Review of European Studies; Vol. 11, No. 1; 2019
ISSN 1918-7173 E-ISSN 1918-7181
Published by Canadian Center of Science and Education

The Long Road Toward Tracking the Trackers and De-biasing: A Consensus on Shaking the Black Box and Freeing From Bias

George Bouchagiar

Correspondence: George Bouchagiar, Tilburg Law School, Tilburg University, The Netherlands.

Received: November 12, 2018 Accepted: January 7, 2019 Online Published: January 23, 2019

doi:10.5539/res.v11n1p27 URL: https://doi.org/10.5539/res.v11n1p27

Abstract

Automated decision making is both promising and threatening. Processing the biggest data possible may lead to societal advances but also violate human rights. There is, then, an acute need to protect individuals without impeding major benefits. Non-human agents may be biased; and they may not lend themselves to easy explanations. Instead of focusing on interpreting models, there seems to be a shift toward a concept of risk assessments. Opaque systems are aimed at predicting, or forecasting, future situations. This challenges human values and ethical principles. Even though incorporating ethics in machines is an old subject of legal discussion, consensus has not yet been reached; for theories and values may be controversial. This paper examines whether there could be an agreement on fundamental principles. A commonly understood basis could allow for fair and proportionate mechanisms to address crucial aspects of partiality and opacity in automated decision making. It could trigger a shift toward a concept of '*tracking the trackers*' and a discussion on a '*right to an unbiased decision maker*'.

Keywords: automated decision making, bias, opacity, predicting, ethics

1. Introduction (To Automated Decision Making – In Health and Elsewhere)

Since the middle of the 20th century, the use of archives of human specimens, collected in the course of healthcare, has led to important advances in medical knowledge. These advances have, in turn, produced crucial benefits for human kind (Korn, 2000). Although automated medical decision making is not new (Gleser & Collen, 1972: 180-181), the speed of information growth is said to have exceeded Moore's law (Chen & Zhang, 2014: 314, 318, 320) and its value is believed to have exceeded oil's (The Economist, 2017: 7). Therefore, large amounts of data are being processed and another 'V', for Value, is attributed to Big Data (Chartier, 2016).

Focusing on societal value, data processed *en masse* can support genius and felicitous medical decisions (Bates et al., 2001; Sintchenko & Coiera, 2003). But, to do so, 'general' information will not suffice. Processing agents need personal data (Gostin & Hadley, 1998; Chamberlayne et al., 1998). They need large samples of them; for small datasets can often result in incorrect scientific findings (Ioannidis, 2005: 0697; Schuemie, Ryan, Hripcsak, Madigan, & Suchard, 2018).

Despite success stories in medical research (González-Ballester, 2018), personal data concerning health constitute '*special categories*' of information and merit specific protection (GDPR, 2016: Article 9(1)). Namely, the genomic sequence can reveal sensitive data, like ancestry, one's ability to metabolize drugs or diseases she might develop (Azencott, 2018). Hence, there is a growing need to protect privacy without impeding scientific and societal advances.

Ethical principles for medical research demand –amongst others– informed consent, a *sine qua non* condition, and confidentiality (World Medical Association, 1964). Informed consent is a challenge. In many cases, physicians need to share patient's data to draw 'everyday' conclusions; it might be unreasonable to obtain informed consent each time data were transmitted between professionals to make a diagnosis. Thus, it could be argued that, in some cases, when confidentiality is maintained, data processing without the data subject's consent constitutes a theoretical harm; it may not be a real danger (Capron, 1991: 189).

But this could be an exception. Healthcare is one among the many fields where non-human agents perform their role as decision makers; and health-related data, like genome, are some among the many sources that can reveal sensitive information. For instance, consumers' buying history can predict health status (Robinson, Yu, & Rieke, 2014: 6). Contemporary algorithms (or 'traditional' dating from the '70s) make automated decisions of utmost importance in any domain (Veale & Edwards, 2018: 399; Edwards & Veale, 2017: 25). They do so, camouflaged by the ambiguity and technical jargon of a firm's terms of service; such terms may not guarantee confidentiality. Needless to say, *informed* consent (Papadimos et al., 2018: 1845) is rarely met; it is often obtained *via rental-car-contracts-like* forms, which make it certain that people will not read them (Hilts, 1995). A number of cases (Lipworth, Mason, & Kerridge, 2017: 486) and numerous jokes (Chesterman, 2017) can prove such claims. This raises tremendous concerns; a system may deny loans and visas, target citizens for scrutiny (Kroll et al., 2017) and anorexics with emetics (Veale & Edwards, 2018: 401) or score customers, job applicants and criminals (Chander, 2017).

Ranking people by numbers may be as old as human society (Dixon & Gellman, 2014: 6). But credit scoring as business practice has existed since, at least, the middle of the 20^{th} century (Gunter, 2000: 445). That was a time when authors found that discriminant analysis could predict how people would repay (Durand, 1941; Dixon & Gellman, 2014: 13; Hui, Li, & Zongfang, 2013: 911). Scholars, then, became seriously concerned that anyone could be classified as high risk at some point in her life; and courts began dealing with subjecting humans to some kind of test (Matthewman, 1984: 1190, 1199; Supreme Court of Illinois, 1966). Today, algorithms may score the previously unscorable; they aim to predict almost anything (Purtova, 2018). Being secret or unregulated, they may invade people's privacy, hide discrimination or deny due process (Dixon & Gellman, 2014: 6, 10). They can, thus, threaten or directly violate an array of fundamental human rights (Mantelero, 2018: 761, 771; Latzer, Hollnbuchner, Just, & Saurwein, 2016).

Yet, major benefits are still being promised. Eco-innovative technologies, like smart plant systems (Smart-Plant, 2018), can foster efficiency, enhance green economy and liberate consumers from household chores (European Commission, 2013). Speech to text applications can simplify life and help people with special needs (Fearn, 2018). Smart lights can optimize energy use and guarantee sustainability (Ritchie, Kessler, & Sargent, 2018). Real-time price monitoring and tracking can offer consumers the best deal (Competitoor, 2018; Prisync, 2018; Competitor Monitor, 2018). The list goes on and it seems that applications can be endless.

The ethos has become 'the more data, the better' (Citron & Gray, 2013: 262) and juggernaut-like firms process the *biggest* data possible (Dixon, 2018). Societal goals can be –and are being– achieved, but also human rights may be –and are being– threatened. Therefore, the challenge seems to be straightforward: how to protect fundamental rights without impeding societal, scientific or other advances of major importance (Žliobaitė, 2017: 1065).

This paper focuses on automated decision making. Like human minds, machines may be biased. Even though justifying decisions could be a way to diagnose bias and avoid partiality, non-human agents do not always lend themselves to easy explanations. Further uncertainties can occur, for people, forced by the *will to know*, tend to work *against* future uncertainty. They focus on a notion of risk assessments; opaque and unintelligible systems are aimed at forecasting future scenarios. A general shift toward regimes of control-by-risk challenges human values and ethical principles. Although humans, assisted by technologies, could perform a teaching role and embed such principles within systems, debates on controversial values render 'agreement on ethics' impossible. A perfect ethical theory to address every case and scenario may never come, albeit, a *free* and *willing* consensus on fundamental principles could be achieved. Fair and proportionate strategies, built upon a commonly understood basis, could address important aspects of opacity and partiality; they could lead the way toward a notion of '*tracking the trackers*' and a potential '*right to an unbiased decision maker*'.

2. Algorithmic Partiality

Decision theory may be concerned with identifying the best decision to make based on the assumption that the decision maker is fully rational –or bounded rational (Wang, Xu, Hamido, & Liu, 2016). It can deal with the question of how an agent may rationally choose among available courses of action (Cooper, 2003: 43). Moreover, decision-making can be regarded as a process that leads to a certain choice. It can be seen as a procedure that comprises the phases of finding occasions for making a decision, finding possible courses of action and, finally, choosing among courses of action (Simon, 1960: 1). And Big Data involve(s) a number of steps, starting from finding instances (such as data acquisition, collection, capture or storage) and resulting in a certain choice, a decision (Janssen, Voort, & Wahyudi, 2017: 339). Therefore, decision making (to some, rational and, to others, biased) may exist in many procedures of Big Data, from collection, compilation or consolidation to data mining and use (FTC, 2016: 3; Wang et al., 2016: 748; Voort, Klievink, Arnaboldi, & Meijer, 2018; Chen, Chiang, & Storey, 2012; O'Malley, 2014).

Even though machines do make decisions based solely on automated processing (Geist, 2016), human decision makers have not yet become *has-beens*. Systems, such as the legal one in some jurisdictions (Harvard Law Review, 2017; Katz, 2013), may be based solely on human judgment. But people are not perfect. Although human imperfection is said to be a core dimension of freedom (Benkler, 2013), in many fields, like healthcare, some accuracy would be needed to benefit society as a whole.

Human decision making has not yet been fully understood. It is not very clear what influences people's decisions. To some, people are self-interested and, thus, make decisions that provide the greatest benefit or satisfaction (Smith, 2005). From this perspective, decision making could be simple; one may perform a cost-benefit analysis and pick the option that would maximize the benefits and minimize the costs. Yet, as imperfect (but free) beings, humans do not always seek to maximize benefits; 'prisoner's dilemma' has proved that they are not always rational (Bravetti & Padilla, 2018). Therefore, human decision making process can depend on several factors and the outcome may not be predicted by a mere cost-benefit analysis. Biases, like imperfect information, processing capabilities, norms or resistance to change, could limit rational decision making (Kahneman, 2011: 255, 417, 430). Thus, people's brain, with its well-hidden bias, may not be transparent; one does not always know how she, herself, makes decisions. As some have aptly put it, it would be *universal madness* to think that one generally knows what she is doing (Katsafanas, 2018: 123-124).

Biases in people's decisions call for objectivity; and some claim that algorithms could fulfil this need. It has been argued that machines are different from human minds (Freitas, 2012; Vogel, 2008) and that data driven systems have the potential of avoiding biases (Rudin, 2016). According to this argument, systems could be less subjected to biases and, thus, come up with more accurate decisions (Rudin, 2016). Rationality and its instances, being captured in mathematical terms *via*, among others, game theory (Cooper, 2003: 14), could be involved. If this were the case, some levels of objectivity could be met (Kumar, 2017: 11).

However, the vast literature devoted to this matter disagrees. Non-human agents *claim* to be neutral, but they are not (Barocas & Selbst, 2016). Prior work has repeatedly revealed that the *unbiased-argument* is an unsustainable belief (Richards & King, 2016: 10-13; Förster & Weish, 2017: 15, 19; Boyd & Crawford, 2011; Mantelero, 2016: 239-240; Crawford & Schultz, 2014: 93-95, 98; Rubinstein, 2013: 76; O'Neil, 2016: 3-5, 130-134, 151). Humans are not perfect and Artificial Intelligence (AI) is a human product (Bostrom, 2014), a creation of human design. Therefore, AI is neither perfect nor objective (Crawford, 2013); the claim that algorithms decide more objectively cannot be taken at face value; human judgment may be built-in (Burrell, 2016: 3).

Thus, impartiality, implicating *inter alia* freedom from bias (McManus-Degnan, 2010: 226), may not be met. It would, then, be reasonable to demand that these agents explain their decisions. This could be a way to diagnose partiality and guarantee freedom from bias.

3. Explaining Automated Decision Making: A Failure?

People may use arguments to justify their claims or beliefs. For instance, a lawyer can present her claim at the trial, describe evidence to support it, detect laws that apply, call experts or refer to previous cases. Similarly, a physician may use arguments to justify a diagnosis.

But could machines use explanations to justify their conclusions?

Explanation techniques have long been proposed (Shankar & Musen, 1999). Explaining why a car stopped when the system detected a red light might not be that meaningful. However, knowing how the machine reached an outcome, when deciding whether one's life should be terminated, would be important.

A good example can be found in *Ms B v. An NHS Hospital Trust* (EWHC, 2002). A 43-year-old woman suffered from hæmorrhage into the spinal column in her neck. She became paralyzed from neck and down and unable to breathe without ventilator support. After unsuccessful treatment, she requested that the ventilator be switched off. Yet, physicians felt they were being asked to kill her and refused (EWHC, 2002: 57). Ms B exercised her right to refuse treatment and brought her case to court. The ventilator ceased and Ms B died (Slowther, 2002).

What is of great interest is that physicians claimed the issue of desirability of her survival as ethical in nature (Foster & Miola, 2015: 509). Were an algorithm in charge, it might not see such human point of view (Taylor, 2015: 514); claim could be different. But, whatever the agent's thoughts and biases, it would be meaningful and important to know and understand its decision making. Same would be the case, when deciding if one should try a specific medical treatment or whether one would be able to pay for a loan or be eligible for parole (Angwin, Larson, Mattu, & Kirchner, 2016; O'Neil, 2016).

Yet, explanations may not be possible.

Namely, there could be obstacles of legal nature. Firms would not reveal their secrets (Reddix-Smalls, 2011). A non-human agent could be secret and its controller could have undertaken steps required to keep it secret; and, virtually, any system may impart value to its owner as a consequence of this very secrecy (Jandoli & Dani: 46). Therefore, trade secrecy laws can apply (Directive 2016/943: Article 2(1)); and they cover a wide array of information, including drawings or calculations relating to, *inter alia*, processes (Jandoli & Dani: 49). It seems that such information, well hidden in firms' back-offices and guarded like the Coca-Cola formula (Harcourt, 2005: 5), would not be easily subjected to audit or review (Reddix-Smalls, 2011). Besides, the –recent but underdiscussed– European Directive on trade secrets does not allow for exceptions as regards researchers –who could scrutinize processes. It only excepts journalists (Directive 2016/943: Articles 1(2)(a), 5(a); Recital (19)). But journalists cannot decipher.

Moreover, explanations could be impossible due to lack of knowledge. Even if firms revealed their secrets, people would lack literacy to comprehend relevant procedures. What Bhargava et al. (2015) describe as the desire and ability to engage constructively in society through and with data, the literacy that enables individuals to understand principles and challenges of data, is most probably missing. Were decisions explained, people might not understand.

In addition to lack of knowledge, there is also *asymmetry* of knowledge, an *information advantage* enjoyed by machines (Matthias, 2004: 182). If people had skills, the very system might not lend itself to easy explanations. There is lack of interpretability (Lisboa, 2013) and experts may not be capable of comprehending how an outcome was reached (Edwards & Veale, 2018: 49).

Besides, if machines were 'generous' and lent themselves to easy explanations, this would not *per se* mean de-biasing. Biases can arise for a number of reasons. They may exist independently of and prior to the creation of a system; they may emerge from resolving issues in design; or, most importantly, they can arise in a context of use, as a result of altering knowledge, population or values (Friedman & Nissenbaum, 1996: 333). Data collected may have been preferentially sampled (Crawford, 2013); they may be inaccurate, failing to represent reality; they can reflect human culture (Caliskan, Bryson, & Narayanan, 2017: 183); they may be hidden in the code; or they can emerge due to a shift in context (Friedman & Nissenbaum, 1996: 335).

Therefore, diagnosing and avoiding bias may be impossible (Veale & Binns, 2017). And it seems that challenges go beyond failure to de-bias. Tempted by the *will to know*, humans tend to focus on predicting technologies. This triggers a shift toward a concept of risk assessments that may, in turn, result in further uncertainties.

4. Working Against Uncertainty

Non-human agents tend to diffuse information in ways difficult to decipher. They may, then, be as opaque as human minds (Castelvecchi, 2016). To many authors, the problem lies in the *black box* function of AI (Pasquale, 2015). Although transparency has been regarded as the key to audit or monitor algorithms (Allnutt, Dickson, & Webster, 2016), the question may not be 'how to whiten the black box'. Rather, it may be 'what would transparency mean, when what comes into light, what is rendered transparent, escapes comprehension?'. In other words, the very value, or even usefulness, of transparency may be challenged, when what is under scrutiny is being updated; it is a later version of what was supposed to become transparent.

It might take some time for people to understand what algorithms can perform in seconds. Humans lack both time and skills; and there is a tremendous discrepancy between cyberspace and cybertime, an incompatibility of space with time in cyber-terms. As Berardi (2014: 242) has put it, the expansion of cyberspace is boundless, albeit, cybertime is not infinitely stretchable; for it is composed of time of attention and subjected to the latter's physical limitations.

But, in human-terms, time has been understood as a dimension where the past is, now, known and the future is, presently, unknown. Future is something that will be, that is not here. It portrays uncertainty (Esposito, 2015: 95). Working *with* the idea that the future is unknown can allow *preparing* for the future (Langley, 2013: 55, 56, 60, 69). Working *against* the notion of uncertainty can allow for techniques aimed at *predicting* the future (Esposito, 2015: 99).

If the beauty of the future is in its uncertainty (Esposito, 2015: 106), it seems that, today, future is aimed at becoming *ugly*. For, now, data may be used with a view to predicting future situations (Pearsall, 2010: 16). Yet, with the objective (or instead) of predicting, agents may render future uncertainties actionable today. If future can be seen as the performative consequence of actors and actions (Latour, 1988: 165), algorithms may produce a 'desired' future. Worse, they can reproduce past –and its injustices.

To oversimplify, let us assume that a system collects data from areas 'A' and 'B'. In 2018, 500 crimes were totally committed in area 'A'; 400 of them were committed by whites; and 100 by colored. But only the latter 100 were recorded.

Here, the 'unfair past' would be in the belief that colored people –whose crimes were recorded– are more dangerous than whites –about whom data are missing. The 'desired' (yet unfair) future would be in focusing on colored; for data recorded suggest that they be high risk.

Let us further assume that in area 'B' 10,000 crimes were committed in 2018. All of them were committed by whites, but no crime was recorded. In the absence of data, 'B' would be thought of as a safe place to live in.

The computer science adage goes 'garbage in, garbage out' (Vayena, Blasimme, & Cohen, 2018). The system *spits back* what was put into it (Ferguson, 2017b: 47; Llenas, 2014). Even though more crimes were committed in 'B', it will consider area 'A' as high risk. Although whites committed more crimes, colored people will be scored as dangerous. The algorithm may, then, suggest that police forces draw attention to area 'A' and the colored. This way, non-human agents not only perpetuate existing injustice, but also create new; a 'desired', and unfair, future.

This scenario is far from being ethically neutral. But it is real and happening.

The development of technical knowledge in predicting technologies, but also its driving force, i.e. the *will to know*, triggered the shift in people's conception of just punishment from notions of reform or rehabilitation to concepts of risk assessment (Harcourt, 2005: 32). Therefore, people have come to believe that it is just for punishment to primarily relate to the statistical probability of reoffending (Harcourt, 2005: 33). Yet, biased inputs and outputs transform 'predictive policing' into '*crime forecasting*'. Like weather forecasting systems, these agents may become more accurate (Ferguson, 2017a: 1137, 1144). But this focus on (potential) utility seems to ignore fairness and justice; it prioritizes predictive performance over interpretability (Edwards & Veale, 2017: 26). This work *against* uncertainty, occurring in various areas, allows for a general shift toward risk assessments, '*razor sharp segmentation games*' or a regime of control-by-risk (Poon, 2009: 656-659).

'Slowing down' algorithms would come counter to what they are being created for. Systems need to be 'fast' to provide benefits. As Mander (1978: 43-45) has put it, if one accepts the existence of cars, she also accepts the existence of roads laid upon landscape, oil to run the cars or firms to pump and sell oil. In the same vein, one could argue that, if people accept algorithmic speed, they also accept inability to control; accepting automated decision making would, then, mean accepting both its benefits and its risks.

But there may be values higher than speed and efficiency (Perry, 1977: 412). So, could there be another way?

It seems impossible to understand a system that remains unintelligible, even when transparent. These agents consume more and more data to become even more powerful; ever-growing, yet invisible to humans, they are black holes, rather than boxes. In this arena, the unpredictable can be forecasted. People may be 'punished' for behaviors they may never express, mistreated for entirely internal attitudes they may never manifest or denied for future scenarios that are not known and may never occur.

Being colored and living in area 'A' (of the above scenario) does not necessarily mean one is dangerous; nor does it necessarily mean that police forces should focus on her. Same can be the case with other domains where algorithms perform their *forecasting* role. Namely, in health, the mere possession of a deficiency or a trait does not necessarily mean one will later become ill or disabled (Matthewman, 1984: 1188). Denying loans, jobs, paroles and any opportunity because in some years an individual may, but also may not, die from smoking, alcohol, cancer or junk food could be unfair or unreasonable.

Laws can transform messy phenomena into ordered, knowable and calculable cases (Dove & Özdemir, 2015: 530). They may regulate, among many others, activities (Friedman, 2016), behaviors (Kim, 2018), markets (Mugarura, 2016: 603, 605), populations of living species (United Nations, 1992c) or the management of the dead (Gaggioli, 2018). They can regulate expressions, like actions through which a state of mind, a belief, a mood, an emotion, an attitude or an idea may manifest itself (Anderson & Pildes, 2000: 1506). If one, holding the belief that 'war is wrong', attacks an embassy for a nation's involvement in war, he can be punished for attacking, for expressing his belief in such way. But he cannot be punished for his very belief that 'war is wrong'. Laws (in jurisdictions we know) do not regulate –nor do they punish people for– entirely internal thoughts or ideas. What regulators left unregulated, people's deepest wish or fear (Battelle, 2005: 1-9; Tene, 2007), is increasingly becoming the 'input', the reason for denial or unfair treatment.

It seems that the attempt to eliminate and work against uncertainty of the future is alarmingly diminishing certainty of today, or people's certainty about themselves. It is challenging people's values, what it means to be transparent or fair. It is, then, challenging ethics.

5. Ethics in Machines? A Need for Consensus

Analyzing algorithmic processing without discussing ethics may be naïve and, to some, deceptive (Mutlu, 2015: 1002).

Ethics refers to values and ways to define right and wrong actions (O'Leary, 2017: 14). Dealing with duty, what people ought to do (Merrill, 2011: 3), it can be understood as the study of one's *responsibility* to discern the rightness or wrongness of her actions; the art or science of one's efforts to live in proper harmonious relationship with others (Kruckeberg, 1989: 11). As such, ethics involves thoughts designed to bring about *good* within a society (Ray, 1996: 48).

Decades ago, some principles were suggested to –bring about such *good* within information and technology societies and– promote, among others, transparency or oversight (U.S. Department of Health, Education and Welfare, 1973). But consensus has not yet been achieved; over the years, several legal instruments have proposed a number –each one its own number– of principles (OECD, 1980; A29WP, 1998; FTC, 2000). The failure to implement these recommendations may have been due to their *legalistic* nature (Cate, 2006: 343, 355, 366). However, regardless of their nature, the very principles these recommendations suggested could be seen as the *building blocks* of information law (Schwartz, 1999: 1670). Therefore, were consensus achieved, it would be desirable that ethical principles (or ethical features; Kraemer, Overveld, & Peterson, 2011: 252) be embedded within *design* specifications of technologies (Dove, Knoppers, & Zawati, 2014).

The idea of incorporating principles in *design* has been developed in a number of frameworks, such as the 'Appropriate

Technology' (Schumacher, 1973), the 'Participatory Design' (Asaro, 2000) or the 'Value Sensitive Design' (Manders-Huits, 2011; Friedman & Kahn, 2003: 1186). Design choices have historically had ethical import. Moses's bridges are a good example. In 1930, Robert Moses was asked to design overpasses for the highway connecting New York to Long Island and Jones Beach (Winner, 1980). Moses designed very low bridges; cars could pass, albeit, the twelve-foot tall buses (that is, public transport oft-used by minorities) could not get through the overpasses (Winner, 1980: 124). Regardless of whether low bridges were intentionally designed to prevent minorities from getting to the beach, Moses and his bridges are often quoted to exemplify potential ethical (or even political) import of design choices (Manders-Huits, 2011: 272-273).

Yet, *design* is just one among the many shape-defining aspects of technology (Flanagan, Howe, & Nissenbaum, 2008: 330). Other crucial steps may include, to name but a few, production, deployment, adoption or distribution. For instance, in case of Stallman's (2010) '*free as in freedom*' concept, it could be fair to argue that what determines free software's shape, its magnitude, and what is of utmost importance (without prejudicing its design) is in its use, adoption and (re)distribution, as described in Stallman's (2015: 3) *freedoms zero to three* (to run a program for any purpose; to study and change it; to redistribute copies of it; and to distribute copies of one's modified versions).

A concept that seems to focus on multiple steps of technology, its whole life cycle, is privacy-by-design. Even though its 'name' refers to design, its interpretation, as provided in *small print*, footnotes, soft laws, press releases (European Commission, 2010: 17; 2015; EDPS, 2018: 4, 10) or literature (Diver & Schafer, 2017), suggests that the relevant principle (here, the protection of privacy) be embedded throughout the entire life cycle of technologies, from the early design stage to their deployment, use and ultimate disposal. Therefore, it would be reasonable and fair to suggest that ethical principles be promoted and incorporated in the many shape-determining steps of technologies.

This could resemble ways people teach ethics. This idea of human teaching could be supported by non-humans; for it is argued that, when people and machines cooperate, results can be better (Kraft, 1964: 100). Or, to cite Katz (2013: 929), the equation could be simple: Humans AND Machines > Humans OR Machines.

Coined some decades ago (Samuel, 1959) and similar to human learning (Informatics Europe & EUACM, 2018: 6), machine learning can make AI respond in different situations (Jordan & Mitchell, 2015; Veale & Binns, 2017; Hastie, Tibshirani, & Friedman, 2009). An agent may, then, become better capable of reaching an outcome. For instance, in fields of healthcare, the ethical goal might be the prevention of a disease. Hence, AI could learn the settings that would be the most effective to prevent, e.g., a possible outbreak of a disease. A disease may have multiple symptoms and similar or same symptoms can be attributable to many diseases (Malmir, Amini, & Chang, 2017: 95). This could confuse even the most experienced physicians (Adlassnig, 1986; Conejar & Kim, 2014). AI could correlate data (King & Mrkonich, 2016; Mayer-Schonberger & Cukier, 2014: 68; Podesta, Pritzker, Moniz, Holdren, & Zients, 2014) and provide patterns that human physicians would never reach. Machines could, then, exclude unrelated symptoms, deliver a more accurate decision and, hence, reach the ethical outcome (this, of course, is a mere example to simplify ways in which machines may learn; it does not take –nor does it aim to take– into consideration further criteria, such as impartiality of input data, that would need to be met to reach accurate outcomes).

Yet, some have questioned this idea of teaching or learning ethics (Lorenc, 2015). It may be true that humans are not always the best teachers; for *selflessness* and *bringing about good* may not be what people desire. For example, few might want to buy a smart car that would look after others and prioritize general good over the driver's own well-being (Metz, 2016; Bonnefon, Shariff, & Rahwan, 2016). Were such preferences taught, machines would acquire, not ethical but, common behavior.

Therefore, if teaching/learning ethics were the answer, the question would remain 'which principles should be implemented?'.

Scholars have long been discussing ethics. Theories range from Aristotle and Plato's ethical virtues (Price, 1989; Polansky, 2014), relations between ethics and metaphysics (Guyer & Wood, 1998; Whittaker, 1916), the *apotheosis* of the average and the science of proportion (Thomas, 1896; Knapp, 2007) to the stakeholder theory, business ethics (Wijnberg, 2000), marketing ethics (Singhapakdi, Vitell, Rao, & Kurtz, 1999), the ethics of data collection (Mutlu, 2015: 1000), military ethics (Mileham, 2015), the ethics of terrorists (Dreisbach, 2011), or 'the ethics of Google Earth' (Sheppard & Cizek, 2009).

No agreement has been reached (Ware, 2015) and contemporary dilemmas seek for an answer (Birnbacher & Birnbacher, 2017). Should a machine maximize happiness by terminating the life of an old man to save ten young women? What if the old man was a genius lawyer and the ten were criminals?

Disagreements can emerge for ethical principles may be controversial; each one may have exceptions or counter examples (Friedman & Kahn, 2003: 1178). For instance, utilitarianism, an influential version of consequentialism, seeks for the greater good (Harris, 2008/2009: 68). But it may permit actions that some would consider unjust, unethical or unfair. The ends could justify the means and the means could *get ugly* (Grau, 2006: 52). Furthermore, by focusing on greater good, utilitarianism is said to ignore individual identity, to *not take seriously* the distinction between persons (Rawls, 2009: 24). In contrast, deontologists focus on the very action; the right action may not be the one that maximizes good. There could be constraints on what one may do, some instances in which maximizing the good may not be right (Rawls, 2009). To commentators, deontological positions are not always clear; for they do not state what these instances are (Harris, 2008/2009: 69). It is further argued that there is no consensus on the very notion. Some deontologists suggest 'requirement', 'permission' or 'prohibition' as basic deontic terms, albeit, others opt for 'obligation' or 'blameworthiness' (Nottelmann, 2008: 325-327).

Further disagreements can arise when theorists attribute values to entities (Davidson, 2013). For instance, a knife may have *utilitarian* value; people use knives to eat. But it can also have *intrinsic* value; it may have been handed down to an individual from her parents. Therefore, she might need to care for what the knife *is*, not for what the knife *does*. Moreover, a ten-year-old kid, doing chores, might have some kind of *utilitarian* value; and the kid has *intrinsic* value, even if she does nothing.

Therefore, one could claim that entities can have *utilitarian* but also *intrinsic* value. Yet, others argue that there is either *utilitarian* or *intrinsic* value, either *price* or *dignity*, not both. According to this argument, if something has price, it can be substituted; if it has dignity, it cannot (Wood, 2018; Vöneky & Wolfrum, 2013: 96). From this perspective, humans have dignity, not price. But still, things are not always clear. Namely, governments are said to assign economic value to people (Davidson, 2013: 174) and firms claim they can measure the economic value of personal data (Malgieri & Custers, 2018).

Many theories and perspectives can lead to endless debates. Were humans the teachers, the ones who would embed ethical principles within systems, uncertainties might occur. Same would be the case with machine learning. Enabling AI to learn and perform ethical actions to reach desired goals would also require some consensus. Humans would need to agree not only on situations to which agents would be exposed, on which they would be trained, but also on the very desired ethical actions and goals.

Instructing machines to do things at which people are *bad* or things that require some intelligence –in narrow tasks– may not be that 'hard'. Namely, with chess (Dickson, 2017; Bostrom, 2014), the machine has to win; people know what winning looks like (Metz, 2016). Yet, aiming to program things that come more naturally to humans, but not to machines, is admittedly challenging (Deng, 2015: 26).

Scientists estimate that AI singularity may soon create its own ethics. Even though these are experts estimating, not dreamers dreaming (Hauer, 2018: 104, 106), such estimates have been regarded as the 'rapture of the nerds' (Popper, 2012). In any case, machines have not yet become super-intelligent (Bostrom, 2014). Nor are they truly autonomous; they will be the day they are instructed to go to work and they instead go to the beach (Markoff, 2015: 333). But algorithms collect and use data to make decisions and guide their 'actions' in line with their goals. This way, they may effectively display some capacity for *choice* and *intentional actions*, an *almost* or *quasi* free will (Malle, 2016: 249). Occupying the position of subject in sentences, they go beyond *chess* or *Go* (Silver et al., 2017); they intervene in real life and are embedded within larger social systems (Ziewitz, 2016: 5, 7). They, thus, carry some form of what humans used to monopolize; agentic power (Beer, 2017: 5).

Seeking for an object, some principles to agree on, in the *teaching/learning argument* and revisiting the dilemma '*should a machine maximize good by killing an old man to save ten young women*', one could argue for utilitarianism. To many, machines could make felicitous correlations to 'calculate greater good'. Yet, such 'good' may have many interpretations (Emanuel & Wertheimer, 2006: 854); and deontologists might be right: killing may not be ethical.

Providing rules is not always for solving dilemmas; in some cases, giving ethical rules for solving dilemmas could be like giving rules for telling jokes (Wijnberg, 2000: 333). Ethical battles may never result in the perfect theory that will provide an explicit list of all desired principles, their exceptions, constraints or instances.

But this cannot be a reason to give up on trying to do what we believe is ethical, right, fair or just. Nor can this be an excuse to stop teaching ethics to our kids. For, whichever theory is correct, there seems to be some consensus, an agreement on some form of *responsibility* to make the world a better place, to *bring about good* or, at least, to *do no harm* (Otto, 2007: 20; Hunt, 2014). We may lack complete knowledge and may never discover the perfect ethical theory, the perfect fit for each case and scenario; we may, thus, not know exactly how to advise our children. But we do (Sutton, 2010). Because they are developing and they have to develop in some ethical way. They cannot just 'exist'. Nor can they simply 'act and behave'. For their actions will affect the world; and the world needs to get better; harm needs to be avoided.

It could be argued that, unlike children, machines cannot have ethics. But, like children, they seem to be developing. Like kids, they seem to have the potential to ‘act and behave’ in ways that affect the world in domains of major ethical, societal or economic significance, from healthcare and education to trade, business, entertainment or security (Malle, Scheutz, Forlizzi, & Voiklis, 2016: 125; Bendel, 2016: 104, 106, 107). These developing beings can influence and shape people’s decisions and lives; they can provide major benefits, but also threaten human rights. Therefore, they cannot simply ‘exist’.

There has to be some kind of consensus on *responsibility*. For, even though machines have been referred to as *thoughtless* (Meyer, 2014), those who craft them (when humans) are (*expected* to be) thoughtful.

6. Toward a Free and Willing Consensus

Each culture may have its own laws, norms, god or time. Diversity is widely acknowledged (Melé & Sánchez-Runde, 2013: 682), not as a condition that will pass away but, as a permanent and desired feature of democracies and, in general, societies, communities, groups or families and their members.

But, in a number of cases, needs have to be satisfied. When shared, such needs can lead to agreements. Namely, people drafted agreements on coffee or sugar (United Nations, 1968; International Coffee Organization, 2007; United Nations, 1992b), for they recognized the importance of such goods to economies that are dependent upon them for the achievement of social goals (International Coffee Organization, 2007: 1).

Furthermore, situations, involving unpleasant or extremely cruel acts and behaviors, can result in common needs and, thus, agreements. For example, *barbarous* acts created the need for *common understanding* of human rights. That is, an acute need that resulted in agreement on freedoms, immunities and benefits that everyone should enjoy in the society in which she lives (Skinner-Thompson, 2018: 10.05[A]; Universal Declaration of Human Rights, 1948). Humanity had to develop on the common basis that some fundamental rights are not created, but deriving from human dignity (United Nations, 1966b).

It was not only sugar, coffee or the human rights. Having so far won the struggle for survival, human beings share common adaptive characteristics (Remoff, 2016: 884, 885) or the intrinsic need to adapt to changing environments. Therefore, agreements have been reached on the Ozone Layer (United Nations, 1985; United Nations Environment Programme, 2018), climate change (United Nations, 1992a), desertification (United Nations, 1994), health (United Nations, 1946) and a number of other issues (United Nations, 2018): from diplomatic relations to the prevention of genocide (United Nations, 1961; 1948); from economic, social, cultural, civil and political rights to rights of children, workers, refugees and people with special needs (United Nations, 1966a; 1966b; 1989; 1990; 1951; 2006); from opium smoking to traffic in women and children (United Nations, 1931/1946; 1921); and from trade, transport and communications to education (United Nations, 1947; 1949a; 2005; 1949b).

Such consensus, satisfying common needs that derive from external factors or the human nature itself, has allowed humans to survive and thrive, to cooperate and become better capable of getting things successfully done. Shared needs led to shared views; shared views, then, performed the vital function of enabling communications to proceed on a *commonly understood* basis (Anderson & Pildes, 2000: 1518).

And it seems that views could be shared in the arena of automated decision making.

Like coffee and sugar, algorithmic procedures are crucial; they are important not just to economies but to people’s lives. This importance could be commonly acknowledged, for people depend on non-human agents to achieve societal advances.

Furthermore, algo-harms may not involve *blood* or *barbarous* behaviors (Bartow, 2006: 62). Yet, unfair or unethical –and, to Meyer (2014), *cruel*– situations can arise. Namely, not all individuals can enjoy what fundamental human rights aim to guarantee, such as the equal and fair prospect to access opportunities available in a society (European Union Agency for Fundamental Rights, 2018: 42). The latter, often referred to as ‘digital discrimination’ (Edelman & Luca, 2014), is not tied to a specific territory and, thus, calls for universal measures. Same seems to be the case with unfair situations that may occur in other non-territorial areas of cyberspace.

There is also the (intrinsic) need to evolve and adapt to rapidly changing technological environments. This has been repeatedly highlighted by policy makers and organizations (A29WP, 2017; European Consumer Consultative Group, 2018; FTC, 2016; Informatics Europe & EUACM, 2018).

Automated decision making is not just a scientific challenge. Leading to debates over what it means to be human, it can be understood as a political, economic, societal, cultural and philosophical challenge. Therefore, it calls for consensus, plans for humans to thrive when machines decide (Larus & Hankin, 2018).

If people do not agree on *good life*, they could agree that making the world a better place or at least removing pain are goods to be defended (First, 2018: 554). There could be a plenitude of shared views and those within differing areas of ethical *taste* could be discussed (Kruckeberg, 1993: 30). Thus, there could be a common understanding based on freedom, the necessity for ethical action and moral agency (Cooley, 2013: 367; Merrill, 2011: 8-9); a consensus to create a more symmetrical atmosphere (Hunt & Tirpok, 1993: 5-6). This could be an *overlapping consensus* (Rawls, 1987: 2-5) on *conventional morality* (Perry, 1977: 388); a *free* and *willing* agreement whose object and grounds would be in an *independent* ethical philosophy beyond controversial parts of theories (Baier, 1948). Instead of relying on positions at a given time or fragile circumstances that may render agreements instable, it could focus on principles and significant aspects of society (Rawls, 1987: 15; Perry, 1977: 388). Having its roots in ethics, it could be a consensus on *responsibility* to discern rightness and wrongness; to specify and prioritize fundamental principles; and to support measures that would guarantee effective enjoyment of rights and liberties associated with these principles (Rawls, 1987: 18).

7. Tracking the Trackers and a 'Right to an Unbiased Decision Maker'

Automated decision making, promising or threatening, may raise ethical questions, challenge legal certainty or be a cause for societal concern. Few might give their informed consent to data processing for medical research, albeit, millions would scream if they were denied access to benefits deriving from such processing. Tremendous challenges emerge from the acute need to protect human rights without impeding societal, scientific and other advances of utmost importance.

Thus far, it has been argued that human decision making may not be accurate; people are biased. But so are automated agents. This creates a need for explanations, to which machines do not easily lend themselves. Like human mind, they may be opaque and unintelligible. Moreover, in the absence of skills and time, it may take years for people to understand tasks run by machines. But this can also apply to tasks performed by humans. An automated agent and a human being can be biased and opaque, albeit, the former is seen as thoughtless while the latter enjoys monopoly over ethics.

However, it seems that we, the 'thoughtful', challenge our own values by focusing on risk assessments and prioritizing (our will to know the future and thus) predictive performance over (the need to explain past or present and hence) interpretability. Then, it has to be humans the ones *responsible* for striking a fair balance; for reaching an agreement on well specified principles and mechanisms to enforce them.

That lack of transparency may be attributed to both human mind and AI is hardly surprising. The very goal of AI has been –from an engineering perspective– to make machines do things that would demand intelligence, were they done by humans, or –from a cognitive science perspective– to design systems that work the way human mind does (Chopra & White, 2011: 5; Citron & Pasquale, 2014: 6). Therefore, techniques, analogous to what humans do when asked to explain their decisions, could address some important aspects of opacity in automated decision making. To the extent that humans are considered interpretable, it may be this kind of interpretability that could apply (Preece, 2018: 67).

Such models have already been proposed. When transparency is unachievable, post-hoc interpretations can explain predictions; instead of revealing how a model works, they perform in a way similar to, for instance, verbal explanations provided by humans (Lipton, 2018). Opaque models may then be interpreted after the fact and without sacrificing predictive performance (Lipton, 2018). However, such techniques are not always perfect; interpretations are given *after* the fact –or the harm. But, in so far as they may accurately interpret procedures, they could be a way to better understand opaque systems. Therefore, instead of targeting investments *solely* in designing opaque models, people could target labor, efforts, time and money in developing systems to interpret these models. People could invest in *tracking the trackers*.

Experimenting in such techniques could result in a more symmetrical atmosphere. It could strike a balance between programming to predict and instructing to interpret; between potential utility and desired fairness. Importantly, such measures could reflect general and commonly accepted principles, such as *proportionality*; the *most far-reaching ground for review* (Chalmers, Davies, & Monti, 2010: 368) and one of the oldest general principles (Court of Justice, 1956). Namely, if a predicting system restricted a human right, it would be fair to demand that such restriction should be proportionate in relation to the public interest pursued; and that any limitation, even a proportionate, should not undermine the very substance of a human right (Schütze, 2015: 98). Such automated procedures could be allowed, if they were appropriate and necessary to achieve *important* objectives legitimately pursued (Cheyne & Alder, 2007: 181-182). Proportionality could put automated decision making to the test; it could analyze its suitability and necessity and determine whether a procedure was suitable but also the least restrictive means to reach a given objective (Schütze, 2015: 204-205). Therefore, it could strike a fair balance between conflicting rights or values (Wright, Hurles, & Firth, 2016). Aiming to *track the trackers*, to examine their suitability and necessity, proportionality could be a way to address crucial aspects of opacity.

In addition to lack of transparency, bias seems to be another common feature shared by humans and agents who aim to work the way people's mind does. In both cases, de-biasing may be impossible. In human-situations, a person bound by the veil of secrecy that protects her deliberations (Hernández, 2014: 128), yet required to be unbiased or impartial, is a judge (Neimanis & Matjusina, 2011; Cooper, 2006). However, judges are not *empty vessels* that litigants fill with content (Meron, 2005: 365); they carry prior experience and, hence, bias. As some have aptly put it, in the best of all possible worlds, even the slightest possibility of bias would suffice to disqualify a judge from hearing a case; but, under such scheme, there would be no one left to adjudicate anything; reality, then, forces people to tolerate some bias (Redish & Marshall, 1986: 492). Thus, authors have suggested that the concept of judicial impartiality could be seen as some kind of *awareness*, an *understanding* by a judge that she carries bias that she has to rationalize when making decisions (Hernández, 2014: 154-155).

In the same vein, scholars have argued that AI could be instructed to acquire some kind of *awareness* of what it does not know; something close to *consciousness* (Raskin & Taylor Rayz, 2016: 393). Regardless of whether a *quasi* consciousness could be attributed to AI, systems could be programmed to diagnose factors that diminish accuracy. Namely, instead of demanding people to detect bias in automated decision making, it might be easier or more reasonable to program a system to identify, for example, missing data (Taylor, 2015: 515).

Furthermore, measures safeguarding rights associated with fundamental principles, such as fair or due process, could to some extent address partiality. Experts could focus on bias minimization as a way toward a *right to an unbiased and accurate decision making*. In human-scenarios, this right is a key fair process value, the *floor* of due process (Redish & Marshall, 1986: 479; Poole, 1998: 236). It can allow for *procedural rationality* that can enable people to engage in rational planning about their situations and make informed choices among options; to have a better chance of '*knowing what is going on*', what is happening to them and why (Summers, 1974: 26-27).

In this context, due and fair process could demand that automated procedures satisfy some standards (Crawford & Schultz, 2014: 93-128). Mechanisms could require systems to confirm their fairness and accuracy in various steps of decision making. This could be a way to introduce a '*right to an unbiased decision maker*', to guarantee the basis of due process. Interestingly, aspects of what has been referred to as 'technological due process' (Citron, 2008: 1301) are provided in existing regulations (GDPR, 2016). For instance, certification mechanisms –already established, yet, on a voluntary basis– (GDPR, 2016: Article 42(3)) could further determine design specifications, expertise involved or performance expected. They could also specify information about the experts who designed a given system (such as field of expertise or years of experience) or the nature of data collected (Edwards & Veale, 2017: 79). Importantly, disclosure of these data would not violate trade secrecy; nor would it run counter to privacy expectations. Thus, such information could be publicly available (GDPR, 2016: Article 42(8)) to allow for review and oversight.

It seems that measures aimed at striking a fair balance between conflicting values or rights but also mechanisms focused on strengthening fair processes could rely on principles already proposed, accepted and analyzed by regulators, policy makers and the academia. Such measures and mechanisms may reveal that the problem can often lie, not in the absence of procedures, rules or principles but, in their (application or) understanding. Namely, proportionality, applied to trade secrecy on a commonly understood basis, could resolve a number of contemporary uncertainties (Directive 2016/943: Articles 1(2), 7(1)(a), 11(2); Recitals 21, 26, 28, 36). Undoubtedly, trade secrecy can have *its place* (Dixon & Gellman, 2014: 7). But who would argue for secrecy that denies due process, violates privacy or prevents justice? When *secrecy mistreats* or *threatens*, proportionality and its strict 'tests' could protect public interest and fundamental rights and freedoms.

8. Discussion

Humans have long been prioritizing themselves over 'others'. Conservation projects are a good example (Doak, Bakker, Goldstein, & Hale, 2014). Setting population goals for non-human species in the low thousands, while condoning for humans a population in the billions, raises the question 'how many humans can, in fairness to other species, live on earth?' (Mathews, 2016: 142).

From this perspective, prioritizing AI over humans (and their rights) seems 'interesting'. Yet, if this prioritization is to satisfy mere economic interests, it may be problematic; if this emphasis is because AI may resemble human mind and, thus, 'must deserve' some particular attention, an anthropo-like-centric view, then it may be hardly surprising –it would most probably be following the *trend* to prioritize humans or those who (aim to) 'think alike'; and, if this is because AI may serve human needs but also desires –the will *to know* what will be–, it could be a cause for concern.

Regardless of whether AI is prioritized for profit or its 'human-like mind', for people's wills or their needs, such emphasis and focus can often be unfair, unjust or unethical. People have to think both seriously and creatively. Fair and proportionate approaches may not always or in any scenario enable people to understand, scrutinize or control everything; nor would they *per se* guarantee absolute accuracy. But they would require that processes be structured to foster appropriate control and reasonably accurate determinations (Perry, 1977: 392); they could work as a tool for achieving long lasting improvements (Behrens, 2005: 6) and, thus, address important aspects of automated decision making. Based on a *free* and *willing* consensus, built upon *common understanding*, such strategies could allow for sharing good practice, for incentivizing changes in behaviors (Behrens, 2005: 6). Instead of supporting solely the regime of *control-by-risk*, where numbers are aimed at reflecting people's *internal thoughts and ideas* to forecast scenarios and judge them, there could be a shift toward, or some particular focus on, a concept where one's *actions* would be regarded as key determinants of her identity and character; a notion where choosing *what to do* would be seen as choosing *who to be*(come) (Lebacqz, 1985: 83; Johannesen, 1988: 62).

Some have argued that if people were able to control everything they would not even wake up in the morning (Kim & Routledge, 2018). This may be true and in compliance with working *with* the idea that the future is unknown. People may never achieve ultimate control over everything, always and everywhere –such control may not be desirable.

But an *appropriate* and *adequate* level of control is desired and could be reached. People, working on a commonly understood basis, could take their *responsibility to discern rightness* that many theories of ethics aim to study. Under a fair and proportionate scheme, one who *controls* her behavior in a *suitable* sense (Fischer & Ravizza, 2000: 12-14) could then be *rightly* subject to reactive attitudes (Oshana, 2002: 263; Strawson, 1993); she could, thus, be rightly held responsible. Such scheme might be a way to bridge responsibility *gaps* (Mittelstadt, Allo, Taddeo, Wachter, & Floridi, 2016: 11) emerging from complete absence of control over algorithms' behavior.

In the past, courts rightly suggested that holding a firm responsible for selling cellular phones to '*teenage punks*' who use the phones to set up drug deals would be unreasonable (U.S. Court of Appeals, 1992). Today, at risk of 'falling' into responsibility *gaps*, claims suggest that those *using* automated systems be held responsible for decisions made, even if they cannot explain (and, perhaps, *control*) the relevant process (Association for Computing Machinery, 2017). Instead of constantly looking for someone on whom to pin the blame, attention could be drawn to strategies that would ensure appropriate control over procedures and behaviors; policies that would guarantee there would be the *prerequisite* for rightly holding someone responsible; schemes that would promote a heightened sense of responsibility or accountability (Perry, 1977: 428).

This article has argued that there may be a number of fundamental principles, on which people could agree, to address algorithmic uncertainties. Strategies to effectively exercise rights related to these principles could range from a due process approach and a 'right to a fair, accurate and unbiased decision maker' to proportionality tests and investing in 'tracking the trackers', in favoring the rights of the judged over the judger (Bambauer & Zarsky, 2018: 35). People could build upon existing principles and mechanisms. They could reach a free and willing consensus that would be the criterion of proportionality (Perry, 1977: 407); an agreement that would respect diversity.

As noted above, each culture, community or family may have its own 'rules, god or time'. But they do have some kind of, or something similar to, 'rules, god or time'; those without 'rules' have something similar to what others call 'rules'; those without 'god' have something to believe in, something close to what is referred to as 'god'; and those without time, if any, know that something happened and expect something to be. If people cannot agree on 'time', they could agree on 'day', 'week' or 'past, present and future'.

The perfect ethical theory that would address any scenario may not come soon. Instead of waiting for it, there could be some consensus, some experimenting in fair processes, before we come to realize that muddling through would have been better than not acting at all.

References

A29WP. (1998). Working Party on the protection of individuals with regard to the processing of personal data. Transfers of personal data to third countries: Applying Articles 25 and 26 of the EU Data Protection Directive (adopted on 24 July 1998; DG XV D/5025/98; WP 12). Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/1998/wp12_en.pdf

A29WP. (2017). Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679 (adopted on 3 October 2017; 17/EN; WP 251).

Adlassnig, K. P. (1986). Fuzzy Set Theory in Medical Diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics, 16,* 260-265. https://doi.org/10.1109/TSMC.1986.4308946

Allnutt, H., Dickson, C., & Webster, R. (2016, September 1). In depth analysis: Automated decision making, connectivity and ethics. *DACBeachcroft*. Retrieved from https://www.dacbeachcroft.com/es/gb/articles/2016/september/in-depth-analysis-automated-decision-making-connectivity-and-ethics/

Anderson, E. S., & Pildes, R. H. (2000). Expressive Theories of Law: A General Restatement. *University of Pennsylvania Law Review, 148*, 1503-1575. https://doi.org/10.2307/3312748

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias - There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*. Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Asaro, P. (2000). Transforming society by transforming technology: The science and politics of participatory design. *Accounting, Management and Information Technologies, 10,* 257-290. https://doi.org/10.1016/S0959-8022(00)00004-7

Association for Computing Machinery. (2017). Statement on algorithmic transparency and accountability. USA: ACM US Public Policy Council. Retrieved from https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

Azencott, C.-A. (2018). Machine Learning and Genomics: Precision Medicine versus Patient Privacy. *Philosophical Transactions of the Royal Society A, 376*. https://doi.org/10.1098/rsta.2017.0350

Baier, K. E. (1948). Objectivity in Ethics. *Australasian Journal of Philosophy, 26,* 147-165. https://doi.org/10.1080/00048404808541243

Bambauer, J., & Zarsky, T. (2018). The Algorithm Game. *Notre Dame Law Review, 94,* 1-48.

Barocas, S., & Selbst, A. (2016). Big Data's Disparate Impact. *California Law Review, 104,* 671-732. https://doi.org/10.2139/ssrn.2477899

Bartow, A. (2006). A Feeling of Unease about Privacy Law. *University of Pennsylvania Law Review*, *155,* 52-62. Retrieved from https://scholars.unh.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1119&context=law_facpub

Bates, D., Cohen, M., Leape, L., Overhage, M., Shabot, M., & Sheridan, T. (2001). Reducing the Frequency of Errors in Medicine Using Information Technology. *Journal of the American Medical Informatics Association, 8,* 299-308. https://doi.org/10.1136/jamia.2001.0080299

Battelle, J. (2005). *The search: How Google and its rivals rewrote the rules of business and transformed our culture*. USA: Penguin Group.

Beer, D. (2017). The Social Power of Algorithms. *Information, Communication & Society, 20,* 1-13. https://doi.org/10.1080/1369118X.2016.1216147

Behrens, R. (2005). Ethics and Proportionality: Getting the Balance Right. *Teaching Public Administration, 25,* 3-8. https://doi.org/10.1177/014473940502500102

Bendel, O. (2016). Considerations about the Relationship between Animal and Machine Ethics. *AI & Society, 31,* 103-108. Retrieved from https://doi-org.proxy.uba.uva.nl:2443/10.1007/s00146-013-0526-3

Benkler, Y. (2013). System and conscience: NSA bulk surveillance and the problem of freedom. *Lecture at the Center for Research on Computation and Society of Harvard University* – 2013, December 4 [Online Lecture]. Retrieved from https://privacytools.seas.harvard.edu/presentations/yochai-benkler-system-and-conscience-nsa-bulk-surveillance-and-problem

Berardi, F. (2014). *And. Phenomenology of the end: Cognition and sensibility in the transition from conjunctive to connective mode of social communication*. Unigrafa, Finland: Aalto University Publication Series.

Bhargava, R., Deahl, E., Letouzé, E., Noonan, A., Sangokoya, D., & Shoup, N. (2015). *Beyond data literacy: Reinventing community engagement and empowerment in the age of data*. Data-Pop Alliance White Paper Series. Data-Pop Alliance (Harvard Humanitarian Initiative, MIT Media Lab and Overseas Development Institute) and Internews. Retrieved from http://datapopalliance.org/wp-content/uploads/2015/11/Beyond-Data-Literacy-2015.pdf

Birnbacher, D., & Birnbacher, W. (2017). Fully Autonomous Driving: Where Technology and Ethics Meet. *IEEE Intelligent Systems, 32*, 3-4. https://doi.org/10.1109/MIS.2017.3711644

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The Social Dilemma of Autonomous Vehicles. *Science, 352,* 1573-1576. https://doi.org/10.1126/science.aaf2654

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. UK: Oxford University Press.

Boyd, D., & Crawford, K. (2011). Six provocations for Big Data. A decade in Internet time: Symposium on the dynamics of the Internet and society. Oxford Internet Institute. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431

Bravetti, A., & Padilla, P. (2018). An Optimal Strategy to Solve the Prisoner's Dilemma. *Scientific Reports, 8*, 1-6. https://doi.org/10.1038/s41598-018-20426-w

Burrell, J. (2016). How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data & Society, 3,* 1-12. https://doi.org/10.1177/2053951715622512

Caliskan, A., Bryson, J., & Narayanan, A. (2017). Semantics Derived Automatically from Language Corpora Contain Human-like Biases. *Science, 356,* 183-186. https://doi.org/10.1126/science.aal4230

Capron, A. M. (1991). Protection of Research Subjects: Do Special Rules Apply in Epidemiology?. *Law, Medicine & Health Care*, *44*, 81-89. https://doi.org/10.1016/0895-4356(91)90180-H

Castell, S. (2018). The Future Decisions of RoboJudge HHJ Arthur Ian Blockchain: Dread, Delight or Derision?. *Computer Law & Security Review, 34,* 739-753. https://doi.org/10.1016/j.clsr.2018.05.011

Castelvecchi, D. (2016). The Black Box of AI. *Nature, 538,* 21-23. https://doi.org/10.1038/538020a

Cate, F. (2006). The failure of fair information practice principles. In J. K. Winn (ed.), *Consumer protection in the age of the 'Information Economy'* (pp. 341-378). Great Britain: Ashgate Publishing.

Chalmers, D., Davies, G., & Monti, G. (2010). *European Union law: Cases and materials* (2nd ed.). New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511841408

Chamberlayne, R., Green, B., Barer, M. L., Hertzman, C., Lawrence, W. J., & Sheps, S. B. (1998). Creating a Population-Based Linked Health Database: A New Resource for Health Services Research. *Revue Canadienne de Santé Publique*, *89,* 270-273. Retrieved from http://journal.cpha.ca/index.php/cjph/article/view/1111/1111

Chander, A. (2017). The Racist Algorithm?. *Michigan Law Review, 115,* 1023-1045. Retrieved from https://repository.law.umich.edu/cgi/viewcontent.cgi?article=1657&context=mlr

Chartier, T. (2016). Vertigo over the Seven V's of Big Data. *The Journal of Corporate Accounting & Finance, 27,* 81-82. https://doi.org/10.1002/jcaf.22145

Chen, C. L. P., & Zhang, C.-Y. (2014). Data-Intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data. *Information Sciences, 275,* 314-347. https://doi.org/10.1016/j.ins.2014.01.015

Chen, H., Chiang, R., & Storey, V. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly, 36,* 1165-1188. https://doi.org/10.2307/41703503

Chesterman, S. (2017, September 2). Privacy and our digital selves. *The Straits Times*. Retrieved from https://www.straitstimes.com/opinion/privacy-and-our-digital-selves

Cheyne, I., & Alder, J. (2007). Environmental Ethics and Proportionality: Hunting for a Balance. *Environmental Law Review, 9,* 171-189. http://dx.doi.org/10.1350/enlr.2007.9.3.171

Chopra, S., & White, L. F. (2011). *A legal theory for autonomous artificial agents*. USA: University of Michigan Press. https://doi.org/10.3998/mpub.356801

Citron, D. K. (2008). Technological Due Process. *Washington University Law Review, 85,* 1249-1313. Retrieved from https://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=1166&context=law_lawreview

Citron, D. K., & Gray, D. (2013). Addressing the Harm of Total Surveillance: A Reply to Professor Neil Richards. *Harvard Law Review Forum, 126,* 262-274. Retrieved from https://harvardlawreview.org/2013/06/addressing-the-harm-of-total-surveillance-a-reply-to-professor-neil-richards/

Citron, D. K., & Pasquale, F. (2014). The Scored Society: Due Process for Automated Predictions. *Washington Law Review, 89,* 1-33. https://heinonline.org/HOL/P?h=hein.journals/washlr89&i=8

Competitoor. (2018). Ebay real-time price monitoring & tracking. Retrieved from https://competitoor.com/ebay-real-time-price-monitoring-tracking/

Competitor Monitor. (2018). Price monitoring software to help you sell better online. Retrieved from https://www.competitormonitor.com/

Conejar, R. J., & Kim, H.-K. (2014). A medical Decision Support System (DSS) for Ubiquitous Healthcare Diagnosis System. *International Journal of Software Engineering and its Applications, 8,* 237-244. https://pdfs.semanticscholar.org/8b63/976d3dad27f93e2c8302992725339635cec6.pdf

Cooley, D. R. (2013). "A Kantian Care Ethics Suicide Duty". *International Journal of Law and Psychiatry, 36,* 366-373.

https://doi.org/10.1016/j.ijlp.2013.06.001

Cooper, J. (2006). Procedural Due Process, Human Rights and the Added Value of the Right to a Fair Trial. *Judicial Review, 11,* 78-91. https://doi.org/10.1080/10854681.2006.11426463

Cooper, W. (2003). *The evolution of reason: Logic as a branch of biology*. New York, USA: Cambridge University Press.

Court of Justice. (1956). Judgment of the Court of 16 July 1956. Case 8/55 (ECLI:EU:C:1956:7). *Fédération Charbonnière de Belgique v High Authority of the European Coal and Steel Community*. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/d1c1196f-2248-4b88-93e4-6c4cf337073e/language-en

Crawford, K. (2013). The Hidden Biases in Big Data. *Harvard Business Review* (April 1, 2013). Retrieved from https://hbr.org/2013/04/the-hidden-biases-in-big-data

Crawford, K., & Schultz, J. (2014). Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms. *Boston College Law Review, 55,* 93-128. Retrieved from https://lawdigitalcommons.bc.edu/cgi/viewcontent.cgi?article=3351&context=bclr

Davidson, M. D. (2013). On the Relation Between Ecosystem Services, Intrinsic Value, Existence Value and Economic Valuation. *Ecological Economics, 95,* 171-177. https://doi.org/10.1016/j.ecolecon.2013.09.002

Deng, B. (2015). The Robot's Dilemma: Working out how to build ethical robots is one of the thorniest challenges in artificial intelligence. *Nature, 523,* 25-26. Retrieved from https://www.nature.com/news/machine-ethics-the-robot-s-dilemma-1.17881

Dickson, B. (2017, May 12). What is narrow, general and super artificial intelligence. *TechTalks.* Retrieved from https://bdtechtalks.com/2017/05/12/what-is-narrow-general-and-super-artificial-intelligence/

Directive 2016/943. (2016). Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure.

Diver, L., & Schafer, B. (2017). Opening the Black Box: Petri Nets and Privacy by Design. *International Review of Law, Computers & Technology, 31,* 68-90. https://doi.org/10.1080/13600869.2017.1275123

Dixon, H. (2018). Regulate to Liberate: Can Europe Save the Internet? *Foreign Affairs, 97,* 28-32. Retrieved from http://proxy.uba.uva.nl:2048/docview/2094378245?accountid=14615

Dixon, P., & Gellman, R. (2014). The scoring of America: How secret consumer scores threaten your privacy and your future. World Privacy Forum. Retrieved from https://www.worldprivacyforum.org/2014/04/wpf-report-the-scoring-of-america-how-secret-consumer-scores-threaten-your-privacy-and-your-future/

Doak, D. F., Bakker, V. J., Goldstein, B. E., & Hale, B. (2014). What is the future of conservation? *Trends in Ecology & Evolution, 29,* 77-81. https://doi.org/10.1016/j.tree.2013.10.013

Dove, E. S., & Özdemir, V. (2015). What Role for Law, Human Rights, and Bioethics in an Age of Big Data, Consortia Science, and Consortia Ethics? The Importance of Trustworthiness. *Laws, 4,* 515-540. https://doi.org/10.3390/laws4030515

Dove, E. S., Knoppers, B. M., & Zawati, M. H. (2014). Towards an Ethics Safe Harbor for Global Biomedical Research. *Journal of Law and the Biosciences, 1,* 3-51. https://doi.org/10.1093/jlb/lst002

Dreisbach, C. (2011). Vicious Duty: The Ethics of Osama Bin Laden. *Think, 10,* 29-39. https://doi.org/10.1017/S1477175611000066

Durand, D. (1941). *Risk elements in consumer instalment financing*. USA: National Bureau of Economic Research. Retrieved from https://EconPapers.repec.org/RePEc:nbr:nberbk:dura41-1

Edelman, B., & Luca, M. (2014). Digital discrimination: The case of Airbnb.com. Harvard Business School. NOM Unit Working Paper No. 14-054. http://dx.doi.org/10.2139/ssrn.2377353

EDPS. (2018). European Data Protection Supervisor. Opinion 5/2018 - Preliminary Opinion on privacy by design (31 May 2018). Retrieved from https://edps.europa.eu/sites/edp/files/publication/18-05-31_preliminary_opinion_on_privacy_by_design_en_0.pdf

Edwards, L., & Veale, M. (2017). Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For. *Duke Law & Technology Review, 16,* 18-84. http://dx.doi.org/10.2139/ssrn.2972855

Edwards, L., & Veale, M. (2018). Enslaving the Algorithm: From a “Right to an Explanation” to a “Right to Better Decisions”?. *IEEE Security & Privacy, 16,* 46-54. https://doi.org/10.1109/MSP.2018.2701152

Emanuel, E., & Wertheimer, A. (2006). Who Should Get Influenza Vaccine When Not All Can? *Science, 312,* 854-855. https://doi.org/10.1126/science.1125347

Esposito, E. (2015). Beyond the Promise of Security: Uncertainty as Resource. *Telos, 170,* 89-107. https://doi.org/10.3817/0315170089

European Commission. (2010). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - A digital agenda for Europe (19 May 2010; COM(2010)245 final). Retrieved from https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:FIN:EN:PDF

European Commission. (2013). H2020-EU.3.5.4. - Enabling the transition towards a green economy and society through eco-innovation (L 347 of 2013-12-11). Retrieved from https://cordis.europa.eu/programme/rcn/664417_en.html

European Commission. (2015, December 15). Agreement on Commission's EU data protection reform will boost Digital Single Market. Retrieved from http://europa.eu/rapid/%20press-release_IP-15-6321_en.htm

European Consumer Consultative Group. (2018). Policy recommendations for a safe and secure use of artificial intelligence, automated decision-making, robotics and connected devices in a modern consumer world. Opinion 16 May 2018. Retrieved from https://ec.europa.eu/info/sites/info/files/eccg-recommendation-on-ai_may2018_en.pdf

European Union Agency for Fundamental Rights. (2018). Handbook on European non-discrimination law. Retrieved from fra.europa.eu/en/publication/2018/handbook-european-law-non-discrimination

EWHC. (2002). Ms B v An NHS Hospital Trust [2002] EWHC 429 (Fam) (22nd March, 2002). Retrieved from http://www.bailii.org/ew/cases/EWHC/Fam/2002/429.html

Fearn, N. (2018, May 10). Best speech to text app of 2018. *TechRadar*. Retrieved from https://www.techradar.com/news/best-speech-to-text-app

Ferguson, A. G. (2017a). Policing Predictive Policing. *Washington University Law Review, 94,* 1109-1189. Retrieved from https://openscholarship.wustl.edu/law_lawreview/vol94/iss5/5/

Ferguson, A. G. (2017b). *The rise of big data policing: Surveillance, race, and the future of law enforcement*. New York: New York University Press. https://doi.org/10.2307/j.ctt1pwtb27

First, D. (2018). Will Big Data Algorithms Dismantle the Foundations of Liberalism? *AI & Society, 33,* 545-555. https://doi.org/10.1007/s00146-017-0733-4

Fischer, J. M., & Ravizza, M. (2000). *Responsibility and control: A theory of moral responsibility* (paperback ed.). USA: Cambridge University Press.

Flanagan, M., Howe, D. C., & Nissenbaum, H. (2008). Embodying values in technology: Theory and practice. In J. (van der) Hoven & J. Weckert, J. (eds), *Information technology and moral philosophy* (pp. 322-353). Cambridge: Cambridge University Press.

Förster, K., & Weish, U. (2017). Advertising critique: Themes, actors and challenges in a digital age. In G. Siegert, M. B. (von) Rimscha, S. Grubenmann (eds), *Commercial communication in the digital age - Information or disinformation?.* de Gruyter GmbH. Retrieved from https://www.degruyter.com/view/product/455036

Foster, C., & Miola, J. (2015). Who’s in Charge? The Relationship Between Medical Law, Medical Ethics, and Medical Morality? *Medical Law Review, 23,* 505-530. https://doi.org/10.1093/medlaw/fwv004

Freitas, (De) J. (2012, May 29). Why is memory so good and so bad? Explaining the memory paradox. *Scientific American*. Retrieved from https://www.scientificamerican.com/article/why-memory-so-good-bad/

Friedman, B., & Kahn, P. H. (2003). Human values, ethics, and design. In J. A. Jacko & A. Sears (eds), *The human-computer interaction handbook* (pp. 1177-1201). Mahwah: Lawrence Erlbaum Associates. Retrieved from https://brandorn.com/img/writing/tech-ethics/human-values-ethics-and-design.pdf

Friedman, B., & Nissenbaum, H. (1996). Bias in Computer Systems. *ACM Transactions on Information Systems (TOIS), 14,* 330-347. https://doi.org/10.1145/230538.230561

Friedman, L. (2016). Law, Technology, and the Butterfly Effect. *Revista Electrónica Direito e Sociedade - REDES, 4,* 103-122. Retrieved from https://revistas.unilasalle.edu.br/index.php/redes/article/view/2318-8081.16.14/pdf

FTC. (2000). Federal Trade Commission. Privacy online: Fair information practices in the electronic marketplace - A report to Congress. Retrieved from https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-fair-information-practices-electronic-marketplace-federal-trade-commission-report/privacy2000.pdf

FTC. (2016). Federal Trade Commission. Big Data: A tool for inclusion or exclusion? Understanding the issues. FTC Report (January 2016). Retrieved from https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf

Gaggioli, G. (2018). International Humanitarian Law: The Legal Framework for Humanitarian Forensic Action. *Forensic Science International, 282,* 184-194. https://doi.org/10.1016/j.forsciint.2017.10.035

Galinsky, K. (1998). *Augustan culture: An interpretive introduction.* Princeton, New Jersey: Princeton University Press.

GDPR. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC.

Geist, E. M. (2016). It's Already too Late to Stop the AI Arms Race - We Must Manage it Instead. *Bulletin of the Atomic Scientists, 72,* 318-321. https://doi.org/10.1080/00963402.2016.1216672

Gleser, M. A., & Collen, M. F. (1972). Towards Automated Medical Decisions. *Computers and Biomedical Research, 5,* 180-189. https://doi.org/10.1016/0010-4809(72)90080-8

González-Ballester, M.-Á. (2018). Machine learning in healthcare and computer-assisted treatment. In E. Gómez (ed.), *Assessing the impact of machine intelligence on human behaviour: An interdisciplinary endeavour* (pp. 45-47). Luxembourg: Publications Office of the European Union; Seville: European Commission.

Gostin, L., & Hadley, J. (1998). Health Services Research: Public Benefits, Personal Privacy, and Proprietary Interests. *Annals of Internal Medicine, 129,* 833-835. https://doi.org/10.7326/0003-4819-129-10-199811150-00017

Grau, C. (2006). There Is No "I" in "Robot": Robots and Utilitarianism. *IEEE Intelligent Systems, 21,* 52-55. https://doi.org/10.1109/MIS.2006.81

Gunter, K. (2000). Computerized Credit Scoring's Effect on the Lending Industry. *North Carolina Banking Institute, 4,* 443-474. Retrieved from https://scholarship.law.unc.edu/cgi/viewcontent.cgi?referer=https://www.google.be/&httpsredir=1&article=1080&context=ncbi

Guyer, P., & Wood, A. W. (eds). (1998). *The Cambridge edition of the works of Immanuel Kant: Critique of pure reason*. USA: Cambridge University Press.

Harcourt, B. E. (2005). Against prediction: Sentencing, policing, and punishing in an actuarial age. University of Chicago Public Law & Legal Theory. Working Paper No. 94. Retrieved from https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1021&context=public_law_and_legal_theory

Harris, M. (2008/2009). Consequentialism, Deontologism, and the Case of Sheva ben Bikhri. *The Torah U-Madda Journal, 15,* 68-94. Retrieved from https://www.jstor.org/stable/40914728

Harvard Law Review. (2017). State v. Loomis - Criminal law - Sentencing Guidelines - Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing - 881 N.W.2d 749 (Wis. 2016). *Harvard Law Review, 130,* 1530-1537. Retrieved from https://harvardlawreview.org/2017/03/state-v-loomis/

Hastie, T., Tibshirani, R., & Friedman J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Berlin, Germany: Springer. https://doi.org/10.1007/978-0-387-84858-7

Hauer, T. (2018). Society and the Second Age of Machines: Algorithms Versus Ethics. *Society, 55,* 100-106. https://doi.org/10.1007/s12115-018-0221-6

Hernández, G. I. (2014). *The international court of justice and the judicial function*. USA: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199646630.001.0001

Hilts, P. J. (1995, January 15). Consensus on ethics in research is elusive. *The New York Times,* p. 24. Retrieved from http://proxy.uba.uva.nl:2048/docview/109441522?accountid=14615

Hui, L., Li, S., & Zongfang, Z. (2013). The Model and Empirical Research of Application Scoring Based on Data Mining Methods. *Procedia Computer Science, 17,* 911-918. https://doi.org/10.1016/j.procs.2013.05.116

Hunt, L. W. (2014). The Global Ethics of Helping and Harming. *Human Rights Quarterly, 36,* 798-819.

https://doi.org/10.1353/hrq.2014.0062

Hunt, T., & Tirpok, A. (1993). Universal Ethics Code: An Idea whose Time has Come. *Public Relations Review, 19,* 1-11. https://doi.org/10.1016/0363-8111(93)90025-8

Informatics Europe & EUACM. (2018). When computers decide: European recommendations on machine-learned automated decision making. Retrieved from https://dl-acm-org.proxy.uba.uva.nl:2443/citation.cfm?id=3185595

International Coffee Organization. (2007). International Coffee Agreement. www.ico.org/ica2007.asp; http://www.ico.org/ica2007.asp; http://www.ico.org/show_faq.asp?show=4

Ioannidis, J. (2005). Why Most Published Research Findings Are False. *PLoS Medicine, 2,* 0696-0701. https://doi.org/10.1371/journal.pmed.0020124

Jandoli, V., & Dani, M. (2019). Trade Secrets in Italy: The Reform. *European Intellectual Property Review, 41,* 46-55.

Janssen, M., Voort, (van der) H., & Wahyudi, A. (2017). Factors Influencing Big Data Decision-Making Quality. *Journal of Business Research, 70,* 338–345. https://doi.org/10.1016/j.jbusres.2016.08.007

Johannesen, R. L. (1988). What should we teach about formal codes of communication ethics? *Journal of Mass Media Ethics, 3,* 59-64. https://doi.org/10.1080/08900528809358310

Jordan, M. I., & Mitchell, T. M. (2015). Machine Learning: Trends, Perspectives, and Prospects. *Science, 349,* 255-260. https://doi.org/10.1126/science.aaa8415

Kahneman, D. (2011). *Thinking fast and slow*. New York: Farrar Straus Giroux.

Katsafanas, P. (2018). Nietzsche's Account of Self-Conscious Agency. *Philosophical Explorations, 21,* 122-137. https://doi.org/10.1080/13869795.2017.1421695

Katz, D. M. (2013). Quantitative Legal Prediction-or-How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry. *Emory Law Journal, 62,* 909-966. Retrieved from http://law.emory.edu/elj/_documents/volumes/62/4/contents/katz.pdf

Kim, H. S. (2018). The Role of Legal and Moral Norms to Regulate the Behavior of Texting while Driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 52,* 21-31. https://doi.org/10.1016/j.trf.2017.11.004

Kim, T. W., & Routledge, B. R. (2018). Informational Privacy, A Right to Explanation, and Interpretable AI. *2018 IEEE Symposium on Privacy-Aware Computing* (pp. 64-74). Washington, DC, USA: IEEE. Retrieved from https://ieeexplore.ieee.org/document/8511831

King, A. G., & Mrkonich, M. (2016). "Big Data" and the Risk of Employment Discrimination. *Oklahoma Law Review, 68,* 555-584. Retrieved from https://digitalcommons.law.ou.edu/olr/vol68/iss3/3/

Knapp, C. (2007). Equality and Proportionality. *Canadian Journal of Philosophy, 37,* 179-201. https://doi.org/10.1353/cjp.2007.0017

Korn, D. (2000). Contribution of the human tissue archive to the advancement of medical knowledge and the public health. In *Research involving human biological materials: Ethical issues and policy guidance, Volume II, Commissioned Papers* (pp. E1-E30). Rockville, MD: U.S. National Bioethics Advisory Commission. Retrieved from https://repository.library.georgetown.edu/bitstream/handle/10822/559357/nbac_biological2.pdf?sequence=1&isAllowed=y

Kraemer, F., Overveld, (van) K., & Peterson, M. (2011). Is there an ethics of algorithms? *Ethics and Information Technology, 13,* 251-260. https://doi.org/10.1007/s10676-010-9233-7

Kraft, I. (1964). Automated Decision Making: A threat, or a Promise?. *Science, New Series, 143,* 99-100. Retrieved from http://science.sciencemag.org/content/143/3602/99.2

Kroll, J. et al. (2017). Accountable Algorithms. *University of Pennsylvania Law Review, 165,* 633-705. Retrieved from https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3/

Kruckeberg, D. (1989). The Need for an International Code of Ethics. *Public Relations Review, 15,* 6-18. https://doi.org/10.1016/S0363-8111(89)80050-5

Kruckeberg, D. (1993). Universal Ethics Code: Both Possible and Feasible. *Public Relations Review, 19,* 21-31. https://doi.org/10.1016/0363-8111(93)90027-A

Kumar, J. (2017). Ethics and legal in AI: Decision making and moral issues - A theme report based on the 2nd meeting of the all-party parliamentary group on Artificial Intelligence. UK: Big Innovation Centre. Retrieved from http://www.biginnovationcentre.com/media/uploads/pdf/ThemeReport2_27Mar.pdf

Langley, P. (2013). Anticipating Uncertainty, Reviving Risk? On the Stress Testing of Finance in Crisis. *Economy and Society, 42,* 51-73. http://dx.doi.org/10.1080/03085147.2012.686719

Larus, J., & Hankin, C. (2018). Regulating Automated Decision Making. *Communications of the ACM, 61,* 5. https://doi.org/10.1145/3231715

Latour, B. (1988). *The pasteurization of France* (translated by A. Sheridan & J. Law). Cambridge, Massachusetts; London, England: Harvard University Press.

Latzer, M., Hollnbuchner, K., Just, N., & Saurwein, F. (2016). The economics of algorithmic selection on the Internet. In J. M. Bauer & M. Latzer (eds), *Handbook on the economics of the Internet* (pp. 395-425). Edward Elgar. Retrieved from https://doi.org/10.4337/9780857939852

Lebacqz, K. (1985). *Professional ethics: Power and paradox*. Nashville: Abingdon Press.

Lipton, Z. C. (2018). The Mythos of Model Interpretability. *ACM Queue - Machine Learning, 16,* 1-28. Retrieved from https://dl.acm.org/citation.cfm?id=3236386.3241340

Lipworth, W., Mason, P., & Kerridge, I. (2017). Ethics and Epistemology of Big Data. *Bioethical Inquiry, 14,* 485-488. https://doi.org/10.1007/s11673-017-9815-8

Lisboa, P. J. G. (2013). Interpretability in machine learning – Principles and practice. In F. Masulli, G. Pasi, & R. Yager (eds), *Fuzzy logic and applications - WILF 2013: Lecture notes in computer science* (volume 8256; Springer, Cham) (pp. 15-21). https://doi.org/10.1007/978-3-319-03200-9_2

Llenas, B. (2014, February 25). Brave new world of 'predictive policing' raises specter of high-tech racial profiling. *FoxNews*. Retrieved from https://www.foxnews.com/world/brave-new-world-of-predictive-policing-raises-specter-of-high-tech-racial-profiling

Lorenc, T. (2015). Artificial Intelligence and the Ethics of Human Extinction. *Journal of Consciousness Studies*, *22*, 194-214. Retrieved from https://www.ingentaconnect.com/contentone/imp/jcs/2015/00000022/f0020009/art00012#

Malgieri, G., & Custers, B. (2018). Pricing Privacy – The Right to Know the Value of Your Personal Data. *Computer Law & Security Review, 34,* 289-303. https://doi.org/10.1016/j.clsr.2017.08.006

Malle, B. (2016). Integrating Robot Ethics and Machine Morality: The Study and Design of Moral Competence in Robots. *Ethics and Information Technology, 18,* 243-256. Retrieved from https://link.springer.com/article/10.1007/s10676-015-9367-8

Malle, B., Scheutz, M., Forlizzi, J., & Voiklis, J. (2016). Which Robot Am I Thinking About? The Impact of Action and Appearance on People's Evaluations of a Moral Robot. *11th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 125-132). New York: IEEE. Retrieved from https://ieeexplore-ieee-org.proxy.uba.uva.nl:2443/document/7451743

Malmir, B., Amini, M., & Chang, S. (2017). A Medical Decision Support System for Disease Diagnosis under Uncertainty. *Expert Systems with Applications, 88,* 95-108. https://doi.org/10.1007/s10676-015-9367-8

Mander, J. (1978). *Four arguments for the elimination of television*. New York: William Morrow.

Manders-Huits, N. (2011). What Values in Design? The Challenge of Incorporating Moral Values into Design. *Science and Engineering Ethics, 17,* 271-287. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3124645/#

Mantelero, A. (2016). Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection. *Computer Law & Security Review, 32,* 238-255. https://doi.org/10.1016/j.clsr.2016.01.014

Mantelero, A. (2018). AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment. *Computer Law & Security Review, 34,* 754-772. https://doi.org/10.1016/j.clsr.2018.05.017

Markoff, J. (2015). *Machines of loving grace: The quest for common ground between humans and robots*. New York: ECCO.

Mathews, F. (2016). From Biodiversity-Based Conservation to an Ethic of Bio-Proportionality. *Biological Conservation, 200,* 140-148. https://doi.org/10.1016/j.biocon.2016.05.037

Matthewman, W. (1984). Title VII and Genetic Testing: Can Your Genes Screen You out of a Job. *Howard Law Journal, 27,* 1185-1220. Retrieved from https://heinonline.org/HOL/P?h=hein.journals/howlj27&i=1197

Matthias, A. (2004). The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata. *Ethics and Information Technology, 6,* 175-183. https://doi.org/10.1007/s10676-004-3422-1

Mayer-Schonberger, V., & Cukier, K. (2014). *Big Data: A revolution that will transform how we live, work, and think*. Boston, New York: Eamon Dolan Mariner Books.

McManus-Degnan, M. (2010). No Actual Bias Needed: The Intersection of Due Process and Statutory Recusal. *Temple Law Review, 83,* 225-252. Retrieved from https://heinonline.org/HOL/P?h=hein.journals/temple83&i=227

Melé, D., & Sánchez-Runde, C. (2013). Cultural Diversity and Universal Ethics in a Global World. *Journal of Business Ethics, 116,* 681-687. https://doi.org/10.1007/s10551-013-1814-z

Meron, T. (2005). Judicial Independence and Impartiality in International Criminal Tribunals. *The American Journal of International Law, 99,* 359-369. https://doi.org/10.2307/1562502

Merrill, J. C. (2011). Overview: Theoretical foundations for media ethics. In A. D. Gordon, J. M. Kittross, J. C. Merrill, W. A. Babcock, & M. Dorsher (eds), *Controversies in media ethics* (3rd ed.) (pp. 3-32). New York, London: Routledge. Retrieved from http://cw.routledge.com/textbooks/9780415963329/data/overview.pdf

Metz, C. (2016, September 6). Self-driving cars will teach themselves to save lives—but also take them. *Wired*. Retrieved from https://www.wired.com/2016/06/self-driving-cars-will-power-kill-wont-conscience/

Meyer, E. (2014, December 24). Inadvertent algorithmic cruelty. Meyerweb. Retrieved from https://meyerweb.com/eric/thoughts/2014/12/24/inadvertent-algorithmic-cruelty/

Mileham, P. (2015). Part 1: Human Conflict and Universal Ethics. *Defense & Security Analysis, 31,* 348-355. https://doi.org/10.1080/14751798.2015.1101221

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The Ethics of Algorithms: Mapping the Debate. *Big Data & Society, 3,* 1-21. https://doi.org/10.1177/2053951716679679

Mugarura, N. (2016). Law as an Equalizer: “Contemporary Global Crises” and its Implication on the Global Village. *International Journal of Law and Management, 58,* 602-617. https://doi.org/10.1108/IJLMA-07-2015-0043

Mutlu, C. E. (2015). Of Algorithms, Data and Ethics: A Response to Andrew Bennett. *Journal of International Studies, 43,* 998-1002. https://doi.org/10.1177/0305829815581536

Neimanis, J., & Matjusina, R. (2011). Judge Impartiality in Comprehensive Judicial Development. *European Integration Studies, 2011,* 87-91. http://dx.doi.org/10.5755/j01.eis.0.5.1081

Nottelmann, N. (2008). Introduction. *Synthese, 161,* 325-337. https://doi.org/10.1007/s11229-006-9087-z

O’Leary, R. (2017). The 2016 John Gaus Lecture: The New Guerrilla Government: Are Big Data, Hyper Social Media and Contracting Out Changing the Ethics of Dissent? *Political Science and Politics, 50,* 12-22. https://doi.org/10.1017/S1049096516002018

O’Malley, M. (2014). Doing What Works: Governing in the Age of Big Data. *Public Administration Review, 74,* 555-556. https://doi.org/10.1111/puar.12260

O’Neil, K. (2016). *Weapons of math destruction - How Big Data increases inequality and threatens democracy*. New York: Broadway Books.

OECD. (1980). Recommendation of the Council of 23rd September 1980. Guidelines governing the protection of privacy and transborder flows of personal data (C(80)58/FINAL).

Oshana, M. A. L. (2002). The Misguided Marriage of Responsibility and Autonomy. *The Journal of Ethics, 6,* 261-280. https://doi.org/10.1023/A:1019482607923

Otto, S. (2007). Nursing Ethics: Request an Ethics Consult to Resolve Dilemmas. *Nursing Critical Care, 2,* 18-20. https://doi.org/10.1097/01.CCN.0000281587.54995.dc

Papadimos, T. et al. (2018). Ethics of Outbreaks Position Statement. Part 1: Therapies, Treatment Limitations, and Duty to Treat. *Critical Care Medicine, 46,* 1842-1855. https://doi.org/10.1097/CCM.0000000000003416

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Cambridge, Massachusetts; London, England: Harvard University Press. https://doi.org/10.4159/harvard.9780674736061

Pearsall, B. (2010). Predictive Policing: The Future of Law Enforcement? *NIJ Journal, 266,* 16-19. Retrieved from https://www.ncjrs.gov/pdffiles1/nij/230414.pdf

Perry, M. J. (1977). Constitutional "Fairness": Notes on Equal Protection and Due Process. *Virginia Law Review, 63,* 383-430. https://doi.org/10.2307/1072356

Podesta, J., Pritzker, P., Moniz, E., Holdren, J., & Zients, J. (2014). Big Data: Seizing opportunities, preserving values. Report, May 1, 2014; The White House. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

Polansky, R. (ed.). (2014). *The Cambridge companion to Aristotle's Nicomachean ethics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CCO9781139022484

Poole, T. (1998). Procedural Fairness in Dismissals: Towards a 'Due Process' Approach? *Industrial Law Journal, 27,* 233-237. https://doi.org/10.1093/ilj/27.3.233

Poon, M. (2009). From New Deal Institutions to Capital Markets: Commercial Consumer Risk Scores and the Making of Subprime Mortgage Finance. *Accounting, Organizations and Society, 34,* 654-674. https://doi.org/10.1016/j.aos.2009.02.003

Popper, P. (2012, October 22). Rapture of the nerds: Will the Singularity turn us into gods or end the human race? *The Verge*. Retrieved from https://www.theverge.com/2012/10/22/3535518/singularity-rapture-of-the-nerds-gods-end-human-race

Preece, A. (2018). Asking 'Why' in AI: Explainability of Intelligent Systems – Perspectives and Challenges. *Intelligent Systems in Accounting, Finance and Management, 25,* 63-72. https://doi.org/10.1002/isaf.1422

Price, A. W. (1989). *Love and friendship in Plato and Aristotle*. New York: Oxford University Press. Retrieved from www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198248996.001.0001/acprof-9780198248996

Prisync. (2018). Competitor price tracker. Retrieved from https://prisync.com/

Purtova, N. (2018). The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law. *Law, Innovation and Technology, 10,* 40-81. https://doi.org/10.1080/17579961.2018.1452176

Raskin, V., & Taylor, R. J. (2016). Towards a Semantic Theory of Robotic Non-Cooperation. *Procedia Computer Science, 94,* 392-397. https://doi.org/10.1016/j.procs.2016.08.060

Rawls, J. (1987). The Idea of an Overlapping Consensus. *Oxford Journal of Legal Studies, 7,* 1-25. https://doi.org/10.1093/ojls/7.1.1

Rawls, J. (2009). *A theory of justice* (revised ed.). Cambridge, Massachusetts: Harvard University Press.

Ray, T. (1996). Differentiating the Related Concepts of Ethics, Morality, Law, and Justice. *New Directions for Teaching and Learning, 1996,* 47-53. https://doi.org/10.1002/tl.37219966609

Reddix-Smalls, B. (2011). Credit Scoring and Trade Secrecy: An Algorithmic Quagmire or How the Lack of Transparency in Complex Financial Models Scuttled the Finance Market. *UC Davis Business Law Journal, 12,* 87-123. Retrieved from https://heinonline.org/HOL/P?h=hein.journals/ucdbulj12&i=93

Redish, M. H., & Marshall, L. C. (1986). Adjudicatory Independence and the Values of Procedural Due Process. *Yale Law Journal, 95,* 455-505. https://doi.org/10.2307/796487

Remoff, H. (2016). Malthus, Darwin, and the Descent of Economics. *American Journal of Economics and Sociology, 75,* 862-903. https://doi.org/10.1111/ajes.12158

Richards, N. M., & King, J. H. (2016). Big Data and the future for privacy - Handbook of research on digital transformations. *SSRN*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2512069

Ritchie, R., Kessler, D., & Sargent, M. (2018, July 21). Best smart lights for your home in 2018. *iMore*. Retrieved from https://www.imore.com/best-smart-lights

Robinson, D., Yu, H., & Rieke, A. (2014). Civil rights, big data, and our algorithmic future. A September 2014 report on social justice and technology by Robinson + Yu. Retrieved from https://centerformediajustice.org/wp-content/uploads/2014/10/Civil-Rights_Big-Data_Our-Future.pdf

Rubinstein, I. (2013). Big Data: The End of Privacy or a New Beginning?. *International Data Privacy Law, 3,* 74-87. https://doi.org/10.1093/idpl/ips036

Rudin, P. (2016, December 2). Decision making: Performed by humans or machines? *Singularity 2030*. Retrieved from https://singularity2030.ch/decision-making-performed-by-humans-or-machines/

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development, 3,* 211-229. https://doi.org/10.1147/rd.33.0210

Schuemie, M., Ryan, P., Hripcsak, G., Madigan, D., & Suchard, M. (2018). Improving Reproducibility by Using High-Throughput Observational Studies with Empirical Calibration. *Philosophical Transactions of the Royal Society A, 376*. https://doi.org/10.1098/rsta.2017.0356

Schumacher, E. F. (1973). *Small is beautiful: Economics as if people mattered*. London: Blond & Briggs. Retrieved from http://www.daastol.com/books/Schumacher%20(1973)%20Small%20is%20Beautiful.pdf

Schütze, R. (2015). *An introduction to European law* (2nd ed.). UK: Cambridge University Press. https://doi.org/10.1017/CBO9781316278314

Schwartz, P. (1999). Privacy and Democracy in Cyberspace. *Vanderbilt Law Review, 52,* 1609-1701. Retrieved from http://www.paulschwartz.net/pdf/VAND-SCHWARTZ.pdf

Shankar, R. D., & Musen, M. A. (1999). Justification of Automated Decision-Making: Medical Explanations as Medical Arguments. *Proceedings of the AMIA Symposium, 1999,* 395-399. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2232527/

Sheppard, S. R. J., & Cizek, P. (2009). The ethics of Google Earth: Crossing thresholds from Spatial Data to Landscape Visualization. *Journal of Environmental Management, 90,* 2102-2117. https://doi.org/10.1016/j.jenvman.2007.09.012

Silver, D. et al. (2017). Mastering the Game of Go without Human Knowledge. *Nature, 550,* 354-359. https://doi.org/10.1038/nature24270

Simon, H. A. (1960). *The new science of management decision*. New York: Harper & Brothers Publishers. https://doi.org/10.1037/13978-000

Singhapakdi, A., Vitell, S. J., Rao, C. P., & Kurtz, D. L. (1999). Ethics Gap: Comparing Marketers with Consumers on Important Determinants of Ethical Decision-Making. *Journal of Business Ethics, 21,* 317-328. https://doi.org/10.1023/A:1005965230277

Sintchenko, V., & Coiera, E. W. (2003). Which Clinical Decisions Benefit from Automation? A Task Complexity Approach. *International Journal of Medical Informatics, 30,* 309-316. https://doi.org/10.1016/S1386-5056(03)00040-6

Skinner-Thompson, S. (2018). *AIDS and the law* (5th ed.). New York: Wolters Kluwer Law & Business.

Slowther, A. (2002). The Case of Ms B and the “Right to Die”. *Journal of Medical Ethics, 28,* 243. https://doi.org/10.1136/jme.28.4.243

Smart-Plant. (2018). Scale-up of low-carbon footprint MAterial Recovery Techniques in existing wastewater treatment PLANTs. Retrieved from http://www.smart-plant.eu/

Smith, A. (2005). *The theory of moral sentiments.* Sao Paolo, Brasil: MetaLibri.

Stallman, R. (2010). *Free as in freedom (2.0): Richard Stallman and the free software revolution* (2nd ed.). USA: Free Software Foundation.

Stallman, R. (2015). *Free software, free society: Selected essays of Richard M. Stallman* (3rd ed.). Free Software Foundation. Retrieved from https://www.gnu.org/doc/fsfs3-hardcover.pdf

Strawson, P. (1993). Freedom and resentment. In J. M. Fischer & M. Ravizza (eds), *Perspectives on moral responsibility* (pp. 45-66). USA: Cornell University Press.

Summers, R. S. (1974). Evaluating and Improving Legal Processes - A Plea for Process Values. *Cornell Law Review, 60*, 1-52. Retrieved from https://scholarship.law.cornell.edu/cgi/viewcontent.cgi?article=3965&context=clr

Supreme Court of Illinois. (1966). 34 Ill. 2d 266 (1966), 215 N.E.2d 286, *Motorola, Inc. v. Illinois Fair Employment Practices Commission* (No. 39297). Retrieved from https://law.justia.com/cases/illinois/supreme-court/1966/39297-5.html

Sutton, A. (2010). Ethics and Law Teaching and Learning in Undergraduate Medicine. *Journal of Medical Ethics, 36,* 511. https://doi.org/10.1136/jme.2010.036350

Taylor, J. (2015). Mapping Human Understanding to Robotic Perception. *Procedia Computer Science, 56,* 514-519. https://doi.org/10.1016/j.procs.2015.07.244

Tene, O. (2007). What Google knows: Privacy and Internet search engines. *SSRN*. https://doi.org/10.2139/ssrn.1021490

The Economist. (2017, May 6). The world’s most valuable resource. *The Economist, May 6th-12th 2017,* 7. Retrieved from https://dl.cainiao.io/file/The-Economist/2017-05-06/The%20Economist%2020170506%20859d1e.pdf

Thomas, W. C. (1896). *Cosmic ethics, or the mathematical theory of evolution*. London: Smith, Elder and Co.

U.S. Court of Appeals. (1992). US Court of Appeals for the Seventh Circuit - 970 F.2d 283 (7th Cir. 1992) *United States of America v. Robert Blankenship and Thomas E. Lawrence* (argued: January 7, 1992; decided: July 21, 1992). Retrieved from https://law.justia.com/cases/federal/appellate-courts/F2/970/283/269857/

U.S. Department of Health, Education and Welfare. (1973). Records, computers, and the rights of citizens - Report of the secretary's advisory committee on automated personal data systems. Washington, D.C.: DHEW Publication. Retrieved from https://www.justice.gov/opcl/docs/rec-com-rights.pdf

United Nations Environment Programme. (2018). *Handbook for the Montreal Protocol on substances that deplete the Ozone Layer* (12th ed.). Ozone Secretariat. Retrieved from http://ozone.unep.org/sites/default/files/MP_handbook-english-2018.pdf

United Nations. (1921). International Convention for the Suppression of the Traffic in Women and Children (concluded at Geneva on 30 September 1921). Retrieved from https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=VII-2&chapter=7&clang=_en

United Nations. (1931/1946). Agreement concerning the Suppression of Opium Smoking (Bangkok, 27 November 1931; Lake Success, New York, 11 December 1946). Retrieved from https://treaties.un.org/Pages/ViewDetails.aspx?src=TREATY&mtdsg_no=VI-9&chapter=6&clang=_en

United Nations. (1946). Constitution of the World Health Organization (New York, 22 July 1946). Retrieved from https://treaties.un.org/doc/Treaties/1948/04/19480407%2010-51%20PM/Ch_IX_01p.pdf

United Nations. (1947). General Agreement on Tariffs and Trade (Geneva, 30 October 1947). Retrieved from https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=X-1-a&chapter=10&clang=_en

United Nations. (1948). Convention on the Prevention and Punishment of the Crime of Genocide (Paris, 9 December 1948). Retrieved from https://treaties.un.org/doc/publication/unts/volume%2078/volume-78-i-1021-english.pdf

United Nations. (1949a). Agreement providing for the provisional application of the Draft International Customs Conventions on Touring, on Commercial Road Vehicles and on the International Transport of Goods by Road (Geneva, 16 June 1949). Retrieved from https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XI-A-1&chapter=11&clang=_en

United Nations. (1949b). Agreement for Facilitating the International Circulation of Visual and Auditory Materials of an Educational, Scientific and Cultural Character (Lake Success, New York, 15 July 1949). Retrieved from https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XIV-1&chapter=14&clang=_en

United Nations. (1951). Convention relating to the Status of Refugees (Geneva, 28 July 1951). Retrieved from https://www.unhcr.org/1951-refugee-convention.html

United Nations. (1961). Vienna Convention on Diplomatic Relations (Vienna, 18 April 1961). Retrieved from http://legal.un.org/ilc/texts/instruments/english/conventions/9_1_1961.pdf

United Nations. (1966a). International Covenant on Economic, Social and Cultural Rights (New York, 16 December 1966). Retrieved from https://treaties.un.org/Pages/ViewDetails.aspx?src=IND&mtdsg_no=IV-3&chapter=4&clang=_en

United Nations. (1966b). International Covenant on Civil and Political Rights (New York, 19 December 1966). Retrieved from https://treaties.un.org/doc/publication/unts/volume%20999/volume-999-i-14668-english.pdf

United Nations. (1968). International Coffee Agreement. Retrieved from https://treaties.un.org/doc/Publication/UNTS/Volume%20893/volume-893-A-9262-English.pdf

United Nations. (1985). Vienna Convention for the Protection of the Ozone Layer (Vienna, 22 March 1985). Retrieved from https://treaties.un.org/Pages/ViewDetails.aspx?src=IND&mtdsg_no=XXVII-2&chapter=27&clang=_en

United Nations. (1989). Convention on the Rights of the Child (20 November 1989). Retrieved from https://www.ohchr.org/en/professionalinterest/pages/crc.aspx

United Nations. (1990). International Convention on the Protection of the Rights of All Migrant Workers and Members of their Families (New York, 18 December 1990). Retrieved from https://treaties.un.org/Pages/ViewDetails.aspx?src=IND&mtdsg_no=IV-13&chapter=4&clang=_en

United Nations. (1992a). United Nations Framework Convention on Climate Change (FCCC/INFORMAL/84; GE.05-62220 (E) 200705). Retrieved from https://unfccc.int/resource/docs/convkp/conveng.pdf

United Nations. (1992b). International Sugar Agreement. Retrieved from http://ec.europa.eu/world/agreements/prepareCreateTreatiesWorkspace/treatiesGeneralData.do?step=0&redirect=true&treatyId=542

United Nations. (1992c). Convention on biological diversity. Retrieved from https://www.cbd.int/doc/legal/cbd-en.pdf

United Nations. (1994). UN Convention to Combat Desertification. Retrieved from https://www.unccd.int/

United Nations. (2005). Convention on the Use of Electronic Communications in International Contracts (New York, 23 November 2005). Retrieved from https://www.uncitral.org/pdf/english/texts/electcom/06-57452_Ebook.pdf

United Nations. (2006). Convention on the Rights of Persons with Disabilities (New York, 13 December 2006). Retrieved from https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html

United Nations. (2018). Multilateral Treaties Deposited with the Secretary-General. Retrieved from https://treaties.un.org/

Universal Declaration of Human Rights. (1948). United Nations General Assembly. Universal Declaration of Human Rights (Paris, 10 December 1948). Retrieved from http://www.un.org/en/universal-declaration-human-rights/

Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine Learning in Medicine: Addressing Ethical Challenges. *PLoS Medicine, 15,* e1002689. Retrieved from https://doi.org/10.1371/journal.pmed.1002689

Veale, M., & Binns, R. (2017). Fairer Machine Learning in the Real World: Mitigating Discrimination without Collecting Sensitive Data. *Big Data & Society, 4,* 1-17. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/2053951717743530

Veale, M., & Edwards, L. (2018). Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling. *Computer Law & Security Review, 34,* 398-404. https://doi.org/10.1016/j.clsr.2017.12.002

Vogel, E. K. (2008, November 4). Why do we forget things?. *Scientific American*. Retrieved from https://www.scientificamerican.com/article/why-do-we-forget-things/

Vöneky, S., & Wolfrum, R. (eds). (2013). *Human dignity and human cloning*. Springer Science and Business Media.

Voort, (van der) H., Klievink, B., Arnaboldi, M., & Meijer, A. (2018). Rationality and Politics of Algorithms. Will the Promise of Big Data Survive the Dynamics of Public Decision Making?. *Government Information Quarterly, In Press,* 1-12. https://doi.org/10.1016/j.giq.2018.10.011

Wang, H., Xu, Z., Hamido, F., & Liu, S. (2016). Towards Felicitous Decision Making: An Overview on Challenges and Trends of Big Data. *Information Sciences, 367/368,* 747-765. https://doi.org/10.1016/j.ins.2016.07.007

Ware, O. (2015). Accessing the Moral Law through Feeling. *Kantian Review, 20,* 301-311. https://doi.org/10.1017/S1369415415000060/

Whittaker, T. (1916). *The theory of abstract ethics*. Cambridge: Cambridge University Press; New York: Putnam’s Sons.

Wijnberg, N. M. (2000). Normative Stakeholder Theory and Aristotle: The Link Between Ethics and Politics. *Journal of Business Ethics, 25,* 329-342. Retrieved from https://doi-org.proxy.uba.uva.nl:2443/10.1023/A:1006086226794

Winner, L. (1980). Do Artifacts Have Politics?. *Daedalus, 109,* 121-136. Retrieved from https://www.jstor.org/stable/20024652

Wood, A. W. (ed.). (2018). *(Immanuel Kant) Groundwork for the metaphysics of morals*. USA: Yale University Press.

World Medical Association. (1964). WMA Declaration of Helsinki – Ethical principles for medical research involving human subjects. Retrieved from https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/

Wright, C. F., Hurles, M. E., & Firth, H. V. (2016). Principle of Proportionality in Genomic Data Sharing. *Nature Reviews: Genetics, 17,* 1-2. https://doi.org/10.1038/nrg.2015.5

Ziewitz, M. (2016). Governing Algorithms: Myth, Mess, and Methods. *Science, Technology, & Human Values, 41,* 3-16. https://doi.org/10.1177/0162243915608948

Žliobaitė, I. (2017). Measuring Discrimination in Algorithmic Decision Making. *Data Mining and Knowledge Discovery, 31,* 1060-1089. Retrieved from https://doi-org.proxy.uba.uva.nl:2443/10.1007/s10618-017-0506-1