

RESEARCH

Open Access

Big Data and discrimination: perils, promises and solutions. A systematic review



Maddalena Favaretto^{*}, Eva De Clercq and Bernice Simone Elger

^{*}Correspondence:
maddalena.favaretto@unibas.ch
Institute for Biomedical
Ethics, University of Basel,
Bernoullistrasse 28,
4056 Basel, Switzerland

Abstract

Background: Big Data analytics such as credit scoring and predictive analytics offer numerous opportunities but also raise considerable concerns, among which the most pressing is the risk of discrimination. Although this issue has been examined before, a comprehensive study on this topic is still lacking. This literature review aims to identify studies on Big Data in relation to discrimination in order to (1) understand the causes and consequences of discrimination in data mining, (2) identify barriers to fair data-mining and (3) explore potential solutions to this problem.

Methods: Six databases were systematically searched (between 2010 and 2017): PsychINDEX, SocIndex, PhilPapers, Cinhal, Pubmed and Web of Science.

Results: Most of the articles addressed the potential risk of discrimination of data mining technologies in numerous aspects of daily life (e.g. employment, marketing, credit scoring). The majority of the papers focused on instances of discrimination related to historically vulnerable categories, while others expressed the concern that scoring systems and predictive analytics might introduce new forms of discrimination in sectors like insurance and healthcare. Discriminatory consequences of data mining were mainly attributed to human bias and shortcomings of the law; therefore suggested solutions included comprehensive auditing strategies, implementation of data protection legislation and transparency enhancing strategies. Some publications also highlighted positive applications of Big Data technologies.

Conclusion: This systematic review primarily highlights the need for additional empirical research to assess how discriminatory practices are both voluntarily and accidentally emerging from the increasing use of data analytics in our daily life. Moreover, since the majority of papers focused on the negative discriminative consequences of Big Data, more research is needed on the potential positive uses of Big Data with regards to social disparity.

Keywords: Big Data, Data analytics, Unfair discrimination, Disparity, Inequality, Ethics

Introduction

Big Data has been described as a “one-size-fits-all (so long as it’s triple XL) answer” [24] to solve some of the most challenging problems in the fields of climate change, healthcare, education and criminology. This may explain why it has become the buzzword of the decade. Big Data is a very complex and extensive phenomenon that has had fluctuating meanings since its appearance in the early 2010’s [86]. Traditionally it has been defined in terms of four dimensions (the four V’s of Big Data): volume, velocity,

variety, and veracity—although some scholars also include other characteristics such as complexity [63] and value [5]—and it consists of capturing, storing, analyzing, sharing and linking huge amount of data created through computer-based technologies and networks, such as smartphones, computers, cameras, sensors etc. [40]. As we live in an increasingly networked world, where new forms of data sources and data creation abound (e.g., video sharing, online messaging, online purchasing, social media, smartphones), the amount and variety of data that is collected from individuals has increased exponentially, ranging from structured numeric data to unstructured text documents such as email, video, audio and financial transactions (SAS-Institute) [72].

Interestingly, due to the fact that traditional computational systems are unable to process and work on Big Data, characteristics of this phenomenon have been described by scholars in strict relation to the technical challenges they raise: volume and velocity, for example, present the most immediate challenge to traditional IT structures since companies do not have the necessary infrastructures to collect, store and process the vast amount of data that is created at increasingly higher speeds; variety refers to the heterogeneity of both structured and unstructured data that is collected from very different sources making storage and processing even more complex; and finally, since Big Data technologies are dealing with high volume, velocity and great variety of qualitatively very heterogeneous data, it is highly improbable that the resulting data set will be completely accurate or trustworthy, creating issues of veracity [5].

Despite the aforementioned issues, we should not forget that Big Data analytics—understood here as the plethora of advanced digital techniques (e.g. data mining, neural networks, deep learning, profiling, automatic decision making and scoring systems) designed to analyze large datasets with the aim of revealing patterns, trends and associations, related to human behavior—play an increasingly important role in our everyday life: the decision to accept or deny a loan, to grant or deny parole, or to accept or decline a job application are influenced by machines and algorithms rather than by individuals. Data analysis technologies are thus becoming more and more entwined with people's sensitive personal characteristics, their daily actions and their future opportunities. Hence it should not come as a surprise that many scholars have started to scrutinize Big Data technologies and their applications to analyze and grasp the novel ethical and societal issues of Big Data. The most common concerns that arise regard privacy and data anonymity [26, 29], informed consent [41], epistemological challenges [28], and more conceptual concerns such as the mutation of the concept of personal identity due to profiling [27] or the analysis of surveillance in an increasing “datafication” or “data-fied” society [7].

One of the most worrying but still under researched aspects of Big Data technologies is the risk of potential discrimination. Although “there is no universally accepted definition of discrimination” [82], the term generally refers to acts, practices or policies that impose a relative disadvantage on persons because of their membership of a salient social or recognized vulnerable group based on gender, race, skin color, language, religion, political opinion, ethnic minority etc. [61]. For the scope of our study we adhere to the aforementioned general conception of discrimination and only distinguish between *direct discrimination* (i.e. procedures that discriminate against minorities or disadvantaged groups on the basis of sensitive discriminatory attributes related to group

membership such as race, gender or sexual orientation) and *indirect discrimination* (i.e. procedures that might intentionally or accidentally discriminate against a minority, while not explicitly mentioning discriminatory attributes) [32]. We also acknowledge the close connection between discrimination and inequality, since a disadvantage caused by discrimination necessarily leads to inequality between the considered groups [75].

Although research on discrimination in data mining technologies is far from new [69], it has gained momentum recently, in particular after the publication of the White House report of 2014 which firmly warned that discrimination might be the inadvertent outcome of Big Data technologies [65]. Since then, possible discriminatory outcomes of profiling and scoring systems have increasingly come to the attention of the general public. In the United States, for example, a system technology used for the assessment of future risk of re-offending among defendants was found to discriminate against black people [23]. Likewise, in the United Kingdom, an algorithm used to make custodial decisions was found to discriminate against people with lower incomes [15]. But more citizen-centered applications, such as the Boston’s Street Bump App, which is developed to detect potholes on roads are also potentially discriminatory. By relying on the use of a smartphone, the App, risks increasing the social divide between neighborhoods with a higher number of older or less affluent citizens and those more wealthy areas with more young smartphone owners [67].

The proliferation of these cases explains why discrimination in Big Data technologies has become a hot topic in a wide range of disciplines, ranging from computer science and marketing to philosophy, resulting in a scattered and fragmented multidisciplinary corpus that makes it difficult to fully access the core of the issue. Our literature review therefore aims to identify relevant studies on Big Data in relation to discrimination from different disciplines in order to (1) understand the causes and consequences of discrimination in data analytics; (2) to identify barriers to fair data-mining and (3) explore suggested solutions to this problem.

Methods

A systematic literature review was performed by searching the following six databases: PsycINFO, SocINDEX, PhilPapers, Cinhal, Pubmed and Web of Science (see Table 1).

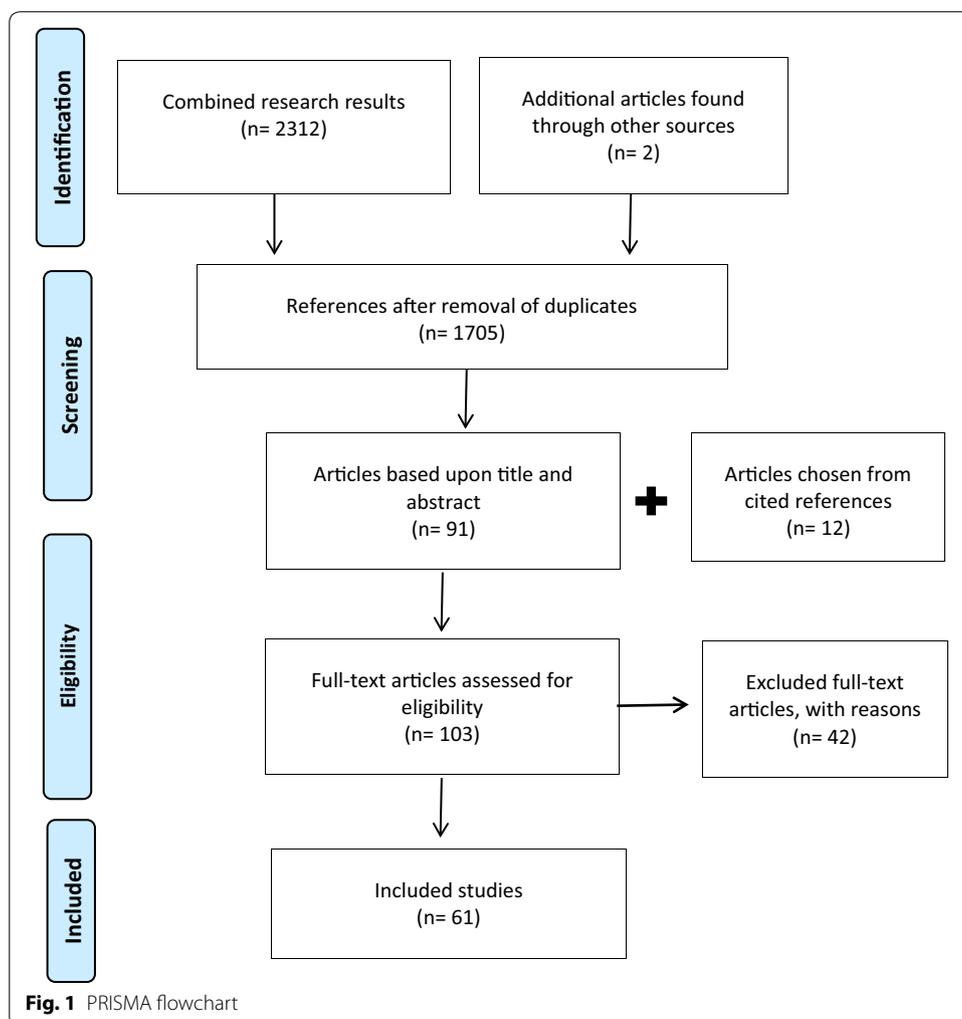
The following search terms were used: “big data”, “digital data”, “data mining”, “data linkage”, “discriminat*”, “*equality”, “vulnerab*”, “*justice”, “ethic*” and “exclusion”. The terms were combined using Boolean logic (see Table 1). The inclusion criteria were: (1) papers published between 2010 and December 2017 and (2) written in English. A

Table 1 Search terms

No.	Matches search terms	PsychInfo	PhilPapers	SocIndex	CINAHL	PubMed	Web of science
1	“Big data” OR “digital data” OR “data mining” OR “data linkage”	2385	179	507	944	13214	23740
2	Discriminat* OR *equality OR vulnerab* OR *justice OR ethic* OR exclusion	69,435	46,349	46,624	38,096	245,604	414,661
3	1 AND 2	156	67	88	55	769	1177

relatively narrow publication window was chosen as “Big Data” has become a buzzword in academic circles only over the last decade and because we wanted to target only those articles that focus on the latest digital technologies for profiling and predictive analysis. In order to obtain a broader understanding of discrimination and inequality related to Big Data, no restriction was placed on the discipline of the papers (medicine, psychology, sociology, computer science, etc.), or on the type of methodology (quantitative, qualitative, mixed methods or theoretical). Books (monographs and edited volumes), conference proceedings, dissertations, literature reviews and posters were omitted.

The search protocol from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method [57] was followed and resulted in 2312 papers (see Fig. 1). Two papers were added that were identified through other sources. The results were scanned for duplicates (609) and 1705 remained. In this phase, we included all articles that mentioned, discussed, enumerated or described discrimination, the digital divide or social inequality related to Big Data (from data mining and predictive analysis to profiling). Therefore, papers that focused mainly on issues of autonomy, privacy and consent were excluded, together with those that merely described means to recognize



or classify individuals using digital technologies without acknowledging the risk of discrimination. Disagreements between the first and second authors were evaluated by a third reviewer who determined which articles were eligible based on their abstracts. In total, 1559 records were excluded.

The first author subsequently scanned the references of the remaining 91 articles to identify additional relevant studies. 12 papers were added through this process. The final sample included 103 articles. During the next phase, the first author read the full texts. After thorough evaluation, 42 articles were excluded because (1) they did not or only superficially referred to discrimination or inequality in relation to Big Data technologies and focused more on risks related to privacy or consent; (2) they discussed discrimination but not in relation to the development of Big Data analytic technologies; (3) they focused on the growing divide between organizations that have the power and resources to access, analyze and understand Big Datasets (“the Big Data rich”) and those that do not (“the Big Data poor”) [4] instead of on the concept of Digital Divide, which is defined as the gap between individuals who have easy access to internet-based technologies and those who do not; or (4) they assessed disparities affecting participation in social media. The subsequent phase of the literature review involved the analysis of the remaining 61 articles. The following information was extracted from the papers: year of publication, country, discipline, methodology, type of discrimination/inequality fostered by data mining technologies, suggested solutions to the discrimination/inequality issue, beneficial applications of Big Data to contrast discrimination/inequality, reference to the digital divide, reference to the concept of the Black Box as an aggravator of discrimination, evaluation of the human element in data mining, mention of the shift from individual to group harm, reference to conceptual challenges introduced by Big Data, and mention of legal shortcomings when confronted with Big Data technologies.

Results

Among the 61 papers included in our analysis, 38 were theoretical papers that critically discussed the relation between discrimination, inequality and Big Data technologies. Of the remaining 23 articles, 7 employed quantitative methods, 3 qualitative methods and 13 computer science methodologies that used a theory to combat or analyze discrimination in data mining and then empirically tested this theory on a data set. To distinguish the latter approach from the more traditional empirical research methods, we classified such studies as “other” (experimental) methods. Most of the papers were published after 2014 ($n=44$), the year of the publication of the White House report on the promises and challenges of Big Data [65]. Almost one-third of the studies ($n=22$) were from the United States, 6 came from the Netherlands, 3 from the United Kingdom and the remaining ones were from Belgium, Spain, Germany, France, Australia, Ireland, Italy, Canada, or Israel. Ten papers were from more than one country (see table). Regarding the scientific discipline, 20 papers were published in papers from the field of Social Sciences, 14 from Computer Science, 14 from Law, 9 from Bioethics and only 2 from Philosophy and Ethics. As to the field of application, a considerable number of papers ($n=24$) discussed discriminatory practices in relation to *various* aspects of daily living such as employment, advertisement, housing, insurance, credit scoring etc., while others focused on *one* specific area.

The majority of the studies ($n=38$) did not provide a definition of discrimination, but instead treated the word as self-explanatory and frequently linked it to others concepts such as inequality, injustice and exclusion. A few defined discrimination as “disparate impact”, “disparate treatment”, “redlining”, “statistical discrimination”, while others gave a more “juridical” definition and referred to the unequal treatment of “legally protected classes”, or directly referred to existing national or international legislation. Only one article discussed the difference between direct and indirect discrimination (see Table 2).

Discrimination and data mining

In order to explore whether and how Big Data analysis and/or data mining techniques can have discriminatory outcomes, we decided to divide the studies according to (a) the possible discriminatory outcomes of data analytics and (b) some of the most commonly identified causes of discrimination or inequality in Big Data technologies.

Forms, targets and consequences of discrimination

Numerous papers assessed the possible various discriminative and unfair outcomes that might result from data technologies (see Table 3).

Among these, a considerable number of papers highlighted the two main forms of discrimination introduced by data mining. In this context, some authors stressed the fact that the aforementioned algorithmic mechanisms might result in involuntary and accidental discrimination [8, 14, 17, 21, 25, 39, 45, 54, 73, 93]. Barocas and Selbst [8], for example, claimed that “when it comes to data mining, unintentional discrimination is the more pressing concern because it is likely to be far more common and easier to overlook” [8] and expressed concern about the possibility that classifiers in data mining could contain unlawful and harmful discrimination towards protected classes and or vulnerable groups. Holtzhausen, along the same lines, argued that “algorithms can have unintended consequences” [39] and might cause real harm to individuals, ranging from differences in pricing, to employment practices, to police surveillance. Some other studies instead highlighted that data mining technologies could result in direct and voluntary discrimination [32, 39, 46]. Here we follow the aforementioned definition of direct discrimination offered by [32] that describes it as discrimination against minorities or disadvantaged groups on the basis of sensitive discriminatory attributes related to group membership such as race, gender or sexual orientation. Holtzhausen, for instance, warned against the discriminatory use of ethnic profiling in housing and surveillance [1, 39] discussed potentially oppressive and discriminatory outcomes of data mining on migration and profiling that impose an automatic and arbitrary classification and categorization upon supposedly risky travelers.

Some papers also defined the potential targets of data mining technologies [46, 58] discussed the increased exploitation of the vulnerable as one of the most worrying consequences of data mining; they claimed that algorithms might identify those who are less capable, such as elder individuals with gambling habits, and prey on them with targeted advertisements or by persuading them “to take out risky loans, or high-rate instant credit options, thereby exploiting their vulnerability” [58]. Leese [48] claimed that discrimination is one of the harms that derives from the massive scale of the profiling of society and that the risk is even higher for vulnerable populations. Four of the reviewed

Table 2 List of included articles

Author, Year, Country	Design	Participants	Discipline	Field of application	Definition of discrimination	Reference to legislation/regulatory text
Ajana (2015) [1], UK	Theoretical		Social Sciences	Migration	Unequal treatment	
Ajunwa et al. (2016) [2], USA	Theoretical		Bioethics	Employment	Not given—self explanatory	
Bakken and Reame (2016) [6], USA	Theoretical		Bioethics	Healthcare research	Not applicable—digital divide	
Barocas and Selbst (2016) [8], USA	Theoretical		Law	Employment	Disparate treatment/disparate impact	
Berendt and Preibusch (2014) [10], Belgium-UK	Other		Computer Science	Various	Juridical—legally protected classes	
Berendt and Preibusch (2017) [11], Belgium-UK	Other		Computer Science	Various	Illegitimate discrimination on grounds of four protected attributes	
Boyd and Crawford (2012) [12], Australia-USA	Theoretical		Social Sciences	Digital divide in research	Not applicable—digital divide	
Brannon (2017) [13], USA	Theoretical		Social Sciences	Social disparity	Not given—inequality	
Brayne (2017) [14], USA	Qualitative	A sample of Employees of LAPD (Officers and Civilians)	Social Sciences	Policing/criminology	Not given—inequality	
Calderis and Verwer (2010) [17], Netherlands	Other		Computer Science	Various	Not given—self explanatory	
Casanas i Comabella and Wanat (2015) [18], UK	Theoretical		Bioethics	Digital divide in research	Not applicable—digital divide	
Cato et al. [19], USA	Theoretical		Bioethics	Healthcare	Not given—injustice	Belmont Report; 1976
Chouldechova (2017) [20], USA	Other	A sample of Caucasian/African American US Defendants	Computer Science	US criminal justice system	Disparate impact	
Citron and Pasquale (2014) [21], USA	Theoretical		Law	Credit scoring	Not given—reference to protected classes	
Cohen et al. (2017) [22], USA	Theoretical		Bioethics	Healthcare	Not given—inequality	
d'Alessandro et al. (2017) [25], USA	Theoretical		Computer Science	Various	Disparate treatment/disparate impact	
de Vries (2010) [27], Belgium	Theoretical		Philosophy	Various	Unwarranted discrimination	
Francis and Francis (2017) [30], USA	Theoretical		Law	Healthcare and healthcare research	Not given—stigmatization and harm	
Hajian and Domingo-Ferrer (2013) [32], Spain	Other		Computer Science	Various	Not given—self explanatory	

Table 2 (continued)

Author, Year, Country	Design	Participants	Discipline	Field of application	Definition of discrimination	Reference to legislation/regulatory text
Hajian et al. (2014) [33], Spain	Other		Computer Science	Various	Unfair or unequal treatment	Australian Legislation 2008; European Union Legislation 2009
Hajian et al. (2015) [34], Italy-Spain	Other		Computer Science	Various	Unfair or unequal treatment	Australian Legislation 2014; European Union Legislation 2014
Hildebrandt and Koops (2010) [35], USA	Theoretical		Law	Ambient intelligence	Unlawful/unfair discrimination	
Hirsch (2015) [36], USA	Theoretical		Law	Various	Not given—elusive concept	
Hoffman (2010) [37], USA	Theoretical		Social Sciences	Employment	Unlawful discrimination on basis of disability	Americans with Disabilities Act (ADA), 1990; Genetic Information Nondiscrimination Act (GINA), 2003; Health Insurance Portability and Accountability Act (HIPAA), 1996
Hoffman (2017) [38], USA	Theoretical		Social Sciences	Employment	Unlawful discrimination on basis of disability	Americans with Disabilities Act (ADA), 1990; Genetic Information Nondiscrimination Act (GINA), 2003; Health Insurance Portability and Accountability Act (HIPAA), 1996
Holzhausen (2016) [39], USA	Theoretical		Social Sciences	Various	Not given—self explanatory	Australian Sex Discrimination Act, 1984; US Equal Pay Act, 1963; US Equal Credit Opportunity Act, 1974; European Council Directive, 2004
Kamiran and Calders (2012) [42], Netherlands-UK	Other		Computer Science	Various	Unfair and unequal treatment	
Kamiran et al. (2013) [43], Netherlands-Saudi Arabia-UK	Other		Computer Science	Various	Unfair and unequal treatment	
Kennedy and Moss (2015) [44], UK	Theoretical		Social Sciences	Society and culture	Not given—self explanatory	
Kroll et al. (2017) [45], USA	Theoretical		Law	Various	Not given—opposite of fair treatment	
Kuempel (2016) [46], USA	Theoretical		Law	Various	Not given—self explanatory	
Le Meur et al. (2015) [47], France	Quantitative	A sample of pregnant women	Bioethics	Healthcare	Not given	Australian Sex Discrimination Act, 1984; US Equal Pay Act, 1963

Table 2 (continued)

Author, Year, Country	Design	Participants	Discipline	Field of application	Definition of discrimination	Reference to legislation/regulatory text
Leese (2014) [48], Germany	Theoretical		Ethics	Aviation/migration	Principle of equality and non discrimination	[60]; European Convention on Human Rights, 1953; Treaty on the Functioning of the European Union, 1958
Lerman (2013) [49], USA	Theoretical		Law	Digital divide in social participation	Social marginalization/exclusion	
Lupton (2015) [51], Australia	Theoretical		Social Sciences	Society	Not given—stigmatization	
MacDonnell (2015) [53], Ireland	Theoretical		Social Sciences	Insurance	Not given	
Mantelero (2016) [54], China-Italy	Theoretical		Social Sciences	Various	Unjust or prejudicial treatment	
Mao et al. (2015) [55], USA	Quantitative	A sample of citizens from Cote D'Ivoire	Social Sciences	Economic development	Not given—related to social and economic disparity	
Newell and Marabelli (2015) [58], UK-USA	Theoretical		Social Sciences	Various	Not given—Harm towards vulnerable individuals	
Nielsen et al. (2017) [58], Brasil-USA	Quantitative	A sample of Twitter users in Brazil	Social sciences	Public health	Not given—self explanatory	
Pak et al. (2017) [60], Belgium	Quantitative	Citizens of Brussels using "Fix My Street" App	Social Science	Urban and social involvement	Not given—social exclusion/disparity	
Peppet (2014) [62], USA	Theoretical		Law	Various	Illegal or unwanted discrimination	
Ploug and Holm (2017) [64], Denmark	Theoretical		Bioethics	Society	Differential treatment and stigmatization	
Pope and Sydnor (2011) [66], USA	Other	Full sample of UI claimants from the State of New Jersey between 1995 and 1997	Computer Science	Employment	Not given—self explanatory	
Romei et al. (2013) [70], Italy	Quantitative	Italian female researchers	Computer Science	Academia	Unjustified distinction of individuals based on their membership	European Union Legislation, 2010
Ruggieri et al. (2010) [71], Italy	Other		Computer Science	Various	Juridical	Australian Legislation, 2010; European Union Legislation, 2010; United Nations Legislation, 2010; U.K. Legislation, 2010; U.S. Federal Legislation, 2010
Sharon (2016) [74], Netherlands	Theoretical		Bioethics	Healthcare and Healthcare Research	Not given—self explanatory	

Table 2 (continued)

Author, Year, Country	Design	Participants	Discipline	Field of application	Definition of discrimination	Reference to legislation/regulatory text
Schermer (2011) [73], Netherlands	Theoretical		Social Sciences	Not Defined	Not given—self explanatory/Stigmatization	
Susewind [76], Germany	Quantitative	Selected Asian countries	Social Sciences	Various	Not given—self explanatory	
Taylor (2016) [78], Netherlands	Qualitative	West Africa Population (Cote d’Azur)	Social Sciences	Surveillance	Not given—self explanatory	
Taylor (2017) [79], Netherlands	Theoretical		Social Sciences	Various	Disparity/inequality/exclusion	
Timmis et al. (2016) [80], UK	Theoretical		Social Sciences	Education	Not given—social exclusion/disparity	
Turow et al. (2015) [81], USA	Theoretical		Social Sciences	Marketing	Social discrimination	
Vaz et al. (2017) [83], Canada	Quantitative		Social Sciences	Urban development	Social inequalities	
Veale (2017) [84], UK	Theoretical		Social Sciences	Various	Not given—opposite of fairness and equality	
Voigt (2017) [85], Canada	Theoretical		Social Sciences	Healthcare	Inequality	
Zarate et al. (2016) [91], USA	Qualitative	Participants of the PGP (Personal Genome Project)	Bioethics	Various	Not given—self explanatory	
Zarsky (2014) [93], Israel	Theoretical		Law	Various	Illusive concept—unfair or Unequal Treatment of the individual	
Zarsky (2016) [92], Israel	Theoretical		Law	Credit scoring	Unfairness and inequality	
Zliobaite and Custers (2016) [95], Finland-Netherlands	Other		Computer Science	Various	Juridical	Race Equality Directive (2000/43/EC), Employment Equality Directive (2007/78/EC), Gender Recast Directive (2006/54/EC), Gender Goods and Services Directive (2006/113/EC)
Zliobaite (2017) [94], Finland-Netherlands	Other		Computer Science	Various	Adversary treatment of people based on belonging to some group	Race Equality Directive (2000/43/EC), Employment Equality Directive (2007/78/EC), Gender Recast Directive (2006/54/EC), Gender Goods and Services Directive (2006/113/EC)

papers also noticed how profiling and data mining technologies are causing a shift in harm from single profiled and classified individuals to larger groups. The papers argued that decisions taken on the aggregation of collected information might have harmful consequences for (a) the entire collectivity of the people involved in the data set [53], (b) for people who were not in the original analyzed dataset [30], and (c) for the general public due to the penetration of data mining practices into each of our every day's activity thanks to big companies like Facebook, Twitter, Google [44]. de Vries [27], has taken this concept a step further and argued that the increased use of machine profiling and automatic classification could lead to a general increase of discrimination in many sectors to a level that might make discrimination perceived as a legitimate practice in a constitutional democracy.

Regarding the consequences of the use of Big Data technologies, social exclusion, marginalization and stigmatization were mentioned in 11 articles. Lupton [51] argued that the disclosure of sensitive data, specifically sexual preference and health data related to fertility and sexual activity could result in stigma and discrimination. Ploug [63] described how health registries for sexual transmittable diseases risk singling out and excluding minorities, Barocas and Selbst [8], Pak et al. [59], and Taylor [78] argued that some individuals will be marginalized and excluded from social engagement due to the digital divide.

According to the literature, Big Data technologies might also perpetuate existing social and geographical historical disparities and inequalities, for example by increasing the exclusion of ethnic minorities from social engagement, worsening the living conditions of the economically disadvantaged, widening the economic gap between poor and rich countries, excluding some minorities from healthcare [13, 14, 60, 79, 80, 85], and/or delivering a fragmented and incomplete picture of the population through data mining technologies [13].

Some papers also highlighted how new means of automated decision making and personalization could create novel forms of discrimination that transcend the historical concept of unlawful discrimination and that are not related to historically protected classes or vulnerable categories. According to Newell and Marabelli [58], individuals could be inexplicably and unexpectedly excluded from certain opportunities, exploited on the basis of their lack of capacities, and be unfairly treated through targeted advertisement and profiling. The reviewed literature pinpointed two main new forms of discrimination: first, economic or marketing discrimination, that is, the unequal treatment of different consumers based on their purchasing habits or inequality in pricing and offers that are given to costumers based on profiling, such as insurance or housing [35, 62, 81]; secondly, discrimination based on health prediction, that is the unequal treatment or discrimination of individuals based on predictive, and not actual, health data [2, 22, 37, 38].

Causes of discrimination

Many papers highlighted the main elements that might cause discrimination or inequality in Big Data technologies (see Table 4).

Table 3 Discriminatory outcomes of Big Data

Discriminatory outcomes	Paper references
1. Forms of discrimination	
1.1. Accidental/involuntary discrimination	Calders and Verwer 2010 [17], Schermer 2011 [73], Citron and Pasquale 2014 [21], Zarsky 2014 [93], Barocas and Selbst 2016 [8], Holtzhausen 2016 [39], Mantelero 2016 [54], Brayne 2017 [14], Chouldechova 2017 [20], d'Alessandro et al. 2017 [25], Kroll et al. 2017 [45]
1.2. Direct voluntary discrimination	Ajana 2015 [1], Holtzhausen 2016 [39], Kuempel 2016 [46]
2. Victims/targets of discrimination	
2.1. Vulnerable groups/populations	Leese 2014 [48], Newell and Marabelli 2015 [58], Kuempel 2016 [46]
2.2. Larger groups	de Vries 2010 [27], Kennedy and Moss 2015 [44], Mantelero 2016 [54], Francis and Francis 2017 [30]
3. Discriminatory consequences	
3.1. Social marginalization and stigma	Lerman 2013 [49], Casanas i Comabella and Wanat 2015 [18], Kennedy and Moss 2015 [44], Lupton 2015 [51], Susewind 2015 [76], Barocas and Selbst 2016 [8], Sharon 2016 [73], Francis and Francis 2017 [30], Pak et al. 2017 [60], Ploug and Holm 2017 [64], Taylor 2017 [79]
3.2. Exacerbation of existing inequalities	Timmis et al. 2016 [80], Brannon 2017 [13], Brayne 2017 [14], Pak et al. 2017 [60], Taylor 2017 [79], Voigt 2017 [85]
3.3. New forms of discrimination	
3.3.1. Economic discrimination	Hildebrandt and Koops 2010 [35], Peppet 2014 [62], Turow et al. 2015 [81]
3.3.2. Health prediction discrimination	Hoffman 2010 [37], Cohen et al. 2014 [22], Ajunwa et al. 2016 [2], Hoffman 2017 [38]

Algorithmic causes of discrimination

Ten papers focused on how algorithmic and classificatory mechanisms might make data mining, classification and profiling discriminatory. These studies underlined that data mining technologies always involve a form of statistical discrimination. Adverse outcomes against protected classes might occur involuntarily due to the classification system. Barocas and Selbst [8] and d'Alessandro et al. [25], for example, pointed out that while the process of locating statistical relationships in a dataset is automatic, computer scientists still have to personally set both the target variable or outcome of interest (“what data miners are looking for”) and the “class labels” (“that divides all the possible outcomes of the target variable in binary and mutually exclusive categories”) [8]. Insofar the data scientist needs to translate a problem into formal computer coding, deciding on the target variable and the class labels is a subjective process. Another algorithmic cause of discrimination is related to biased data in the model. In order to develop automatization, data mining models need datasets to train on,

Table 4 Causes of discrimination in data analytics

Causes of discrimination	Related articles
1. Algorithmic causes	
1.1. Definition of the target variable	Barocas and Selbst 2016 [8], d'Alessandro et al. 2017 [25]
1.2. <i>Data issues</i> Training data (Historically biased data sets)	Kamiran and Calders 2012 [42], Barocas and Selbst 2016 [8], Brayne 2017 [14], d'Alessandro et al. 2017 [25]
1.3. <i>Data issues</i> Training data (manual assignment of class labels)	Barocas and Selbst 2016 [8], d'Alessandro et al. 2017 [25]
1.4. <i>Data issues</i> Data collection (Overrepresentation and underrepresentation)	Barocas and Selbst 2016 [8], d'Alessandro et al. 2017 [25]
1.5. Proxies	Schermer 2011 [73], Kamiran and Calders 2012 [42], Barocas and Selbst 2016 [8], Zliobaite and Custers 2016 [95], d'Alessandro et al. 2017 [25]
1.6. Feedback loop	Mantelero 2016 [54], Brayne 2017 [14], d'Alessandro et al. 2017 [25]
1.7. Overfitting	Kamiran and Calders 2012 [42], Mantelero 2016 [54]
1.8. Feature selection	Barocas and Selbst 2016 [8]
1.9. <i>Cost function</i> Error by omission	d'Alessandro et al. 2017 [25]
1.10 <i>Masking</i> Proxies	Peppet 2014 [61], Zarsky 2014 [93], Barocas and Selbst 2016 [8], Zliobaite and Custers 2016 [95], Kroll et al. 2017 [45]
2. Digital divide	
2.1. Skills	Boyd and Crawford 2012 [12], Casanas i Comabella and Wanat 2015[18]
2.2. Resources	Barocas and Selbst 2016 [8], Pak et al. 2017 [60]
2.3. Geographical location	Casanas i Comabella and Wanat 2015 [18], Barocas and Selbst 2016 [8], Pak et al. 2017 [60]
2.4. Age	Casanas i Comabella and Wanat 2015 [18]
2.5. Income	Barocas and Selbst 2016 [8], Pak et al. 2017 [60]
2.6 Gender	Boyd and Crawford 2012 [12]
2.7. Education	Boyd and Crawford 2012 [12]
2.8 Race	Bakken and Reame 2016 [6], Sharon 2016 [74]
3. Data linkage	Susewind 2015 [76], Cato et al. 2016 [19], Zarate et al. 2016 [91], Ploug and Holm 2017 [64]

since they learn to make classifications on the basis of given examples. Schermer [73] argued that if the training data is contaminated with discriminatory or prejudiced cases, the system will assume them as valid examples to learn from and reproduce discrimination in its own outcomes. This contamination could derive from historically biased datasets [14] or from the manual assignment of class labels by data miners [8]. An additional issue with the training data might be the data collection bias [8] or sample bias [25]. Bias in the data collection can present itself as an *underrepresentation* of specific groups and/or protected classes in the data set, which might result in unfair or unequal treatment, or also an *overrepresentation* in the data set which might result in a “disproportioned attention to a protected class group, and the increased scrutiny may lead to a higher probability of observing a target transgression” [25]. Within this context, Kroll and colleagues mentioned the phenomenon of “overfitting” where “models may become too specialized or specific to the data used for training” and, instead of finding the best possible decision rule overall, they simply

learn the most suited rule to the training data thus perpetrating its bias [45]. Another possible algorithmic cause of discriminatory outcomes is proxies for protected characteristics such as race and gender. A historically recognized proxy for race, for example, is ZIP or post-code and “redlining” is defined as the systematic disadvantaging of specific, often racially associated, neighborhoods or communities [73]. On this note, Zliobaite and Custers [95] highlighted how, in data mining, the elimination of sensitive attributes from the data set does not help to avoid discriminative outcomes as the algorithm could automatically identify unpredictable proxies for protected attributes. Two papers discussed feedback loop and systematic loop as a possible cause of unfair predictions [14, 25]. These involve the creation of a negative vicious cycle where certain inputs in the data set induce statistical deviations that are learned and perpetuated by the algorithm in a self-fulfilling loop of cause and consequence. An example might help to clarify this mechanism: police crime notification in certain urban areas will increase police patrol activity since crime notification is considered predictive of increased criminal activity. However, intensive patrolling will result in an increasingly higher rate of criminal activity reports in that area, irrespective of the true crime rate of that neighborhood with respect to others. “Feature selection” is another possible cause of discrimination identified by Barocas and Selbst [8]. This is a process that is used by those who collect and analyze the data to decide what kind of attributes or features they want to observe and take into account in their decision making processes. The authors argued that the selection of attributes always involves a reductive representation of the more complex real world object, person, or phenomena that it aims to portray insofar as it cannot take into account all the attributes and all the social or environmental factors related to that individual [8].

d’Alessandro identified two additional possible causes of discrimination lined to model misspecification, that is “the functional form of feature set of a model under study not being reflective of the true model” [25]. These are “cost function” misspecification and “error by omission”. “Cost function” misspecification is defined as the failure to consider the additional weight given to the event or attribute of interest (e.g. criminal record) by the data scientist. d’Alessandro argued that since “discrimination is enforced when a protected class receives an unwarranted negative action”, if a “false positive error could cause significant harm to an individual in a protected class”, the weight of the attribute, namely its asymmetry with respect to others, has to be taken into account [25]. “Error by omission” is another form of cost function misspecification that occurs when terms that penalize discrimination are ignored or left out from the model. Simply put, it means that the model does not take into account the differences in how the algorithm classifies protected and non-protected classes [25].

Finally, the reviewed articles also highlighted how algorithmic analysis can become an excellent and innovative tool for direct voluntary discrimination. This practice, defined as “masking”, involves the intentional exploitation of the mechanisms described above to perpetrate discrimination and unfairness. The most common practice of masking is the intentional use of proxies as indicators of sensitive characteristics [8, 45, 62, 93, 95].

Digital divide

We identified nine papers that discussed the digital divide, that is, the gap between those who have continuous and ready access to internet, computer and smartphones and those who do not, as a possible cause of inequality, injustice or discrimination. Lack of resources or computational skills, older age, geographical location, and low income were identified as.

possible causes of this digital divide [8, 18, 60]. Two papers [49, 74] discussed the “big data exclusions” referring to those individuals “whose information is not regularly collected or analyzed because they do not routinely engage in data-generating practices” [49]. On the same note, Bakken and Reame [6] argued that data is mainly gathered from white, educated people leaving out racial minorities such as Latinos. Boyd and Crawford discussed the creation of new digital divides, arguing that discrimination may arise due to (1) differences in information access and processing skills—the Big Data rich and the Big Data poor, and due to (2) gender differences insofar most researchers with computational skills are men [12]. Lastly, Cohen et al. [22] described how the commercialization of predictive models will leave out vulnerable categories such people with disabilities or limited decision-making capacities and high risk patients.

Data linkage and aggregation

Four papers discussed data linkage, that is, the possibility of automatically obtaining, linking, and disclosing personal and sensitive information as an important cause of discrimination. Two articles [19, 91] described how the use of electronic health records could result in the automatic disclosure of sensitive data without the patient’s explicit agreement or to re-identification. Others [64, 74] also highlighted that discrimination is not created by a data collection system (such as social and health registries) in itself, but is made easier by the linkage and aggregation potentiality embedded in the data.

Suggested solutions

The literature has suggested several different strategies to prevent discrimination and inequality in data analytics, ranging from computer based and algorithmic solutions to the incorporation of human involvement and supervision (see Table 5).

Practical computer science and technological solutions

Some articles authored by IT specialists suggested practical computer science solutions, namely the development of discrimination-aware methods to be applied during the development of the algorithmic models. These techniques include: *pre-processing* methods that involve the sanitization or distortion of the training data set to remove possible bias in order to prevent the new model from learning discriminatory behaviors (e.g. [33, 43]; *in-processing* techniques that provide for the modification of the learning algorithm through the application of regularization to probabilistic discriminative models [43]) such as the inclusion of sensitive attributes to avoid discriminatory predictions [66, 95] or the addition of randomness to avoid overfitting or hidden model bias [45]; *post-processing* methods that involve the auditing of the extracted data mining models for discriminative patterns and eventually their sanitization [34]. Along these lines, [25] suggested the implementation of an overall discrimination-aware auditing process that

involves the coherent combination of all pre-, in-, and post-processing methods to avoid discrimination. Many papers indicated how the implementation of transparency of data mining processes could help avoid injustice and harm. Practical suggestions to reinforce transparency in data mining include the development of interpretable algorithms that will give explanations on the logical steps behind a certain classification [45, 73], and the creation of transparent models that will allow individuals to see in advance how their behavior and choices will be interpreted by the algorithm or the infrastructure [21, 35]. Another solution was the enhancement of proper privacy preserving strategies since it's impossible to eradicate the likelihood of discriminative practices in data mining if discrimination-preventing data mining is not integrated with privacy-preserving data mining models [34]. Lastly, one paper suggested the promotion of exploratory fairness analysis that could be used to build up knowledge of the mechanisms and logics behind machine learning decisions [84].

Legal solutions

Implementation of legislation on data protection and discrimination was another common suggestion among the papers from the USA. Kuempel [46] suggested that the harmonization of stronger data protection legislation across different sectors in the US, could help contrast discrimination in under regulated areas, such as online marketing and data brokering. One author [62] argued that policies to constrain data use should be put into place. Such constraints should limit or deny the disclosure of sensitive data in specific contexts (e.g. health data in employment) or even deny specific uses of data in contexts where sensitive data is already disclosed if such use might cause harm to the individual (e.g. the use of health data to increase premiums in insurance). Finally, one article [35] suggested the idea of “code as law”, that is a transition from written-law to computational law, implying the articulation of specific legal norms in digital technologies through the use of software.

Human-centered solutions

Keeping the human in the loop of data mining was another recommendation. According to some papers, human oversight and supervision is critical to improve fairness since humans could notice where important factors are unexpectedly overlooked or sensitive attributes are improperly correlated [11, 25]. Other solutions that include human involvement were: (a) the participation of trusted third parties to either store sensitive data and rule on their disclosure to companies [84] or supervise and assess suspicious data mining and classification practices [54]; (b) the engagement of all relevant stakeholders involved in a decision making or profiling process—such as health care institutions, physicians, researchers, subjects of research, insurance companies, and data scientists—in a multidisciplinary discussion towards the creation of a theoretical overarching framework to regulate data mining and promote the implementation of fair algorithms [22]; (c) the implementation of strategies to educate data scientists in building proper models, such as the creation of a knowledge base platform for fairness in data mining that could be investigated by data scientists in case they stumbled upon problematic correlations; and (d) the implementation of flexibility and discretion in EHR

Table 5 Suggested solutions to discrimination in Big Data

Suggested solutions	Paper references
1. Computer science and technical solutions	
1.1. Pre-processing	Kamiran and Calders 2012 [42], Hajian and Domingo-Ferrer 2013 [33], Kamiran et al. 2013 [43], Hajian et al. 2014 [32]
1.2. In-processing	Calders and Verwer 2010 [17], Pope and Sydnor 2011 [66], Kamiran et al. 2013 [43], Zliobaite and Custers 2016 [95], Kroll et al. 2017 [45]
1.3. Post-processing	Hajian et al. 2015 [34]
1.4. Mixed methods	d'Alessandro et al. 2017 [25]
1.5. Implementation of transparency	Hildebrandt and Koops 2010 [35], Schermer 2011 [73], Citron and Pasquale 2014 [21], Kroll et al. 2017 [45]
1.6. Privacy preserving strategies	Hildebrandt and Koops 2010 [35], Hajian et al. 2015 [34]
1.7. Exploratory fairness analysis	Veale and Binns 2017 [84]
2. Legal solutions	Hildebrandt and Koops 2010 [35], Hoffman 2010 [37], Citron and Pasquale 2014 [21], Peppet 2014 [62], Hirsch 2015 [36], Kuempel 2016 [46], Hoffman 2017 [38]
3. Human based solutions	
3.1. Human in the loop	Zarsky 2014 [93], Berendt and Preibusch 2017 [11], d'Alessandro et al. 2017 [25]
3.2. Third parties	Mantelero 2016 [54], Veale and Binns 2017 [84]
3.3. Multidisciplinary involvement	Cohen et al. 2014 [22], Taylor 2016 [77, 78], Taylor 2017 [79]
3.4. Education	Zarsky 2014 [93], Veale and Binns 2017 [84]
3.5. Implementing EHR flexibility	Hoffman 2010 [37]

disclosing system to avoid stigma from the disclosure of personal and private information [37].

Obstacles to fair data mining

Many papers described algorithmic decision making as a black box system where the input and the output of the algorithm are visible but the inner process remains unknown [13, 21, 25], resulting in lack of transparency regarding the methods and the logic behind scoring and predictive systems [35, 48, 54, 92]. Reasons behind

the opacity of automated decision making are multiple: first, algorithms might use enormous and very complex data sets that are uninterpretable to regulators [25], who frequently lack the required computer science knowledge to understand algorithmic processes [73]; second, automatic decision making might intrinsically transcend human comprehension since algorithms do not make use of theories or contexts as in regular human based decision-making [58]; and finally, algorithmic processes of firms or companies might be subject to intellectual property rights or covered by trade secret provisions [35]. If there is no transparent information on how algorithms and processes work it is almost impossible to [44] evaluate the fairness of the algorithms or discover discriminatory patterns in the system [45].

Human bias was identified as another main obstacle to fair data mining. Human subjectivity is at the very core of the design of data mining algorithms since the decisions regarding which attributes will be taken into account and which will be ignored are subject to human interpretation [12], and will inevitably reflect the implicit or explicit values of their designers [1].

Algorithmic data mining also poses considerable conceptual challenges. Many papers claimed that automatic decision making and profiling are reshaping the concept of discrimination, beyond legally accepted definitions. In the United States (US), for example, Barocas and Selbst [8] claimed that algorithmic bias and automatization are blurring notions of motive, intention and knowledge, making it difficult for the US doctrine on disparate impact and disparate treatment to be used to evaluate and persecute causes of algorithmic discrimination. One article [48], discussing European Union (EU) regulation, argued that it is necessary to rethink discrimination in the context of data driven profiling, since the production of arbitrary categories in data mining technologies and the automatic correlation of the individual's attributes by the algorithm differ from traditional profiling, which is based on the establishment of a causal chain developed by human logic. Some articles have also pointed out that concepts like "identity" and "group" are being transformed by data mining technologies. de Vries argued that individual identity is increasingly shaped by profiling algorithms and ambient intelligence in terms of increased grouping created in accordance with algorithms' arbitrary correlations, which sort individuals into a virtual, probabilistic "community" or "crowd" [27]. This typology of "group" or "crowd" differs from the traditional understanding of groups, since the people involved in the "group" might not be aware of (1) their membership to that group, (2) the reasons behind their association with that group and, most importantly, (3) the consequences of being part of that group [54]. Two other concepts are being reshaped by data technologies. The first is the concept of border [1], which is no longer a physical and static divider between countries but has become a pervasive and invisible entity embedded in bureaucratic processes and the administration of the state due to Big Data surveillance tools such as electronic passports and airport security measures. The second is the concept of disability, which needs to be broadened to include all diseases and health conditions, such as obesity, high blood pressure and minor cardiac conditions, which might result in discriminatory outcomes from automatic classifiers through algorithmic correlation with more serious diseases [37, 38].

The final barrier that was pinpointed in the literature is of a legal nature. According to some authors, current antidiscrimination and data protection legislation, both in the EU and in the US, are not well equipped to address cases of discrimination stemming from digital technologies [8]. Kroll et al. [45] claimed that current antidiscrimination laws might legally prevent users of algorithms from revising to inspecting algorithms after the discriminatory fact has happened, making the development of ex-ante anti-discriminatory models even more pressing. Kuempel [46] argued that data protection legislation is too sectorial and does not provide sufficient safeguards from discrimination in sectors like marketing. Some papers focused on the implications of the implementation of European data protection regulations, specifically the new General Data Protection Regulation (GDPR) of May 2018. The authors emphasized that data protection requirements, such as data gathering minimization and the limitation of use of personal data, might result in barriers into the development of antidiscrimination models that demand the inclusion of sensitive data in order to avoid discriminatory outcomes [35, 95] (see Table 6).

Beneficial adoption of Big Data technologies

Finally, many papers also described how data mining technologies could be an important practical tool to counteract or prevent inequality and discrimination (see Table 7).

Data mining is said to promote objectivity in classification and profiling because decisions are made by a formal, objective and constant algorithmic process with a more reliable empirical foundation than human decision-making [8]. This feature of objectivity could limit human error and bias. According to some of the literature, automatic data mining could also be used to discover and assess discriminatory practices in classification and data mining. Through the construction of discrimination-aware algorithmic models (e.g. [10, 71]), individuals who suspect that they are being discriminated against could be helped to identify and assess direct/indirect discrimination, favoritism or affirmative action, and decision makers (such as employers, insurance companies managers and so on) could be protected against wrongful discrimination allegations. Some of the papers also highlighted that the potential of Big Data technologies to integrate socioeconomic data, mobile data and geographical data could promote equitable and beneficial implementations in various sectors. In healthcare, for example, the integration of healthcare data with spatial contextual information might help identifying areas and groups that require health promotion [47]; moreover the use of Big Data, profiling and classification could foster equity with regard to health disparities in research, since it could promote the implementation of tailored strategies that take into account an individual's ethnicity, living conditions and general lifestyle [6]. Economic and urban development is another area in which data mining could help foster equity. The integration of analysis from mobile phone activity and socio-economic factors within geographical data could help monitoring and assessment of social structural inequalities to promote the implementation of more equitable city development and growth [55, 83, 85]. Migration could also

benefit from the use of Big Data technologies, as it can provide scholars and activists with more accurate data regarding migration flows and thus prepare and enhance humanitarian processes [1]. Finally, two papers also discussed the positive influence of social media [59] analyzed how text mining could be used to assess the level and diffusion of discrimination related to people affected by Human Immunodeficiency Virus Infection (HIV) and Acquired Immune Deficiency Syndrome (AIDS) in popular social media like Facebook and at the same time implement awareness-raising campaigns to spread tolerance. Another article [18] claims that social media could be used to enhance the participation of people receiving pediatric palliative care, a particularly vulnerable group, in research.

Discussion

The majority of the reviewed papers (49 out of 61) date from the last 5 years. This shows that although Big Data has been a trending buzzword in the scientific literature since 2011 [16], the problem of algorithmic discrimination has become of prime interest only recently, in conjunction with the publication of the White House report of 2014 [65]. Hence, scholarly reflection on this issue has appeared rather late, leaving potentially discriminatory outcomes of data mining unaddressed for a long time. Moreover, in line with other studies [56], our review indicates that while a theoretical discussion on this

Table 6 Barriers to fair data analytics

Obstacles to fair data analytics	Paper references
1. Black box	Hildebrandt and Koops 2010 [35], Ruggieri et al. 2010 [71], Schermer 2011 [73], Berendt and Preibusch 2014 [10], Citron and Pasquale 2014 [21], Cohen et al. 2014 [22], Leese 2014 [48], Zarsky 2014 [93], Kennedy and Moss 2015 [44], Newell and Marabelli 2015 [58], Turow, McGuigan et al. 2015 [81], Mantelero 2016 [54], Zarsky 2016 [92], Brannon 2017 [13], Brayne 2017 [14], d'Alessandro et al. 2017 [25], Kroll et al. 2017 [45], Taylor 2017 [79]
2. Human bias	Boyd and Crawford 2012 [12], Kamiran and Calders 2012 [42], Citron and Pasquale 2014 [21], Zarsky 2014 [93], Ajana 2015 [1], Ajunwa et al. 2016 [2], Barocas and Selbst 2016 [8], Berendt and Preibusch 2017 [11], Brayne 2017 [14], d'Alessandro et al. 2017 [25], Veale and Binns 2017 [84], Voigt 2017 [85]
3. Conceptual challenges	de Vries 2010 [27], Hoffman 2010 [37], Lerman 2013 [49], Leese 2014 [48], Zarsky 2014 [93], Ajana 2015 [1], Hirsch 2015 [36], MacDonnell 2015 [53], Barocas and Selbst 2016 [8], Kuempel 2016 [46], Mantelero 2016 [54], Francis and Francis 2017 [30], Hoffman 2017 [38], Kroll et al. 2017 [45], Taylor 2017 [79]
4. Inadequate legislation	Hildebrandt and Koops 2010 [35], Hoffman 2010 [37], Ruggieri et al. 2010 [71], Lerman 2013 [49], Citron and Pasquale 2014 [21], Peppet 2014 [62], Barocas and Selbst 2016 [8], Kuempel 2016 [46], Zliobaite and Custers 2016 [95], Hoffman 2017 [38], Zliobaite 2017 [94]

topic is finally emerging, empirical studies on discrimination in data mining, both in the field of law and social sciences, are largely lacking. This is highly problematic especially in light of the new forms of disparate treatment that arise with the increased “datafication” of society. Price and health prediction discrimination (e.g. in insurance policies), for example, are not illegal but might become ethically problematic if persons are denied access to essential goods or services based on their income or lifestyle. More evidence-based studies on the possible harmful use of these practices are urgently needed if we want to understand the complexity of this problem in depth. In addition, it is interesting to notice that no paper examined discrimination in relation to the four V’s of Big Data, as they focused more on the classificatory and algorithmic issues of data analytics. It is thus important that future studies also take into account the issue of harmful discrimination related to the specific problems related to the unique characteristic of Big Data, such as the veracity of the data sets and the constraints related to the high volume of data, and the velocity of their production.

Although the majority of papers were theoretical in nature, the term discrimination was presented as self-explanatory and linked to other notions such as injustice, inequality and unequal treatment, with the exception of some papers in law and computer science. This overall lack of a working definition in the literature is highly problematic, for several reasons.

First given that data mining technologies are purposely created to classify, discern, divide and separate individuals, groups or actions [8], discussing the problem of unfair discrimination in absence of a clear definition is creating confusion. The discrimination operated in data-mining, in fact, is not in itself illegal or ethically wrong as long as it limits itself to making a distinction between people with different characteristics [35]. For example distinguishing between minors and adults is a socially and legally accepted practice of “neutral discrimination”; based on a straightforward distinction of age (in most countries set at 18 years old) individuals are dissimilarly treated: adults have

Table 7 Beneficial adoption of data analytics

Beneficial adoption of Big Data	Paper references
1. Promotion of objectivity in classification	Zarsky 2014 [93], MacDonnell 2015 [53], Barocas and Selbst 2016 [8], Brayne 2017 [14]
2. Uncover and assess discriminatory practices	Ruggieri et al. 2010 [71], Romei and Ruggieri et al. 2013 [69], Berendt and Preibusch 2014 [10]
3. Integration of data for promotion of equality and social integration	
3.1. Healthcare	Le Meur et al. 2015 [47], Bakken and Reame 2016 [6]
3.2. Economic growth and urban development	Mao et al. 2015 [54], Vaz et al. 2017 [83], Voigt 2017 [85]
3.3. Migration	Ajana 2015 [1], Taylor 2016 [77, 78]
4. Beneficial use of social media	Casanas i Comabella and Wanat 2015 [18], Nielsen et al. 2017 [59]

different rights and duties than minors, they can drive and vote, they are judged differently in a court of law and so on. Moreover, even efforts to achieve social equality sometimes imply a sort of differential treatment; for example in the case of gender equality, divergent treatment of individuals based on gender is allowed if such treatment is adopted with the long term goal of evening out social disparities [87]. Hence, if researchers want to discuss the problem of discrimination in data-mining, a distinction between harmful and unfair versus neutral or fair discrimination is of utmost importance.

Second, without an adequate definition of discrimination, it is difficult for computer scientists and programmers to appropriately implement algorithms. In fact, to avoid unfair practices, measure fairness and quantify illegal discrimination [43], they need to translate the notion of discrimination into a formal statistical set of operations. The need for this expert knowledge may explain why, compared to other researchers in the field, computer scientists have been at the forefront of the search for a viable definition.

Still, despite the need for a working definition of discrimination, we should not forget that it remains an elusive ethical and social notion which cannot and should not be reduced to a “petrified” statistical measurement. As seen in our review, data-mining has given rise to novel forms of differential treatment. To properly understand the implications of these new discriminatory practises, a reconceptualization of the notion of fair and unfair discrimination might be needed. To keep the debate on discrimination in Big Data open it is important to keep humans in the loop.

Practices of automatic profiling, sorting and decision making through data mining have been introduced with the *prima facie* concept that Big Data technologies are objective tools capable of overcoming human subjectivity and error resulting in increased fairness [3]. However, data mining can never be fully human-free, not only because humans always risk undermining the presumed fairness and objectivity of the process with subconscious bias, personal values or inattentiveness, but also because they are crucial in order to avoid improper correlations and thus to ensure fairness in data mining. It thus seems that Big Data technologies are deeply tied to this dichotomous dimension where humans are both the cause of its flaws and the overseers of its proper functioning.

One way of keeping the human in the loop is through legislation. Our results, however, show that although legal scholars have tried to address possible unfair discriminatory outcomes of new forms of profiling, Big Data poses important challenges to “traditional” antidiscrimination and privacy protection legislation because core

notions, such as motive and intention, are no longer in place [8]. A recurring theme in many papers was that legislation always lags behind technological developments and that while gaps in legal protection are somehow systemic [35], an overarching legal solution to all unfair discriminatory outcomes of data mining is not feasible [45].

In our review, very few papers offered a pragmatic legal solution to the problem of unfair discrimination in data-mining: for example one study advocated for a generally applicable rule [46], while another suggested the production of a set of precedents built in time through a case by case adjudication [36]. Both solutions are incompatible with the reality and needs of data management because they are either too rigid [46] or too specialized and protracted [36].

This poor outcome is probably the result of the technically complex nature of data mining and the intrinsically tricky legal designation of what represents unfair discrimination that should be prohibited by law. The new European *General Data Protection Regulation* (GDPR) is exemplary in this regard. Two key features of the GDPR are: data minimization (i.e. data collection and processing should be kept to a minimum) and purpose limitation (i.e. data should be analysed and processed only for the purpose it was collected for). Since both these principles are inspired from data privacy regulations established in the 1970s, they fail to take into account two crucial points that have been reiterated by many computer science, technical and legal scholars in the past few years [31]: first, with Big Data technologies, information is not collected for a specific, limited and specified purpose, rather it is gathered to discover new and unpredictable patterns and correlations [53]; second, antidiscrimination models require the inclusion of sensitive data in order to detect and avoid discriminatory outcomes [95].

The difficulties encountered in adequately regulating discrimination in Big Data, especially from a legal point of view, could be partly related to a diffuse lack of dialogue among disciplines. The reviewed literature in fact pinpointed that while on the one hand, unfair discrimination is a complex philosophical and legal concept that stores difficulties for trained data scientists [20], Big Data, on the other, is quite a technological field so philosophers, social scientists and lawyers do not always fully understand the implications of algorithmic modelling for discrimination [73].

This mutual lack of understanding highlights the urgent need for a multidisciplinary collaboration between fields, such as philosophy, social science, law, computer science and engineering. The idea of collaboration between disciplines due to the spreading of digital technologies is not new. An example of this can be found in the conception of “code as law” first proposed by both Reidenberg and Lessing in the late 1990s, which implies the design of digital technologies to support specific norms and laws such as privacy and antidiscrimination [50, 68]. As shown by our results (e.g. [25, 42, 43]), the “code as law” proposal has been steadily implemented in computer science practice by many scholars who want to implement antidiscrimination rules in algorithmic models to avoid unfair harmful outcomes. Some papers, however, recommended a broader and overarching dialogue among disciplines [22, 31, 45]. Nonetheless, concrete means to put this multidisciplinaryity into practice were lacking in the literature.

Finally, a few studies highlighted that Big Data technologies may tackle discrimination and promote equality in various sectors, such as healthcare and urban development [6, 18, 47]. Such interventions, however, might have the opposite effect and create other types of social disparities by widening the divide between people who have access to digital resources and those who do not, on the basis of income, ethnicity, age, skills, and geographical location. The significant number of papers that identified the digital divide as a major cause of inequality indicates how, despite all the efforts made to enhance digital participation across the globe [89, 90], social disparities due to lack of access to digital technologies are increasing in many sectors including health [88], public participation/engagement [9] and public infrastructure development [60, 79]. Scholars are rather sceptical about finding a solution to this problem due to the ever-changing technological landscape that creates new inclusion difficulties [89, 90]. Still, due to the potential promising beneficial applications of Big Data technologies, more studies should focus on the analysis and implementation of such fair uses of data-mining while considering and avoiding the creation of new divides.

In conclusion, more research is needed on the conceptual challenges that Big Data technologies raise in the context of data mining and discrimination. The lack of adequate terminology regarding digital discrimination and the possible presence of latent bias might mask persistent forms of disparate treatment as normalized practices. Although a few papers tackled the subject of a possible conceptual revision of discrimination and fairness [79], no study has done so in an exhaustive way.

Limitations

A total of 61 peer-reviewed articles in English qualified for inclusion and were further assessed. It might thus be possible that studies in other languages and relevant grey literature have been overlooked. Aside from these limitations, this is the first study to comprehensively explore the relation between Big Data and discrimination from a multi-disciplinary perspective.

Conclusions

Big Data offers great promise but also poses considerable risks. The literature review highlights that unfair discrimination is one of the most pressing, but at the same time an often underestimated issue in data mining. A wide range of papers proposed solutions on how to avoid discrimination in the use of data technologies. Though most of the suggested strategies were practical computational/algorithmic methods, numerous papers recommended human solutions. Transparency was a commonly suggested solution to enhance algorithmic fairness. Improving algorithmic transparency and resolving the black box issue might thus be the best course to undertake when dealing with discriminatory issues in data analytics. However, our study results identify a considerable number of barriers to the proposed strategies, such as technical difficulties, conceptual challenges, human bias and shortcomings of legislation, all of which hamper the implementation of such fair data mining practices. Due to the risk of discrimination in data mining and predictive analytics and the strikingly shortage of empirical studies on the topic that our review has brought to light, we argue that more empirical research is needed to assess how discriminatory practices are deliberately and accidentally

emerging from their increased use in numerous sectors such as healthcare, marketing and migration. Moreover, since most studies focused on the negative discriminatory consequences of Big Data, more research is needed on how data mining technologies, if properly implemented, could also be an effective tool to prevent unfair discrimination and promote equality. As more reports from the press are emerging on the positive use of data technologies to assist vulnerable groups, future research should focus on the diffusion of similar beneficial applications. However, since even such practices are creating new forms of disparity between those who can access digital technologies and those who do not, research should also focus more on the implementation of practical strategies to mitigate the Digital Divide.

Abbreviations

US: United States; EU: European Union; HIV: human immunodeficiency virus; AIDS: acquired immunodeficiency syndrome.

Authors' contributions

MF collected the data, performed the analysis and drafted the manuscript. EDC supported with data analysis, contributed in writing the manuscript and revised the initial versions of the manuscript. BE provided general guidance, proof-read the manuscript, suggested necessary amendments and helped in revising the paper. All authors read and approved the final manuscript.

Acknowledgements

We thank Dr. David Shaw for his valuable contribution of the project.

Competing interests

The authors declare that they have no competing interests.

Availability of data materials

The datasets used for the current study are available from the corresponding author on reasonable request.

Funding

The funding for this study was provided by the Swiss National Science Foundation in the framework of the National Research Program "Big Data", NRP 75 (Grant-No: 407540_167211).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 1 October 2018 Accepted: 22 January 2019

Published online: 05 February 2019

References

1. Ajana B. Augmented borders: Big Data and the ethics of immigration control. *J Inf Commun Ethics Soc.* 2015;13(1):58–78.
2. Ajunwa I, Crawford K, Ford JS. Health and Big Data: an ethical framework for health information collection by corporate wellness programs. *J Law Med Ethics.* 2016;44(3):474–80.
3. Anderson C. End of theory: the data deluge makes the scientific method. 2008. <https://www.wired.com/2008/06/pb-theory/> Accessed 2 Dec 2017.
4. Andrejevic M. Big Data, big questions| the Big Data divide. *Int J Commun.* 2014;8:17.
5. Anuradha J. A brief introduction on Big Data 5Vs characteristics and Hadoop technology. *Procedia Comput Sci.* 2015;48:319–24.
6. Bakken S, Reame N. The promise and potential perils of Big Data for advancing symptom management research in populations at risk for health disparities. *Annu Rev Nurs Res.* 2016;34:247–60.
7. Ball K, Di Domenico M, Nunan D. Big Data surveillance and the body-subject. *Body Soc.* 2016;22(2):58–81.
8. Barocas S, Selbst AD. Big Data's disparate impact. *California Law Rev.* 2016;104(3):671–732.
9. Bartikowski B, Laroche M, Jamal A, Yang Z. The type-of-internet-access digital divide and the well-being of ethnic minority and majority consumers: a multi-country investigation. *J Business Res.* 2018;82:373–80.
10. Berendt B, Preibusch S. Better decision support through exploratory discrimination-aware data mining: foundations and empirical evidence. *Artif Intell Law.* 2014;22(2):175–209.
11. Berendt B, Preibusch S. Toward accountable discrimination-aware data mining: the Importance of keeping the human in the loop—and under the looking glass. *Big Data.* 2017;5(2):135–52.
12. Boyd D, Crawford K. Critical questions for Big Data: provocations for a cultural, technological, and scholarly phenomenon. *Inf Commun Soc.* 2012;15(5):662–79.

13. Brannon MM. Datafied and Divided: techno-dimensions of inequality in American cities. *City Community*. 2017;16(1):20–4.
14. Brayne S. Big Data surveillance: the case of policing. *Am Sociol Rev*. 2017;82(5):977–1008.
15. Burgess M. UK police are using AI to inform custodial decisions—but it could be discriminating against the poor. 2018. <http://www.wired.co.uk/article/police-ai-uk-durham-hart-checkpoint-algorithm-edit>. Accessed 12 Apr 2018.
16. Burrows R, Savage M. After the crisis? Big Data and the methodological challenges of empirical sociology. *Big Data Soc*. 2014;1(1):2053951714540280.
17. Calders T, Verwer S. Three naive Bayes approaches for discrimination-free classification. *Data Min Knowl Disc*. 2010;21(2):277–92.
18. Casanas i Comabella C, Wanat M. Using social media in supportive and palliative care research. *BMJ Support Palliat Care*. 2015;5(2):138–45.
19. Cato KD, Bockting W, Larson E. Did I tell you that? Ethical issues related to using computational methods to discover non-disclosed patient characteristics. *J Empirical Res Hum Res Ethics*. 2016;11(3):214–9.
20. Chouldechova A. Fair prediction with disparate impact: a Study of bias in recidivism prediction instruments. *Big Data*. 2017;5(2):153–63.
21. Citron DK, Pasquale F. The scored society: due process for automated predictions. *Wash L Rev*. 2014;89:1.
22. Cohen IG, Amarasingham R, Shah A, Bin X, Lo B. The legal and ethical concerns that arise from using complex predictive analytics in health care. *Health Aff*. 2014;33(7):1139–47.
23. Courtland R. Bias detectives: the researchers striving to make algorithms fair. *Nature*. 2018;558(7710):357.
24. Crawford K. Think again: Big Data. *Foreign Policy*. 2013;9.
25. d'Alessandro B, O'Neil C, LaGatta T. Conscientious classification: a data scientist's guide to discrimination-aware classification. *Big Data*. 2017;5(2):120–34.
26. Daries JP, Reich J, Waldo J, Young EM, Whittinghill J, Ho AD, Seaton DT, Chuang I. Privacy, anonymity, and Big Data in the social sciences. *Commun ACM*. 2014;57(9):56–63.
27. de Vries K. Identity, profiling algorithms and a world of ambient intelligence. *Ethics Inf Technol*. 2010;12(1):71–85.
28. Floridi L. Big Data and their epistemological challenge. *Philos Technol*. 2012;25(4):435–7.
29. Francis JG, Francis LP. Privacy, confidentiality, and justice. *J Soc Philos*. 2014;45(3):408–31.
30. Francis LP, Francis JG. Data reuse and the problem of group identity. *Stud Law Polit Soc*. 2017;73:141–64.
31. Goodman BW. A step towards accountable algorithms? algorithmic discrimination and the european union general data protection. In: 29th conference on neural information processing systems (NIPS 2016), Barcelona, Spain. 2016.
32. Hajian S, Domingo-Ferrer J. A methodology for direct and indirect discrimination prevention in data mining. *IEEE Trans Knowl Data Eng*. 2013;25(7):1445–59.
33. Hajian S, Domingo-Ferrer J, Farras O. Generalization-based privacy preservation and discrimination prevention in data publishing and mining. *Data Min Knowl Disc*. 2014;28(5–6):1158–88.
34. Hajian S, Domingo-Ferrer J, Monreale A, Pedreschi D, Giannotti F. Discrimination-and privacy-aware patterns. *Data Min Knowl Disc*. 2015;29(6):1733–82.
35. Hildebrandt M, Koops B-J. The challenges of ambient law and legal protection in the profiling era. *Mod Law Rev*. 2010;73(3):428–60.
36. Hirsch DD. That's unfair! or is it? Big Data, Discrimination and the FTC's unfairness authority. *Ky Law J*. 2015;103:345–61.
37. Hoffman S. Employing e-health: the impact of electronic health records on the workplace. *Kan JL Pub Pol'y*. 2010;19:409.
38. Hoffman S. Big Data and the Americans with disabilities act. *Hastings Law J*. 2017;68(4):777–93.
39. Holtzhausen D. Datafication: threat or opportunity for communication in the public sphere? *J Commun Manag*. 2016;20(1):21–36.
40. Howie T. The Big Bang: how the Big Data explosion is changing the world. 2013.
41. Ioannidis JP. Informed consent, Big Data, and the oxymoron of research that is not research. *Am J Bioethics*. 2013;13(4):40–2.
42. Kamiran F, Calders T. Data preprocessing techniques for classification without discrimination. *Knowl Inf Syst*. 2012;33(1):1–33.
43. Kamiran F, Zliobaite I, Calders T. Quantifying explainable discrimination and removing illegal discrimination in automated decision making. *Knowl Inf Syst*. 2013;35(3):613–44.
44. Kennedy H, Moss G. Known or knowing publics? Social media data mining and the question of public agency. *Big Data Soc*. 2015. <https://doi.org/10.1177/2053951715611145>.
45. Kroll JA, Huey J, Barocas S, Felten EW, Reidenberg JR, Robinson DG, Yu HL. Accountable algorithms. *Univ Pa Law Rev*. 2017;165(3):633–705.
46. Kuempel A. The invisible middlemen: a critique and call for reform of the data broker industry. *Northwestern J Int Law Business*. 2016;36(1):207–34.
47. Le Meur N, Gao F, Bayat S. Mining care trajectories using health administrative information systems: the use of state sequence analysis to assess disparities in prenatal care consumption. *BMC Health Serv Res*. 2015;15:200.
48. Leese M. The new profiling: algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union. *Secur Dialogue*. 2014;45(5):494–511.
49. Lerman J. Big Data and its exclusions. *Stan L Rev Online*. 2013;66:55.
50. Lessing L. Code and other laws of cyberspace. New York: Basic Books; 1999.
51. Lupton D. Quantified sex: a critical analysis of sexual and reproductive self-tracking using apps. *Cult Health Sex*. 2015;17(4):440–53.
52. Lyon D. Surveillance, snowden, and big data: capacities, consequences, critique. *Big Data Soc* 2014;1(2): 2053951714541861.
53. MacDonnell P. The European Union's proposed equality and data protection rules: an existential problem for insurers? *Econ Aff*. 2015;35(2):225–39.

54. Mantelero A. Personal data for decisional purposes in the age of analytics: from an individual to a collective dimension of data protection. *Comput Law Secur Rev.* 2016;32(2):238–55.
55. Mao HN, Shuai X, Ahn YY, Bollen J. Quantifying socio-economic indicators in developing countries from mobile phone communication data: applications to Cote d'Ivoire. *EPJ Data Sci.* 2015. <https://doi.org/10.1140/epjds/s13688-015-0053-1>.
56. Mittelstadt BD, Floridi L. The ethics of Big Data: current and foreseeable issues in biomedical contexts. *Sci Eng Ethics.* 2016;22(2):303–41.
57. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev.* 2015;4(1):1.
58. Newell S, Marabelli M. Strategic opportunities (and challenges) of algorithmic decision-making: a call for action on the long-term societal effects of 'datification'. *J Strategic Inf Syst.* 2015;24(1):3–14.
59. Nielsen RC, Luengo-Oroz M, Mello MB, Paz J, Pantin C, Erkkola T. Social media monitoring of discrimination and HIV testing in Brazil, 2014–2015. *AIDS Behav.* 2017;21(Suppl 1):114–20.
60. Pak B, Chua A, Vande Moere A. FixMyStreet Brussels: socio-demographic inequality in crowdsourced civic participation. *J Urban Technol.* 2017;24(2):65–87.
61. Parliament E. Charter of fundamental rights of the European Union, Office for Official Publications of the European Communities. 2000.
62. Peppet SR. Regulating the internet of things: first steps toward managing discrimination, privacy, security and consent. *Tex L Rev.* 2014;93:85.
63. Perry JS. (2017). What is Big Data? More than volume, velocity and variety. <https://developer.ibm.com/dwblog/2017/what-is-big-data-insight/>. Accessed 21 Jan 2018.
64. Ploug T, Holm H. Informed consent and registry-based research—the case of the Danish circumcision registry. *BMC Med Ethics.* 2017. <https://doi.org/10.1186/s12910-017-0212-y>.
65. Podesta J. Big Data: Seizing opportunities, preserving values. Washington D. C.: White House, Executive Office of the President; 2014.
66. Pope DG, Sydnor JR. Implementing anti-discrimination policies in statistical profiling models. *Am Econ J Econ Pol.* 2011;3(3):206–31.
67. Reich J. Street bumps, Big Data, and educational inequality. 2013. http://blogs.edweek.org/edweek/edtechresearcher/2013/03/street_bumps_big_data_and_educational_inequality.html. Accessed 4 Mar 2018.
68. Reidenberg JR. Lex informatica: the formulation of information policy rules through technology. *Tex L Rev.* 1997;76:553.
69. Romei A, Ruggieri S. Discrimination data analysis: a multi-disciplinary bibliography. *Discrimination and privacy in the information society.* Berlin: Springer; 2013. p. 109–35.
70. Romei A, Ruggieri S, Turini F. Discrimination discovery in scientific project evaluation: a case study. *Expert Syst Appl.* 2013;40(15):6064–79.
71. Ruggieri S, Pedreschi D, Turini F. Integrating induction and deduction for finding evidence of discrimination. *Artif Intell Law.* 2010;18(1):1–43.
72. SAS-Institute. Big Data. What it is and why it matters.
73. Schermer BW. The limits of privacy in automated profiling and data mining. *Comput Law Secur Rev.* 2011;27(1):45–52.
74. Sharon T. The Googlization of health research: from disruptive innovation to disruptive ethics. *Personal Med.* 2016;13(6):563–74.
75. Shin PS. The substantive principle of equal treatment. *Leg Theory.* 2009;15(2):149–72.
76. Susewind R. What's in a name? Probabilistic inference of religious community from South Asian names. *Field Methods.* 2015;27(4):319–32.
77. Taylor L. The ethics of Big Data as a public good: which public? Whose good? *Philos Trans A Math Phys Eng Sci.* 2016. <https://doi.org/10.1098/rsta.2016.0126>.
78. Taylor L. No place to hide? The ethics and analytics of tracking mobility using mobile phone data. *Environ Plann D-Soc Space.* 2016;34(2):319–36.
79. Taylor L. What is data justice? The case for connecting digital rights and freedoms globally. *Big Data Soc.* 2017. <https://doi.org/10.1177/2053951717736335>.
80. Timmis S, Broadfoot P, Sutherland R, Oldfield A. Rethinking assessment in a digital age: opportunities, challenges and risks. *Br Edu Res J.* 2016;42(3):454–76.
81. Turow J, McGuigan L, Maris ER. Making data mining a natural part of life: physical retailing, customer surveillance and the 21st century social imaginary. *Eur J Cult Stud.* 2015;18(4–5):464–78.
82. Vandenhoe W. Non-discrimination and equality in the view of the UN human rights treaty bodies. *Intersentia nv.* 2005.
83. Vaz E, Anthony A, McHenry M. The geography of environmental injustice. *Habitat Int.* 2017;59:118–25.
84. Veale M, Binns R. Fairer machine learning in the real world: mitigating discrimination without collecting sensitive data. *Big Data Soc.* 2017. <https://doi.org/10.1177/2053951717743530>.
85. Voigt K. Social justice, equality and primary care: (How) Can 'Big Data' Help? *Philos Technol.* 2017. <https://doi.org/10.1007/s13347-017-0270-6>.
86. Ward JS, Barker A. Undefined by data: a survey of Big Data definitions. 2013. arXiv preprint [arXiv:1309.5821](https://arxiv.org/abs/1309.5821).
87. Weisbard PH. ABC of women workers' rights and gender equality. *Feminist Collections.* 2001;22(3–4):44.
88. Weiss D, Rydland HT, Øversveen E, Jensen MR, Solhaug S, Krokstad S. Innovative technologies and social inequalities in health: a scoping review of the literature. *PLoS ONE.* 2018;13(4):e0195447.
89. Yu B, Ndumu A, Mon L, Fan Z. An upward spiral model: bridging and deepening digital divide. In: *International conference on information.* Berlin: Springer; 2018.
90. Yu B, Ndumu A, Mon LM, Fan Z. E-inclusion or digital divide: an integrated model of digital inequality. *J Documenta-tion.* 2018;74(3):552–74.

91. Zarate OA, Brody JG, Brown P, Ramirez-Andreotta MD, Perovich L, Matz J. Balancing benefits and risks of immortal data. *Hastings Cent Rep*. 2016;46(1):36–45.
92. Zarsky T. The trouble with algorithmic decisions: an analytic road map to examine efficiency and fairness in automated and opaque decision making. *Sci Technol Hum Values*. 2016;41(1):118–32.
93. Zarsky TZ. Understanding discrimination in the scored society. *Wash L Rev*. 2014;89:1375.
94. Zliobaite I. Measuring discrimination in algorithmic decision making. *Data Min Knowl Disc*. 2017;31(4):1060–89.
95. Zliobaite I, Custers B. Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models. *Artif Intell Law*. 2016;24(2):183–201.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
