et al.,

2016). Consistent with the findings of a previous meta-analysis (Boulianne, 2009), it also found a positive—and moderate—relation between (1) the use of social media for informational purposes, i.e. seeking, gathering, and sharing of various kinds of information, and political participation, and (2) the use of social media for expressivist purposes, i.e. expressing oneself, articulating opinions, ideas, and thoughts, and political participation. Another meta-analysis similarly established a positive relationship between social media use and participation in political and civic life. However, the meta-analysis found it to be questionable whether such participation is a causal effect of social media use (Boulianne, 2015). Other studies suggest that online viewing and sharing of news are positively related to political knowledge (Beam, 2016).

Minimally, the evidence cited suggests that *certain kinds of* online data sharing can, *under certain conditions*, serve valuable social and democratic purposes. If so, then we have reason to protect—perhaps even promote—these kinds of online data sharing. However, this requires that privacy concerns be accommodated because such concerns may lead to withdrawal from online data sharing. A recent study found that privacy concerns defined as concerns about losing control of personal data on social networks are negatively related to social media participation measured as the frequency of interactions on SNS (Bright et al., 2021). A similar study found that privacy concerns in relation to publicly available data are negatively associated with the amount and depth of personal data sharing on SNS (Gruzd & Hernández-García, 2018).

What has been attempted here is to underpin two simple claims, namely that online engagement in the form of data sharing *can* serve valuable social and democratic purposes and that privacy concerns *can* lead to disengagement from online data sharing. If both claims are true, they provide us with a reason to address the concerns of individuals regarding privacy. They imply that disengagement from online data sharing is an undesirable solution to such privacy concerns. The upshot for AI profiling based on publicly available data is this. If individuals sometime in the future, when AI predictive profiling has become more widespread, become aware that their publicly available data can be used for profiling in unpredictable ways, there is a risk that this may lead to disengagement from online data sharing, also by prospective politicians and decision-makers. This is undesirable as it may impede the full realization of the social and democratic benefits of data sharing.

While we have focused here on social and democratic aspects, there is also an extant literature on the health-related effects of social media usage. A recent umbrella review comparing five meta-analyses found that social media use is associated with higher levels of both well-being, i.e. happiness, positive affect, life satisfaction, and ill-being, i.e. i.e. depression, anxiety disorder, distress, and negative affect (Valkenburg et al., 2022). While such studies certainly add important nuances to the evidence on the effects of social media use, they do not rule the possibility that social media use may serve valuable social and democratic purposes. If privacy concerns—and in particular concerns related to AI profiling—can lead to disengagement from the relevant kinds of data sharing, there are substantial reasons for protecting such data sharing from AI profiling.

5.2 Other Trumping Concerns

To fully substantiate the right not to be subjected to AI profiling based on publicly available, personal data, it remains to be considered whether other concerns may outweigh this right. These concerns may be related to specific purposes and contexts of AI profiling. Revisiting the three cases may therefore be helpful.

AI predictive profiling makes an individual significantly more exposed to unacceptable forms of social control and stigmatization and self-stigmatization, and it may lead to withdrawal from online data sharing otherwise conducive to social thriving and democratic health. As such, AI predictive profiling is a threat to basic human autonomy and wellbeing, as well as social life and democracy. These harms impose a burden of proof on the defender of the right to conduct such profiling in the three cases. For each case, the defender of AI profiling based on publicly available data must show either (1) that the harms of AI profiling are unlikely to obtain in the three cases and/or (2) that AI profiling without consent is a proportional measure, where this requires (A) that the benefits of doing AI profiling without consent outweigh the potential harms and (B) that AI profiling without consent is strictly necessary in order to obtain the benefits.

Consider the friends case. That household profiling should be harmless is not a credible proposition. There are no a priori reasons to think that profiling of friends and family members is not as likely to lead to overreaching social control and stigmatization as in any other case. Note that the effects of stigmatization do not require stigmatizing behaviour by friends and family. It only requires that a certain feature is stigmatized in the wider society. Consider now the benefits and the necessity of conducting the AI profiling. The benefit of the profiling is an accurate diagnosis that A may use to motivate B and B's family to seek further health care assistance. Note that this benefit could likely have been achieved by less invasive alternative means. A could simply have presented B and B's family with the suspicion of B having mental health issues. Note also that whether or not the AI profiling is necessary, it is still an open question, who should decide the weight of the benefits and harms? There are at least two reasons why B should decide the weight of the benefits and harms, i.e. that B should have a right to provide or deny informed consent to AI profiling based on publicly available data. Firstly, because the benefits and harms of the AI profiling are relative to B's interests in getting the profiling data or maintaining privacy, B is ceteris paribus the best positioned to determine B's interests and their weight. Secondly, because the potential harms to B—i.e. social control and stigmatization—for all we know are significant harms. They are ways of impacting others that run counter to fundamental values in our society, and it seems to be an equally fundamental principle that generally—if not always—when individuals are at risk of suffering significant harms, they should have the right to protect themselves against these harms.

In the public servant case, there is also the risk of overreaching social control and stigmatization. Psychiatric diagnoses are sticky labels that will form the way individuals are handled in the public system in the short- and long-term, and this carries a latent risk of producing overreaching social control. It may also lead to stigmatization and not least self-stigmatization. In this case, however, there are potential

benefits both for the profiled individual and the public authorities. Thus, the diagnosis may not only come to benefit client B in terms of more adequate health care and social benefits, but it may also increase the efficiency of the social services by making more readily available personal health care information needed for an adequate distribution of social benefits. AI profiling based on publicly available data may be considered necessary in the sense of being a precondition of a maximally efficient public administration, but it is not necessary in the sense that there are no alternative ways of getting access to the information produced by the AI profiling. Thus, the benefits of such AI profiling for public administration and the wider society may ultimately be marginal. In any case, it seems as if such benefits are morally incomparable to the list of potential harms of AI profiling, including the potential harms to B. As argued above, the potential harms to client B speaks in favour of granting B the right to provide or deny informed consent to such profiling.

The prime minister case may for two interrelated reasons be taken to present a more fundamental problem for the attempt to ground a sui generis right not to be AI profiled. Thus, it may be argued that information about the mental health of the prime minister candidate is of public interest and that the public therefore have a right to this information. Relatedly, it may be argued that the AI profiling is covered by the right to freedom of expression. After all the concerned citizen is in this case profiling the candidate with the intent of voicing a concern over the fitness of the candidate for political office. Access to information of public interest and the right to free speech are undeniably instrumental to a flourishing democracy. However, what we have been arguing in this article is that AI profiling is not only a continuous threat to individual autonomy, wellbeing, and social life, but that it may also threaten democratic processes by potentially leading people to withdraw from the use of social media. Decision-makers, including the prime minister candidate, may withdraw from social media exchanges with the public, and in the longer run, the threat of being AI profiled for all sorts of dispositions may prevent people from engaging in politics altogether. Therefore, it remains to be shown how and why a right not to be AI profiled based on publicly available data limits freedom of expression and information in any profound sense, and why such limitation outweighs the negative effects of AI profiling on human autonomy and wellbeing as well as on social flourishing and democracy. Simply flagging the right to freedom of expression and information cannot reasonably be thought to do the job.

Relevant for the attempt to weigh benefits and harms in all three cases is the accuracy of the predictions of mental disorders AI model *Deepmood*. Thus, it may be asserted that a predictive accuracy, i.e. sensitivity and specificity, above that of health care professionals could and would increase the benefits to individuals and society in all three cases—and vice versa. In short, the accuracy of AI models should be a parameter for any attempt to weigh benefits and harms. The evidence cited in the opening lines of the article suggests that available models for predicting mental disorders perform better than unassisted physicians. A recent meta-analysis of the accuracy of AI diagnostic systems in the context of medical imaging and histopathology found the performance of deep-learning models to be equivalent to that of healthcare professionals (Liu et al., 2019). Others have pointed out that the presumption of accuracy may not hold (Hofmann, 2022). Two observations should

be made here. Firstly, as argued in a previous section, there are reasons to believe that an increased accuracy will likely drive stronger attempts of social control and increased stigmatization. One may also hypothesize that increased accuracy will drive increased disengagement from online data sharing. Whether or not the benefits and harms will increase with an increased accuracy is ultimately to be settled empirically. For present purposes, it should simply be noted that an increased accuracy does not necessarily make a difference for the balancing of benefits and harms in the three examples. Secondly, as argued right above, there are strong reasons for believing that the weighing of benefits and harms should in many cases befall the individual. The weight of benefits and harms—including the chance of obtaining the benefits and suffering the harms (accuracy)—cannot be separated from the interests of individuals and not in cases where the individual may come to suffer significant harms.

In conclusion, the arguments presented here suggest that individuals should have a right not to be AI profiled in all three cases. While the right not to be AI profiled based on publicly available data admittedly is a pro tanto right, what is claimed here is that none of the three cases seem to qualify as exemptions from this right.

6 The GDPR and the Right not to be AI Profiled

For the overall purpose in this article of establishing the need for a sui generis legal right not to be AI profiled, it remains to be considered whether current legislation already incorporates such a right. Within the EU, privacy rights in relation to automated profiling and data use are covered in the EU General Data Protection Regulation (GDPR) (General Data Protection Regulation, 2016). The GDPR contains no explicit provisions granting individuals a sui generis right not to be AI profiled. It may, however, entail a prohibition against this type of profiling and hence indirectly establish such a right in relation to the three cases.

6.1 Two Approaches to Deciding the Implications of the GDPR

AI profiling for mental health disorders based on publicly available social media data is a type of data processing involving both personal data and special categories of personal data, i.e. health data. The GDPR contains a general prohibition against the processing of special categories of personal data such as health data, but lists several exemptions in article 9, including the use of special categories of personal data that have manifestly been made public by the data subject. Further processing of personal data must abide by the principles in article 5. In short, personal data must (a) be processed lawfully, fairly, and transparently; (b) be collected and used in relation to a specific purpose; (c) be limited to what is necessary for the specific purpose; (d) be accurate; (e) be anonymized or erased when the processing of identifiable data is no longer necessary; and (f) be processed securely. Article 6 lists lawful types of data processing and includes among these processings necessary to carry out a task of public interest or in the exercise of official authority. The fairness and

transparency requirements are specified in articles 12 and 13. They may for present purposes be summarized as the requirement that the data subject is appropriately informed about the processing of personal data.

In the attempt to decide if the GDPR entails a right not to be AI profiled in the three cases, different approaches may be taken. One approach would be to analyze if the relevant examples of AI profiling can or cannot satisfy the aforementioned processing principles of the GDPR. If they cannot, the GDPR would effectively entail a right not to be AI profiled. Another approach—the one pursued here—would be to consider the scope of the principles and the extent to which the GDPR allows for them to be restricted. If the three cases fall outside the scope of the principles or they may be relevantly restricted, the GDPR would not ensure a right not to be AI profiled in the three cases.

6.2 The Three Cases and the Scope and Potential Restrictions of the GDPR Processing Principles

The AI profiling in the friends case presumably falls outside the scope of the GDPR. According to article 2, the GDPR does not apply to the processing of personal data in relation to purely personal or household activities that have no connection to professional or commercial activities. In the friends case, the processing seems most adequately classified as a case of processing for the sake of personal and household activities. It clearly has no connection to professional or commercial activities. In consequence, the processing requirements do not apply to the friends case, and the mental health profiling would be wholly lawful.

Pertinent to the public servant case, the GDPR allows for national adaptions of the processing principles in article 5 (see above). Thus, article 6 stipulates that member states may introduce more specific provisions with respect to data processing necessary for the performance of a task carried out in the public interest or in the exercise of official authority. Article 23 allows member states to restrict legislatively the processing obligations in articles 12–22 if this is necessary for safeguarding other important objectives of the general public interest, including financial interests related to public health and social security. With reference to the preamble no. 10, article 6 and article 23, the Danish Parliament for instance in late 2021 adopted a legislative proposal permitting the national tax authorities to collect 'blindly' all types of personal data publicly available on social media platforms and to use these data for risk profiling of all citizens (Lov om ændring af lov om et indkomstregister, skatteindberetningsloven og skattekontrolloven, 2021). The only requirement is that the authorities consider the data and profiling *necessary* for carrying out a task, e.g. the detection of taxation fraud. Analogously, the AI profiling in the public servant case may be considered necessary for carrying out a task of the social services and for maintaining an efficient public administration, and such profiling could reasonably be seen as protecting a financial interest related to social security. The conditions for restricting the processing principles are thus likely to be met in cases like the public servant case, and the GDPR therefore does not ensure individuals a right not to be AI profiled in such cases.

Finally, article 85 of the GDPR requires member states to reconcile the data protection regulation with the right to freedom of expression and information, where this includes the right to process data for journalistic purposes. As also stated in article 85, such processing may require that member states provide for exemptions and derogations from, among others, the processing principles in article 5. The GDPR thus allows for exemptions in cases like the prime minister case, where the mental health of the candidate arguably is of public interest. Hence, the GDPR does ensure political candidates a right not to be AI profiled in such cases.

6.3 Advantages of an Explicit Sui Generis Right not to be AI Profiled

Without considering the implications of the processing principles for the three cases, it has nonetheless been argued here that the GDPR does not ensure a right not to be AI profiled in the three cases. But even if the processing principles could be shown to entail a right not to be AI profiled in the three cases and even if all of the above arguments suggesting that the processing principles may be restricted in relation to these cases fail, there may still be reasons for introducing an explicit sui generis right not to be AI profiled into AI legislation. First, the cases do not exhaust the space of the potential use of profiling, which may for instance also be used by employers for recruitment and selection processes and for various other forms of workforce management (Hunkenschroer & Luetge, 2022; Kim & Bodie, 2021). Second, the severity of the potential harms of AI profiling based on publicly available data suggest that in the future, where AI predictive profiling may have become publicly accessible, there is a need of legislation that is effective in regulating our profiling behaviour. One may venture that a rather simple right that unambiguously and unquestionably rules out most cases of AI profiling is more effective than a data protection regulation with negotiable implications and the possibility of special national adaptions. By introducing such a right, an important and complex legal question would a priori have been settled, and this arguably makes a difference for the regulatory effect of legislation.

7 Conclusion—Future Research

This article has substantiated the right not to be subjected to AI profiling based on publicly available data. This right entails that such profiling cannot be conducted without the explicit informed consent of the profiled subject. The line of argument pursued in this article raises several questions that warrant further discussion.

Why a right? Even if the gravity of the potential harms of AI profiling is acknowledged, it may still be considered an open question whether such harms should be addressed through a right or some other type of regulation prohibiting such profiling in one way or another. As was previously argued, the right not to be AI profiled without informed consent may be taken to follow from an

individual's right to protect their interests and their right to protect themselves against harm. It should also be noted, however, that a right of this kind so to speak 'democratises' regulation. It does not confine AI regulation to designated institutions and professions endowed with the competence and obligation to monitor and control AI development and use, but rather, it gives every individual an instrument for the control of AI use. In so doing, a right may increase the regulatory effectiveness over and above alternative types of regulation. This argument cannot be fully developed in this article but should be explored in future work.

Why a pro tanto right? From the outset of this article, the right not to be AI profiled has been suggested as a conditional right. Yet the potential exemptions from this right have not been specified. On the contrary, it has been argued that some of the obvious exemptions—e.g. freedom of expression and information—do not constitute reasonable exemptions. This question also warrants further work. The reason for making this choice here is twofold. We have in this article at no point argued that there are no exemptions to this right. Relatedly, it seems that it is always possible to design situations constituting exemptions to such rights. For instance, threats to national security or threats of terrorism could be candidates for making up such exemptions.

Why informed consent? The right not to be AI profiled can be waived through informed consent. Some would object that the provision of informed consent to data processing is highly routinized, i.e. provided as an unreflected, habitual act, and that the right not to be AI profiled as a consequence may lack any real power to protect individuals (Ploug & Holm, 2013, 2015). The extent to which the problems of routinization may be overcome through careful design of the consent situation is a matter for future (empirical) work. The position taken in this article is that the weighing of the potential benefits and harms of AI profiling cannot be done ethically adequate without the involvement of the individuals potentially being profiled. The legitimacy of subjecting individuals to AI profiling at least partly derives from their involvement in the decision to do the profiling. As such informed consent is simply ethically necessary. The requirement of informed consent may not be sufficient to protect individuals from the potential harms of AI profiling. Other types of regulation protecting individuals from such harms may certainly be needed. Not least—as some have argued—because the requirement of informed consent may be a burden to individuals (Vredenburgh, 2022).

Finally, in the course of substantiating a sui generis right not to be AI profiled based on publicly available data, we have suggested several behavioural mechanisms and patterns that underlie social control and stigmatization and provided relevant evidence to the best of our ability. There is a need for further empirical studies on these matters. Note, however, that in the attempt to build our arguments around actual behavioural patterns and mechanism, this article constitutes a contribution to evidence-based policymaking in the field of privacy rights.

Abbreviations *AI*: Artificial intelligence; *GDPR*: EU General Data Protection Regulation; *EU*: European Union; *SNS*: Social networking sites

Acknowledgements I would like to thank Hanne Marie Motzfeld and Søren Holm for their valuable and insightful comments.

Author Contribution Not applicable.

Data Availability Not applicable.

Declarations

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The author declares no competing interests.

References

Ahn, J. (2012). Teenagers' experiences with social network sites: Relationships to bridging and bonding social capital. *The Information Society, 28*(2), 99–109.

Amini, H., Mohammadi, E., & Kosseim, L. (2022). Quick and (maybe not so) easy detection of anorexia in social media: To explainability and beyond, 141–158. In F. Crestani, D. E. Losada, & J. Parapar (Eds), *Early Detection of Mental Health Disorders by Social Media Monitoring: The First Five Years of the eRisk Project* . Springer International Publishing.

Antheunis, M. L., Schouten, A. P., & Krahmer, E. (2016). The role of social networking sites in early adolescents' social lives. *The Journal of Early Adolescence, 36*(3), 348–371.

Appel, M., & Weber, S. (2021). Do mass mediated stereotypes harm members of negatively stereotyped groups? A meta-analytical review on media-generated stereotype threat and stereotype lift. *Communication Research, 48*(2), 151–179.

Beam, M. A. (2016). Clicking vs. sharing: The relationship between online news behaviors and political knowledge. *Computers in Human Behavior*, 59, 215–220.

Beauchamp, T. L., & Childress, J. F. (2001). *Principles of biomedical ethics* (5th ed). Oxford University Press, New York.

Bhugra, D. (1989). Attitudes towards mental illness. *Acta Psychiatrica Scandinavica, 80*(1), 1–12.

Bishop, R. C. (2003). On separating predictability and determinism. *Erkenntnis, 58*, 169–188.

Boulianne, S. (2009). Does internet use affect engagement? A meta-analysis of research. *Political Communication, 26*(2), 193–211.

Boulianne, S. (2015). Social media use and participation: A meta-analysis of current research. *Information, Communication & Society, 18*(5), 524–538.

Brandtzæg, P. B. (2012). Social networking sites: Their users and social implications A longitudinal study. *Journal of Computer-Mediated Communication, 17*, 467–488.

Bright, L. F., Lim, H. S., & Logan, K. (2021). "Should I Post or Ghost?": Examining how privacy concerns impact social media engagement in US consumers. *Psychology & Marketing, 38*(10), 1712–1722.

Coppersmith, G., Leary, R., Crutchley, P., & Fine, A. (2018). Natural language processing of social media as screening for suicide risk. *Biomedical Informatics Insights, 10*, 1–11.

Corrigan, P. W., Larson, J. E., & Rüsch, N. (2009). Self-stigma and the "why try" effect: Impact on life goals and evidence-based practices. *World Psychiatry, 8*(2), 75–81.

Crandall, C. S., & Coleman, R. (1992). Aids-related stigmatization and the disruption of social relationships. *Journal of Social and Personal Relationships, 9*(2), 163–177.

Druss, B. G., Zhao, L., Von Esenwein, S., Morrato, E. H., & Marcus, S. C. (2011). Understanding excess mortality in persons with mental illness: 17-year follow up of a nationally representative US survey. *Medical Care, 49*(6), 599–604. JSTOR.

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "Friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication, 12*(4), 1143–1168.

Evans, J. P., & Burke, W. (2008). Genetic exceptionalism. Too much of a good thing? *Genetics in Medicine, 10*(7), Art. 7.

Feinberg, J. (1980). The child's right to an open future, 76-97. In Feinberg, J. (1992). *Freedom and Fulfillment: Philosophical Essays*. Princeton University Press, New Jersey.

Frederick, D. (2014). Pro-tanto versus absolute rights. *The Philosophical Forum, 45*(4), 375–394.

Garrett, J. R., Lantos, J. D., Biesecker, L. G., Childerhose, J. E., Chung, W. K., Holm, I. A., Koenig, B. A., McEwen, J. E., Wilfond, B. S., & Brothers, K. (2019). Rethinking the "open future" argument against predictive genetic testing of children. *Genetics in Medicine, 21*(10), 2190–2198.

Gkotsis, G., Oellrich, A., Velupillai, S., Liakata, M., Hubbard, T. J. P., Dobson, R. J. B., & Dutta, R. (2017). Characterisation of mental health conditions in social media using Informed Deep Learning. *Scientific Reports, 7*(1), Art. 1.

Goffman, E. (1963). *Stigma: Notes on the management of spoiled identity*. Simon & Schuster, New York.

Goldstein, J. (1991). The stigmatization of smokers: An empirical investigation. *Journal of Drug Education, 21*(2), 167–182.

Gruzd, A., & Hernández-García, Á. (2018). Privacy concerns and self-disclosure in private and public uses of social media. *Cyberpsychology, Behavior, and Social Networking, 21*(7), 418–428.

Herek, G. M. (1999). AIDS and stigma. *American Behavioral Scientist, 42*(7), 1106–1116.

Hilbert, A., Rief, W., & Braehler, E. (2008). Stigmatizing attitudes toward obesity in a representative population-based sample. *Obesity, 16*(7), 1529–1534.

Hofmann, B. (2022). Too much, too mild, too early: Diagnosing the excessive expansion of diagnoses. *International Journal of General Medicine, 15*, 6441–6450.

Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-enabled recruiting and selection: A review and research agenda. *Journal of Business Ethics, 178*, 977–1007.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399.

Kahlow, J. A., Coker, M. C., & Richards, R. (2020). The multimodal nature of Snapchat in close relationships: Toward a social presence-based theoretical framework. *Computers in Human Behavior, 111*, 106409.

Keller, J. (2005). In genes we trust: The biological component of psychological essentialism and its relationship to mechanisms of motivated social cognition. *Journal of Personality and Social Psychology, 88*(4), 686–702.

Kim, P., & Bodie, M. T. (2021). Artificial intelligence and the challenges of workplace discrimination and privacy. *35 ABA Journal of Labor and Employment Law 289 Saint Louis U Legal Studies Research Paper No, 2021–26*, 289–315.

Kim, J., Lee, J., Park, E., & Han, J. (2020). A deep learning model for detecting mental illness from user content on social media. *Scientific Reports, 10*(1), Art. 1.

Kumar, A., Sharma, A., & Arora, A. (2019). Anxious depression prediction in real-time social data. *Proceeding of International Conference on Advanced Engineering, Science, Management and Technology,* 1–7.

Kümpel, A. S., Karnowski, V., & Keyling, T. (2015). News sharing in social media: A review of current research on news sharing users, content, and networks. *Social Media + Society, 1*(2), 2056305115610141.

Leino-Kilpi, H., Välimäki, M., Dassen, T., Gasull, M., Lemonidou, C., Scott, A., & Arndt, M. (2001). Privacy: A review of the literature. *International Journal of Nursing Studies, 38*(6), 663–671.

Link, B. G. (1987). Understanding labeling effects in the area of mental disorders: An assessment of the effects of expectations of rejection. *American Sociological Review, 52*(1), 96–112. JSTOR.

Link, B. G., & Phelan, J. C. (2001). Conceptualizing stigma. *Annual Review of Sociology*, 27, 363–385.

Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., Mahendiran, T., Moraes, G., Shamdas, M., Kern, C., Ledsam, J. R., Schmid, M. K., Balaskas, K., Topol, E. J., Bachmann, L. M., Keane, P. A., & Denniston, A. K. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. *The Lancet Digital Health, 1*(6), e271–e297.

Lov om ændring af lov om et indkomstregister, skatteindberetningsloven og skattekontrolloven, nr. L 73, 1 (2021). https://www.ft.dk/ripdf/samling/20211/lovforslag/l73/20211_l73_som_fremsat.pdf

McGuire, A. L., Fisher, R., Cusenza, P., Hudson, K., Rothstein, M. A., McGraw, D., Matteson, S., Glaser, J., & Henley, D. E. (2008). Confidentiality, privacy, and security of genetic and genomic test information in electronic health records: Points to consider. *Genetics in Medicine, 10*(7), 495–499.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44*(12), 1469–1481.

Mojtabai, R., Olfson, M., Sampson, N. A., Jin, R., Druss, B., Wang, P. S., Wells, K. B., Pincus, H. A., & Kessler, R. C. (2011). Barriers to mental health treatment: Results from the National Comorbidity Survey Replication. *Psychological Medicine, 41*(8), 1751–1761.

Myers, A., & Rosen, J. C. (1999). Obesity stigmatization and coping: Relation to mental health symptoms, body image, and self-esteem. *International Journal of Obesity, 23*(3), 221–230.

Nozick, R. (1969). Coercion, 440-472. In Morgenbesser, W. (Ed). *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*. St Martin's Press.

Peretti-Watel, P., Legleye, S., Guignard, R., & Beck, F. (2014). Cigarette smoking as a stigma: Evidence from France. *International Journal of Drug Policy, 25*(2), 282–290.

Ploug, T. (2020). In Defence of informed consent for health record research—Why arguments from 'easy rescue', 'no harm' and 'consent bias' fail. *BMC Medical Ethics, 21*(1), 75.

Ploug, T., & Holm, S. (2013). Informed consent and routinisation. *Journal of Medical Ethics, 39*(4), 214–218.

Ploug, T., & Holm, S. (2015). Routinisation of informed consent in online health care systems. *International Journal of Medical Informatics, 84*(4), 229–236.

Ploug, T., Holm, S., & Gjerris, M. (2015). The stigmatization dilemma in public health policy-the case of MRSA in Denmark. *BMC Public Health, 15*(1), 640.

Pouwels, J. L., Valkenburg, P. M., Beyens, I., van Driel, I. I., & Keijsers, L. (2021). Social media use and friendship closeness in adolescents' daily lives: An experience sampling study. *Developmental Psychology, 57*(2), 309.

Powers, P. (2007). Persuasion and coercion: A critical review of philosophical and empirical approaches. *HEC Forum, 19*(2), 125–143.

Rangel, U., & Keller, J. (2011). Essentialism goes social: Belief in social determinism as a component of psychological essentialism. *Journal of Personality and Social Psychology, 100*(6), 1056.

Reece, A. G., & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science, 6*(1), 15.

Reece, A. G., Reagan, A. J., Lix, K. L. M., Dodds, P. S., Danforth, C. M., & Langer, E. J. (2017). Forecasting the onset and course of mental illness with Twitter data. *Scientific Reports, 7*(1), Art. 1.

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 L 119, (2016). http://data.europa.eu/eli/reg/2016/679/oj

Rummens, S., & Cuypers, S. E. (2010). Determinism and the paradox of predictability. *Erkenntnis, 72*(2), 233–249.

Skoric, M. M., Zhu, Q., Goh, D., & Pang, N. (2016). Social media and citizen engagement: A meta-analytic review. *New Media & Society, 18*(9), 1817–1839.

Stuart, H. (2006). Mental illness and employment discrimination. *Current Opinion in Psychiatry, 19*(5), 522–526.

Stuber, J., Meyer, I., & Link, B. (2008). Stigma, prejudice, discrimination and health. *Social Science & Medicine, 67*(3), 351–357.

Stuber, J., Galea, S., & Link, B. G. (2009). Stigma and smoking: The consequences of our good intentions. *Social Service Review, 83*(4), 585–609.

Valkenburg, P. M., Meier, A., & Beyens, I. (2022). Social media use and its impact on adolescent mental health: An umbrella review of the evidence. *Current Opinion in Psychology, 44*, 58–68.

Voorhees, B. W. V., Fogel, J., Houston, T. K., Cooper, L. A., Wang, N.-Y., & Ford, D. E. (2005). Beliefs and attitudes associated with the intention to not accept the diagnosis of depression among young adults. *The Annals of Family Medicine, 3*(1), 38–46.

Vredenburgh, K. (2022). The right to explanation*. *Journal of Political Philosophy, 30*(2), 209–229.

Wertheimer, A. (1990). *Coercion*. Princeton University Press.

Yzerbyt, V., Rocher, S., & Schadron, G. (1997). Stereotypes as explanations: A subjective essentialistic view of group perception, 20-50. In R. Spears, P. J. Oakes, N. Ellemers, & S. A. Haslam (Eds.). *The social psychology of stereotyping and group life*. Blackwell Publishing.

Zirikly, A., Resnik, P., Uzuner, Ö., & Hollingshead, K. (2019). CLPsych 2019 Shared task: Predicting the degree of suicide risk in Reddit posts, 24-33. *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*. Minneapolis, Minnesota.