Research article
DOI: https://doi.org/10.21202/jdtl.2023.16

Algorithmic Discrimination and Privacy Protection

Elena Falletti

Università Carlo Cattaneo – LIUc
Castellanza, Italy

Keywords

Algorithm,
artificial intelligence,
data protection,
digital technologies,
discrimination,
law,
personal data,
privacy,
private life,
regulation

Abstract

Objective: emergence of digital technologies such as Artificial intelligence became a challenge for states across the world. It brought many risks of the violations of human rights, including right to privacy and the dignity of the person. That is why it is highly relevant to research in this area. That is why this article aims to analyse the role played by algorithms in discriminatory cases. It focuses on how algorithms may implement biased decisions using personal data. This analysis helps assess how the Artificial Intelligence Act proposal can regulate the matter to prevent the discriminatory effects of using algorithms.

Methods: the methods used were empirical and comparative analysis. Comparative analysis allowed to compare regulation of and provisions of Artificial Intelligence Act proposal. Empirical analysis allowed to analyse existing cases that demonstrate us algorithmic discrimination.

Results: the study's results show that the Artificial Intelligence Act needs to be revised because it remains on a definitional level and needs to be sufficiently empirical. Author offers the ideas of how to improve it to make more empirical.

Scientific novelty: the innovation granted by this contribution concerns the multidisciplinary study between discrimination, data protection and impact on empirical reality in the sphere of algorithmic discrimination and privacy protection.

Practical significance: the beneficial impact of the article is to focus on the fact that algorithms obey instructions that are given based on the data that feeds them. Lacking abductive capabilities, algorithms merely act as obedient executors of the orders. Results of the research can be used as a basis for further research in this area as well as in law-making process.

For citation

Falletti, E. (2023). Algorithmic Discrimination and Privacy Protection. *Journal of Digital Technologies and Law, 1*(2), 387–420. https://doi.org/10.21202/jdtl.2023.16

Contents

Introduction

The 18th-century Industrial Revolution was the precursor to social change and claimed rights through automation in the production process. It fostered the advent of goods manufacturing by creating places dedicated to the production, shifting activity from the countryside to urban centres. It brought about a social transformation of society by displacing the masses into an alienating reality: the factories erased mentalities, behaviours, and customs linked to the previous agricultural life rooted in the culture and collective memory (Tunzelmann, 2003; Freeman et al., 2001).

From the perspective of the relationship between law and technology, this change has had a hard legal and social impact since technological evolution promotes a unifying thrust by standardising the rules of conduct of the new production process. Such a transition represents a shift towards an articulated system of norms, which, in the function of technical precepts, obliges each subject involved to homologate their behaviour according to technical standards that take on legal value.

This context allows us to make a historically diachronic comparison between the automation introduced in the 19th century in British factories and today's algorithmic automation. This comparison concerns the measurement of time during work performance. As has been the case in the past, the first area where innovation in human activity manifests itself and develops is the field of work. This circumstance applies first and foremost

to automation: on the one hand, the performance of a job, a biblical punishment, allows individuals to live an independent life; on the other, automating work allows entrepreneurs and producers to have a series of savings, to the detriment of workers, whose role is comparable to that of a mere machine (Marx, 2016). The Marxist curse of 'alienated labour' is also realised (Marx, 2016; Claeys, 2018).

The measurement of labour time in order to cut costs is one of the most significant concerns when managing an organisation so complex as a business: the automation of the production process in the 19th century extended the workday beyond daylight hours (thanks to electricity) and detached it from weather conditions, separating labour from agricultural rhythms, whereas algorithmic automation creates an even greater extreme since the algorithm can manage production processes without human control, achieving the kind of alienation evoked by Marx of the individual worker estranged from other workers (Marx, 2016).

By applying Marxist theory to the length of working days (Marx, 2016; Claeys, 2018), illo tempore as well as now, a paradoxical effect can be observed: on the one hand, capitalists expand their profits by increasing both the labour capacity of workers and the surplus value of production; however, this method results in marginal returns, as the lengthening of shifts exhausts employees. Such is true concerning the manufacturing capitalist and the digital platform, which hires workers to perform algorithms-managed tasks. In both situations, the increase in surplus value produced by the engaged labour force involves the physical control of workers. While in the past, said control was made problematic because, unlike other production factors, the power of labour is (or was) constituted by its embodiment in people, workers who, in turn, maintain (or maintained) their power to resist being treated as commodities (Woodcock, 2020; Altenried, 2020).

In the «Digital Revolution era», such correspondence between physicality and work is severely undermined by the solitude of work performance. Indeed, the worker has the platform managed by algorithms as their only reference. They are isolated from other colleagues and have as one means of a direct relationship with the app that allows interaction between them and the algorithm. Algorithmic discrimination in contemporary workplaces sees workers deprived of proper legal and wage recognition, and it enables the perpetuation of categorical and legal distinctions that reflect those related to work (Fuchs, 2014).

The Industrial Revolution changed the paradigm of equality and, mirroring it, that of discrimination since the advent of capitalism had brought about a change in social reality and lifestyles (Schiavone, 2019). In philosophical discourse, equality acquired a new centrality that broke away from the religious and philosophical perspective and took on the social and political points of view embodied in the great revolutions of the time, namely the American and French (Schiavone, 2019).

Equality, inequality, and, therefore, discrimination moved out of the formal condition of only slavery, which still existed, and entered the social facts about which a broader public opinion began to form beyond that of the old aristocratic and religious orders. People began

to reflect that discrimination and poverty are related to technologies that shattered social categories and weakened them.

The shattering and weakening of social categories are happening again today. Indeed, technology seems to increase discrimination against the most vulnerable, and it appears to do so because it allows for the emergence of «expansive elements of impersonal equality concerning any individual or gender difference» (Schiavone, 2019).

There would then be a little effect according to which equality would turn into a form of impersonality since it would be necessary to ignore, or even nullify, the individual characteristics that make each person unrepeatable (Ainis, 2015).

However, discriminatory cases could occur in this context of homogeneous serialisation of individuals, particularly in the automated processing of personal data. In this sense, automated decision-making algorithms find the justification of efficiency in their use precisely in serialising issues and overlapping facts. In fact, in the perspective of algorithmic automation, a kind of indifference, i.e., a pretended neutrality, on the possible origin of inequality, whether physical or moral, is envisioned because the nomenclature of choices elaborated upstream by the author of the algorithm concerns cataloguing, and thus a classification based on the processing of concepts. This procedure allows for a choice during the formation of the algorithm as to how to classify each element of the procedure based on the purposes to which the algorithm is directed, i.e., the beliefs of the programmer. Such an operation cannot be neutral but follows criteria for categorisation according to the intended purpose. Since this process is not neutral, nor can it be since it is choice-based, it contains potential discrimination.

On this issue, scholars have investigated whether and how it is possible to adopt measures of different elements related to fairness, and they are specifically examining the mathematical relationships between them (Garg et al., 2020; Pessach & Shmueli, 2020; Krawiec et al., 2023). For example, group calibration, positive and/or class balancing show that «except in narrow specific cases", no conditions can simultaneously satisfy the situations dealt with in the experiment (Kleinberg et al., 2016; Pleiss et al., 2017; Abdollahpouri et al., 2020). Other studies have addressed related issues on incompatibilities between fairness criteria, which in specific contexts (such as psychological tests applied to recidivism) can lead to significant indirect discrimination when recidivism differs among the groups examined (Chouldechova, 2017; Lippert-Rasmussen, 2022). Likewise, in criminal risk analysis, it has been observed that it is «generally impossible to maximise accuracy and fairness simultaneously, and it is impossible to satisfy all types of fairness simultaneously» (Berk et al., 2021).

Under the condition of technological transformation, as was the case in the mid-18th century, and now as well, the discrimination origin could be discerned due to the vulnerability of people involved in the automation process. Indeed, there are similarities between the early Industrial Revolution and contemporary artificial intelligence use.

In both cases, there is a subjugation of the vulnerable classes who cannot (and/or were not) escape such a technological paradigm shift. It is a condition that occurs due to belonging to the vulnerable classes, and, at the same time, it is the same vulnerability that forces such weak parties to be subaltern.

In Western Legal Tradition countries, particularly in the United States, this double connotation is most profound for those suffering from the nefarious effects related to the slave heritage, which is firmly rooted in that social reality, despite various attempts to overcome this legacy. Such endeavours have remained unsatisfactory since, in large parts of society, there is significant racial discrimination introjected by ADMs, particularly in areas such as risk assessment software.

1. Discrimination and personal data treatment

As far as privacy is concerned, it should be regarded as the manifestation of the individual's right to personal integrity, both physical and mental, against influences from third parties, whether physical subjects, legal entities, or the state itself. Privacy has become the bulwark of personality protection, both on- and offline. It is a shield against the reconstruction of individuality by third parties, public or private, as a result of the tracking of data that each person leaves behind during his/her day through geolocation carried out by the apps on his/her smartphone, payments made, and the sharing of materials and places. Privacy defends individual identity and reputation (thanks to the affirmation of the right to be forgotten, in which privacy is an element) as much in real personal life as in virtual.

In this perspective, the role played by discrimination is more insidious and difficult to demonstrate, and it concerns its manipulative aspect: from the moment that black boxes collect personal data on a massive scale, they shatter (Messinetti, 2019; Vale et al., 2022) the personality and identity of each individual by effectively annihilating them into a mass of information; by this very fact, they accomplish an operation of manipulation of individual data that, depending on the perspective of observation, entails two opposite results.

On the one hand, if each black box is formed by entering the data collected as neutrally as possible, it reflects social reality. Therefore, like a mirror, the same discriminatory processes result in reporting our society's distortions.

Instead, on the other hand, if the input of data into the black box is cleansed of discriminatory content, it could represent a distorted and manipulated image of the same reality, which returns a black box that adheres to the ideals of those who collected the data. However, this could be even more dangerous because it no longer depicts reality but a view of the same, which, in addition to being biased and unreal, is also tied to a value judgment oriented to the purposes of those who manage the black box and its results.

Therefore, the formation of the black box, which is the basis for elaborate automated decision-making processes, must be carried out with maximum transparency and attention to the objectives of the promoters. These are processes in which the multidisciplinary skills of data scientists, mathematicians, philosophers, and jurists are needed, especially those who turn to the study of comparative systems to learn and understand guidelines and results obtained in other legal systems that have dealt with the subject.

Although the protection of the right to privacy is traditionally focused on state and government authorities, the invasiveness of private entities focused on the lucrative interest provided by the commercial exploitation of their users' profiles is steadily growing (Zuboff, 2019; York, 2022) and emerges in situations that may be only seemingly unexpected. The confrontation between technological evolution and personal protection has resulted in four different mutations:

(a) on the one hand, the need to ensure the protection of the sphere of privacy, already acquired by the emancipation of the individual from central powers (Bellamy, 2014), of all citizens of industrialised societies (Rodotà, 2012), both individually and as a mass (Canetti, 1960);

(b) this need is consequent to the introduction of the electronic processor and telematic transmission of information related to the individual in the production process;

(c) the collection and storage of such data by private entities represents a paradigm shift (Kuhn, 1962) that has a legal interest (Rodotà, 1995; Alpa & Resta, 2006; Angiolini, 2020);

d) the latter activity allows for the profiling of users, which is also a harbinger of a paradigm shift related to the subject of law that is pitted against the entity, a legal person endowed with such technological power (Pellecchia, 2018).

Using computers and data telematic transmission combination has allowed for storing large amounts of data through their storage in magnetic tapes and, later on, in increasingly sophisticated digital media. With this transformation, private entities have joined the state in the massive collection of individuals' data, commonly called «big data» (Mayer-Schonberger & Cukier, 2013). Such collection has evolved by disseminating more advanced, refined, complex, and, therefore, intrusive technological tools among consumers.

It involves the collection of personal data from locations such as the Internet (the digital platforms and social networks), mobile devices (smartphones, tablets, smartwatches), satellite geolocators, radio frequency identifiers (RFID), sensors, cameras, or other tools that are capable of implementing other new emerging technologies (Gressel et al., 2020).

The right to privacy manifests itself as negative freedom, that is, in not being subjected to interference in one's private life by persons or entities, whether private or institutional. However, the right to personal data protection is embodied in the positive freedom to exercise control over the processing and to circulate of information about one's person (Rodotà, 2014).

What is legally evaluable in a «negative» sense (protection of privacy toward external interference) or positive way (protection of personal data through control over information) is that a black box takes on an indistinct connotation. Indeed, what is processed

is the information extracted from subjects during their lifetime, regardless of its relation to judgmental classifications of various conceptual natures.

Imagining being inside a black box, immersed in a virtual place where the algorithmic nature of automated decisions takes shape through matrix calculations, one realises that it is impossible to follow the logical path of the data. Such operations feed on the knowledge at their disposal (Cockburn et al., 2018; Bodei, 2019), without dealing with its provenance or origins.

Suppose it is true that this massive amount of data («big data,» precisely) represents the knowledge used by black boxes; it becomes essential to identify who has the authority to manage it and the power to dispose of it. The questions referred to can be summarised as follows:

a) Who knows what? It is a question related to the distribution of knowledge and whether barriers can legitimately be placed on access to such information sources.

b) Who decides what is possible to know? (and, therefore, how the collected data can be accessed). This question concerns the performance of the role of each subject part of the information chain: the subject «source» of the data (considered an individual even though the data collected relevant to this individual is processed massively), the institutions in charge of controlling the use of the data themselves (such as the National Privacy Authorities), as well as the operators who draw their black boxes from such data.

c) Who decides who decides? Who exercises the power of control over the sharing or subtraction of collected and knowable data?

The element that serves as the fulcrum of the balance concerning each issue concerns the effective implementation of privacy protection. For example, regarding question (a), privacy, viewed as a barrier to the intrusiveness of others by private and public administration entities into the private and personal sphere, is a limitation on the possibility of collection of discriminatory elements, and thus the continuation of discrimination itself, at the time of black box formation.

About question (b), regarding which authority can allow for the massive collection of data, the roles of responsibility and guarantee must be played, on the one hand, by the Data Protection Officers of each entity, either public or private, involved in the collection and processing of data, respectively. On the other hand, the role of guarantors of fairness is played by European and national institutions.

Finally, concerning question (c) as to who has the power to determine who the decision-makers are, it seems to be appropriate that this role should be played by the state, understood as the representative body of the consociates, subject to the rule of law and the principle of separation of powers. However, it is self-evident that the power achieved by platforms in data management represents a factual monopoly of problematic management by nation-states related to the asserted extraterritoriality of economic entities to apply national law (Tarrant & Cowen, 2022; Witt, 2022; Parona, 2021).

It is a context provoked by the interaction of subjects in a stateless society, such as the Internet. On the one hand, there is no substantial difference between social and political relations as the individual user considers him/herself as the unit of measure of his/her world. On the other hand, it dissolves in the information magma present in online life.

On the contrary, dictating the rules through their contractual conditions are non-state entities that can apply unilaterally established sanctions without counterbalance and according to their discretion. In this context, the data collection process is influenced by discriminatory elements, especially if illicitly collected, leading to biased, distorted, and thus non-neutral results, which are products of illegitimate and factually wrong decisions.

In a sense, a study of black boxes and the discrimination absorbed in them could be said to be a mirror in which reality itself is reflected. Such a mirror is helpful, especially from a legal point of view, to find tools to remedy it.

In this area, case law is given a definite role in resolving the juxtaposition between the need for public safety and privacy protection.

2. Bias and discrimination in automated decision-making systems

In computer science, «bias» refers to a length bias in bits transmission. In legal-computer science, this term refers to discriminatory situations on the part of algorithmic models, which «may lead to the detection of false positives (or negatives) and, consequently, produce discriminatory effects to the detriment of certain categories of individuals» (Parona, 2021). Such bias may depend on the set of training data, the relevance or accuracy of the data, and the types of algorithms used (Hildebrandt, 2021) concerning the fineness and speed of the results. From the perspective of social reality, bias-related discrimination may refer to unfair treatment or illegal discrimination. On this point, it is crucial to distinguish computer bias from the impact of unfair or unjust bias that consists of illegal discrimination, depending on how the collected data are formed and how they interact with each other and, simultaneously, with the surrounding reality.

It is noted in the doctrine that the interacting biases in automated decisions are mainly three (Hildebrandt, 2021):

(a) The first concerns the machine learning posed by machine learning algorithms. It is an inductive and unavoidable bias that, although neither positive nor negative in itself, cannot be considered neutral concerning the reality in which it interacts.

(b) The second concerns the ethically problematic bias because it allows for the distribution of goods, services, risks, opportunities, or access to information configured in ways that may be morally problematic. Some examples could involve excluding people, pushing them in a particular direction, or toward specific behaviours.

(c) The third type of bias is the most obvious, even to the less observant eye. It could be based on illegitimate situations or behaviours, i.e. when the machine learning algorithm focuses on individuals or categories of subjects based on illegitimate and discriminatory motives.

It has been discussed whether bias (c) may involve a subset of ethical biases (Hildebrandt, 2021). Indeed, discrimination based on gender is illegal, but not everyone considers it unethical, such as charging male drivers a higher insurance premium. After all, they are considered more reckless drivers, while female drivers appear more cautious. Such discrimination, while illegal, is not necessarily an ethical problem.

Biases (a) and (b) can relate to behaviours observable by sensors or online tracking systems or infer by the same automated algorithm. In observation, bias affects the training data, while bias affects the system's output in inference. In both cases, the output (i.e., the result) is affected, so it lacks neutrality concerning the reality to which it relates. The use of machine learning (Hildebrandt, 2021) inevitably produces bias since, as in the case of human cognition and perception are already characterised, those of machine learning are not objective, contrary to what one might think. This situation calls for caution and critical capacity, especially when the cognitive results of machine learning appear reliable (Hildebrandt, 2021).

One of the seminal texts on the subject, namely «Machine Learning» written by Tom Mitchell (Mitchell, 2007; Kubat & Kubat, 2017) makes explicit that bias, understood as variance, is necessary to demonstrate the importance and usefulness of refined tests. From these, it is possible to derive the realisation that bias, whether it consists of variance or bias, can cause errors (Hildebrandt, 2021), which can become embedded in the various stages of the decision-making process. The latter may include steps such as collecting or classifying «training data» until the goal of automated decision-making is achieved. Such errors may consist in the translation of factual contingencies into the programming language or in the circumstance that the source data are themselves incomplete (technically defined as «low hanging fruit»), however, without their incompleteness being apparent, and this may be caused by the context in which the machine learning program is operating, i.e., on a simulated or actual model (Hildebrandt, 2021).

The amount of data used by the machine learning program could cause errors since this software could process «spurious» correlations or patterns because of biases, understood in the sense of variance, inherent in the original data, precisely because of the reference to a certain idea of outcome toward which the creators of machine learning themselves were oriented.

However, a third situation could exist. It refers to an issue that is both fundamental and elusive and occurs when data is correctly assimilated by machine learning («ML»). However, it refers to real situations having distorting effects, such as following events that involved social developments that resulted in the exclusion of vulnerable groups, i.e., exclusions of subjects due to characteristics of a physical or behavioural nature.

In the case of erroneous or spurious raw materials, the collected data have ab origine flaws that can cause misinterpretation of correlations; the bias is rooted in real life, and thus data extraction will confirm or even reinforce existing biases.

This situation cannot be resolved by adopting ML, although some argue that ML can help highlight such bias or its causes. So, focusing the critical lens on what may cause such bias is necessary. Indeed, it should be undertaken to determine whether the data collection procedure helps the inclusion of bias in the raw materials or whether they occur in a noncausal distribution while maintaining great caution in working with ML tools at the risk of uncritically accepting their results (Hildebrandt, 2021).

The accuracy of machine learning results depends on the knowledge on which it works and interacts, despite the limitation (Marcus, & Davis, 2019; Brooks, 2017; Sunstein, 2019) and the models used.

The issue concerns the ability to understand the difference between cognitive biases present in humans (whose intelligence can adapt and make abductive and unexpected reasoning) and machine learning, whose intelligence instead always depends on the data, assumptions, and characteristics of ML feeding. The ML can process inductive and deductive inferences but not abductive ones. Indeed, abductive reasoning demands an «intuitive leap» that starts from a set of elements and, from these, elaborates an explanatory theory of these elements, which the available data must verify (Hildebrandt, 2021). To test such hypotheses, programming the machine learning model is concerned with the creative ability to recompose the abductive step to test such hypotheses inductively. If such an operational hypothesis were confirmed, the system could use the abductive method as the basis for deductive reasoning. In this regard, the experiential feedback (feedback) of machine learning is decisive as it is fundamental and crucial (Hildebrandt, 2021).

It follows from this that the quality of the samples themselves reflects the quality of the training of databases. If the user provides the system with data (or samples) that are distorted or characterised by poor quality, the behaviour of the machine learning algorithm remains negatively affected by this, producing poor-quality results (Gallese, 2022). In addition, it should be remembered that machine learning algorithms tend to lose the ability to generalise and are thus prone to exaggeration (Gallese, 2022).

This circumstance is well-known to programmers, but jurists elude it. It is identified by observing the relationship between the training and test errors. It occurs when the improvement on the error related to the training set causes a worse error on the test data set. In this case, the network on which machine learning rests proves to be «overfitting» (Gallese, 2022). It happens when the model fits the observed data (the sample) because it has too many parameters relative to the number of observations and thus loses touch with the reality of the data. From this, it can be understood that human input in the preliminary decisions on data collection, classification, and processing is inescapable and essential for obtaining a reasonable and acceptable result.

In other words: if the machine learning algorithm produces erroneous, discriminatory, or wrong results, the responsibility lies with those who organised the dataset and set up the algorithm. Machine learning obeys instructions that it merely executes.

In these situations, it could occur or be exacerbated even in the case of continuous feeding with new categories of data, leading to the system's imbalance, so some categories are more represented than others, with a significant influence on the impact on the future behaviour of that AI (Gallese, 2022). However, there is a situation in which the problem remains unsolvable: this is the case of so-called generalised class incremental learning (Gallese et al., 2020) in which the machine learning method receives new data that may, in principle, belong to new classes or cases never considered before. In this peculiar situation, the algorithm must be able to reconfigure its inner workings (e.g., in the case of the deep learning machine, where the algorithm must adapt the architecture and recalibrate all parameters) (Gallese et al., 2020). It would prevent any realistic possibility of predicting the future behaviour of the automated system.

3. The twofold nature of risk assessment algorithms

Automated decision-making algorithms highlight a relevant aspect of civil coexistence that is changing in its course and, thus, in its nature. It is the relationship between authorities (whether public or private, however able to influence people's individual and collective lives) and consociates (the people who, consciously or unconsciously, find themselves subject to or the source of approaching personal data).

Every stage or element of life (daily and in the entire existential journey) has become the object of automated collection, profiling, and decision-making. Biometric recognition algorithms investigate our identity and feelings through so-called. «affective computer learning» (Guo et al., 2020), i. e., using data on the most obvious physical characteristics (such as eye colour, complexion, hair colour) or the collection of fingerprints contained in the chips of identity identification documents (such as passports), or through gait recognition, or images capturing expression during a leisurely walk or a work activity (Crawford, 2021) or a run to catch the last subway under the prying eye of a surveillance camera.

These informational data make it possible to answer interesting questions like, who are you? or where are you going? Nevertheless, one may wonder to what extent the collection and processing of data in response to such questions are legitimate.

To develop answers to such questions, it seems helpful to examine risk assessment algorithms. They present fascinating aspects from an evolutionary point of view, as they were the first programs to be used for predicting recidivism, initially in probation and parole (Oswald et al., 2018). As pointed out in the doctrine, using predictive analytics algorithms in a judicial context can satisfy three needs, two of which are more general, especially on a cost-benefit assessment and access to public safety and legal protection resources level, and one placed on an individual level (Oswald et al., 2018).

From a justice administration efficiency perspective, specifically, the use of such software could have certain advantages, such as:

(a) facilitating overall strategic planning of forecasts and priorities in combating criminal activities;

(b) evaluating profiles of specific activities related to crime reduction;

(c) assessing the prediction of recidivism in the case of individuals.

On the latter aspect, the application of risk assessment algorithms concerns the need to balance necessity (pursued primarily by the government) with the principle of proportionality in light of respect for human rights so that the protection of the rights of the individual can be fairly balanced against the needs of the community (Oswald et al., 2018).

From a legal perspective, the major critical issues affecting this approach concern, first of all:

1. opacity and secrecy, linked first and foremost with the intellectual property protections of the algorithm itself (Oswald et al., 2018), but questionable, as contrary to the principle of transparency, especially in the case of the use of ADMs in the judicial sphere, such as the calculation of criminal recidivism;

2. the use of such automatisms does not meet the criteria of protection of principles of constitutional relevance, such as the principle of the natural judge predefined by law, the right to be heard in court and to a fair trial, the motivation of the decision (contained in Article 111 of the Italian Constitution and Article 6 of the European Convention for the Protection of Human Rights and Fundamental Freedoms) as well as, evidently, with the provision of Article 22 GDPR, which guarantees the right to the explanation of the automated decision.

On the subject of limitations on freedom, both the decision flow adopted by the algorithm and the database on which it bases its decision-making process must be transparent and accessible and the inadequacy of even one of the aforementioned constitutes a violation of the principles of due process under Articles 111 Const and 6 ECHR.

Automated decisions with inadequate results obtained through procedures that do not adhere to the principles related to respect for the protection of personal data and fundamental rights must be able to be challenged before the ordinary courts in order to obtain intelligible and adequate decisions to be understood by the person concerned. It is the right of the person subjected to the automated decision to be informed about the correctness of the decision that affects him or her, both in a formal sense (i. e., compliance with the safeguards for his or her protection prepared by the multilevel legislative framework) and in a substantive sense (i.e., why the matter was decided in the dispositive sense and how the conclusions were reached in the automatically decided matter). The logical path should not give rise to essential doubt as to whether the decision maker, in this case, the risk assessment algorithm, erred in law in making a rational decision based on relevant grounds (Oswald et al., 2018).

Is it possible to argue that ADMs require a higher decision-making standard than a human decision-maker? (Oswald et al., 2018). For example, in the legal sphere, a judicial decision

issued by a judge without a statement of reason is considered nonexistent[1], i.e., null and void[2], both under Article 132, No. 5 of the Code of Criminal Procedure, 546 of the Code of Criminal Procedure and Article 111 of the Constitution (Chizzini, 1998; Taruffo, 1975; Massa, 1990; Bargi, 1997). A judicial decision must be reasoned with a statement of the relevant facts of the case and the legal reasons for the decision (Taruffo, 1975).

However, it is noted that historical and comparative experience shows that per se, the aforementioned obligation to give reasons in fact and law is not an indispensable element of the exercise of the judicial function as both the experience of popular juries, judges of fact (Chizzini, 1998; Taruffo, 1975; Massa, 1990; Bargi, 1997), and the finding that in various systems, an obligation to give reasons is absent (Chizzini, 1998). The obligation to state reasons arises from the Jacobean Constitutions as an endoprocessual function relating to the need for control over the exercise of judicial power, consistent with the delineation of the nomofilactic role of the Supreme Court in the implementation of the principle of subordination of the judge to the law (Taruffo, 1975).

These words bring to mind the ancient Montesquiean precept that wants the judge bouche de la loi instead of bouche du roi (Petronio, 2020). In the case of the jurisprudential use of risk assessment algorithms, who fills the shoes of the «roi»? Especially in light of sensitivity, i.e., the ease with which the results of recidivism proceedings conducted through risk assessment algorithms can be manipulated. One may wonder if the algorithm cannot become the bouche de la loi. Scholars affirm that:

«to ensure that the judge, even the supreme judge, does not go his or her own way but complies with the law, which is the same for all even if it is to be applied on a case-by-case basis taking into account the multiplicity of cases to be judged, it becomes necessary for the judgment to be reasoned, that is, to give an account of why that particular solution» (Petronio, 2020).

The statement of reasons represents the manifestation of the competence, independence, and responsibility of the decisional measure issued by the judge; it is the instrument of the legal protection of the parties or the defendant against the presence of hidden factors such as bias, error, and irrationality. Against such critical issues concerning automated decision-makers, it is not possible to carry out a similar transparency operation, given the established opacity of the decisional phase.

Although in the promoters' intentions, the use of algorithms is justified by the replacement of evaluation systems in the area of bail calculation and bail granting (Israni, 2017), in reality, algorithms incorporate commonly shared stereotypes, particularly on ethnic, economic, and sexual grounds (Starr, 2014), as well as presenting ethical and constitutionality issues.

The Wisconsin Supreme Court dealt with using an algorithm developed to assist the work of judges in matters of probation and the risk of recidivism of arrested persons. The case

1 Cass., 19-7-1993, n. 8055, GC, 1993, I, 2924; Cass., 8-10-1985, n. 4881, NGL, 1986, 254.

2 Cass., 27-11-1997, n. 11975.

can be summarised as follows: in February 2013, Eric Loomis was arrested while driving a car used in a shooting. Shortly after his arrest, he pleaded guilty to contempt of court and did not contest the fact that he had taken possession of a vehicle without the owner's consent. As a result, the defendant was sentenced to six years in prison. This decision was worthy of attention because the district court used a proprietary algorithm developed by Northpointe, Inc1 in the decision-making phase of a fourth-generation intelligent risk assessment software called Correctional Offender Management Profiling for Alternative Sanctions (COMPAS). This algorithm was believed to have predicted a person's risk of recidivism based on a complex analysis involving information gathered from a survey of 137 questions divided into several sections and information corresponding to individual public criminal records (Rebitschek et al., 2021; Bao et al., 2021; Wang et al., 2022).

The Wisconsin Supreme Court held that such a tool did not present constitutionality problems about the defendant's due process if the software processed individual cases using accurate information (Israni, 2017)[3]. It should be noted that the manufacturer of the COMPAS software refused to disclose at trial the methodology and guidelines used by the software for its decisions (Custers, 2022), even though the risk assessment score developed by the algorithm was cited in the judgment, since the algorithm believed the defendant was at high risk of reoffending, the court denied him parole and handed down a six-year sentence (Israni, 2017).

The Wisconsin Supreme Court rejected the constitutionality concerns related to the violation of the defendant's due process, especially concerning the lack of transparency and accuracy of the algorithm's formulation, which would prevent the defendant from being confident of the impartiality of the decision-making process to which he is subjected.

Such verification is prevented (a usual circumstance in cases of judicial requests for access to decision-making algorithms) by the rules of intellectual property and trade secret protection as an expression of a form of the inscrutability of the algorithm (Vogel, 2020) that becomes absolute sovereign: the concretisation of the voix du roi, a king anointed by the oil of intellectual property instead of divine sanctity.

Nevertheless, in this context, intellectual property rules protect the economic interests of those who develop machine learning technology. In balancing the rights of asks to access the decisional mechanism, particularly the rights of the defence (and in certain decisional contexts, claims to due process) protected by different multilevel sources, from the Constitution to international conventions, which ones prevail?

North American jurisprudence justifies continued opacity, confirming the constitutionality of using COMPAS, despite the Wisconsin Supreme Court placing numerous restrictions on its use. The algorithm cannot be used to determine whether a detainee should be incarcerated, i.e., to calculate the length of his or her detention (Israni, 2017).

[3] State v. Loomis, 881 N.W.2d 749, 767 (Wis. 2016)

The COMPAS algorithm must be justified at sentencing and for any scoring processing on recidivism prediction, always with the caveat regarding its limited decisional utility. Because the federal Supreme Court declined to issue a writ of certiorari, the Wisconsin Supreme Court's decision is final[4].

It is the questionable factual circumstance that the Wisconsin Supreme Court has allowed an algorithm, about which there are ethical and constitutional doubts related to its non-transparency, to be placed alongside ordinary judges in exercising jurisdiction. In the U.S. court system, the protection of confidentiality in patent matters would be prioritised to maintain the patent owner's competitive advantage over individual due process rights and freedoms. According to the doctrine, this is a misunderstanding that would be difficult to overcome without the intervention of the federal Supreme Court (Israni, 2017).

Addressing the application of no less than five psychological and actuarial risk assessment tools, the Supreme Court of Canada offered a reversal of the rationale regarding whether these decision-making mechanisms can be used in the assessment of the level of risk of recidivism presented by native offenders, the case is, the Ewert v Canada decision[5]. In this case, the Supreme Court of Canada ruled that the Correctional Service of Canada (CSC) breached its statutory duty under Section 24 of the Corrections and Conditional Release Act (CCRA) for the exclusive use of accurate information in risk assessments (Scassa, 2021).

According to the Canadian Court, the Correctional Service of Canada has to account for the systemic discrimination indigenous peoples face in the criminal justice system, in general, and in prisons. However, the SCC found that despite using tools that may be discriminatory against indigenous individuals, there was no violation of the Canadian Charter of Rights and Freedoms (Russell, 1983; Epp, 1996).

This decision drew several critical comments given that it is primarily Native people, particularly women, who suffer systemic discrimination because of how the prison system is organised: they experience detention to a greater extent and for longer periods. They do not enjoy culturally appropriate programs or forms of rehabilitation that could bring Native individuals back into their communities where they might have greater support.

While the Canadian Court has recognised such discrimination, motivated by the fact that «identical treatment can lead to serious inequalities»,[6] it has been pointed out that there is a significant connection between the risk assessment tools used in prison and prisoners' freedom. An inmate's excessive risk rating significantly impacts the individual's freedom: he or she is placed in more restrictive environments, and the chances of early release are reduced. Despite all this, the Court decided that neither the person's right to life,

4 Loomis v. Wisconsin, 137 S. Ct. 2290 (2017)

5 Ewert v Canada [2018 SCC 30]. https://canliiconnects.org/en/commentaries/62360

6 *Ibid*.

liberty, and security (s. 7 of the Charter) nor the equality provisions (s. 15) had been violated. According to the Court, there was no evidence that such discrimination was absorbed by the software risk assessment algorithms and, thus, discriminated against the plaintiff.

The Supreme Court of Canada acknowledged that risk classification tools to determine prisoner recidivism are inaccurate and provided the appellant with a statement. However, it is troubling that despite a lengthy analysis of how inaccurate they are and how this may impact indigenous individuals, the Court chose not to declare them unconstitutional. It is an even more harmful and insidious decision than that of the Wisconsin Supreme Court, which,

in any case, had placed procedural limits on the use of such software. For example, some possible expedients would allow the algorithm to limit the economic-ethnic-social influence regarding the entry of the personal data of the individual being screened by the program. Furthermore, it would be possible to omit the entry of the «ZIP code», i.e., data related to the defendant's residence, since these show the potential income ranges of areas. Based on the principle of presumption of innocence, this data is irrelevant for calculating the possible risk of his recidivism, the objective of using such software.

4. The Artificial Intelligence Proposal

The Proposal for a regulation published by the European Commission on 21 April 2021 represents the first completed attempt to regulate AI in general terms, despite the context in which such a regulation would have to be regulated: on the one hand, the fragmentary nature or lightness of the regulation (especially in the United States and China, the two leading global competitors in this field), and on the other, the difficulty of foreseeing and balancing the development of a sector that is a harbinger of interests that may even be divergent ones, and subject to very rapid developments (Rosa, 2021; Alpa, 2021; Scherer, 2015). The regulation framing the development of AI must ensure the protection of fundamental rights and the rule of law while being flexible enough to adapt to technological changes that are yet to be foreseeable. In other words, the regulation is required to 'square the circle' at the national or European level and serve as a model at the global level.

The critical points concern the balance between the uniformity of the discipline and its updating, given the not remote possibility of its rapid obsolescence, due to the autonomous developments of black boxes, machine learning, deep learning and neural networks, which in turn represent a source of risks that cannot be foreseen ex-ante, in contradiction with one of the fundamental principles of the rule of law, the provision of general and abstract pro-future regulations (Scherer, 2015).

In any case, one observes the acknowledgement by the drafters of the text of the known jurisprudential experience on the subject, particularly concerning the recognition of the subordinate role of the weaker user to the role of the platforms. However, it shows a growing understanding of the discriminatory phenomena linked to automated algorithms.

The European Commission has considered such complexities related to potentially present risk factors that cannot be predicted a priori by introducing two flexibility mechanisms in the regulatory framework. This strategy is developed in three main points:

(a) The Artificial Intelligence Act proposal is accompanied by several annexes (Annexes) that form an integral part of it, characterising its discipline. These annexes, for instance, outline the categories of high-risk devices (high-risk AI systems) for which the legislation is detailed and a specific compliance procedure is envisioned. Such annexes are so relevant for the European Commission's regulatory approach to AI that the procedure for their amendment adheres to Article 290 TFEU, allowing technical regulatory standards to be approved (Battini, 2018). In order that appropriate and timely solutions can be found to the application issues related to the use of high-risk systems, according to Article 74 of the AIA, a committee is planned to intervene through amendments to the regulation, even outside the ordinary regulatory procedure required for the formal revision of the regulation (Casonato et al., 2021; Veale et al., 2021; Stuurman et al., 2022);

b) The AIA proposal itself envisions a general obligation of a five-yearly review, the first within five years of its entry into force, precisely because of the instability of the subject matter, with the consequent regulatory and legal adaptation to the evolutions it has achieved.

c) In order for the review mechanism of the regulation in question to be as thoughtful as possible, the Proposal provides in Title V for the implementation of the mechanism of so-called 'sandboxes', i.e., functional spaces set up by the Member States, for a limited period, and under the control of the national authorities, where it is possible to experiment and test innovative artificial intelligence systems with a view to their introduction on the market.

Given the combination of complexity and opaqueness in the mode of operation, significant unpredictability, and autonomy in forming results based on input data, the need to regulate artificial intelligence has become almost imperative. Such regulation must address security risks and guarantee and reinforce the protection of fundamental rights against legal uncertainty to stem the fragmentation of regulation and distrust in both the tool and the human ability to control it.

The proposed AIA regulation is part of the Union's strategy to strengthen the single digital market, which uses harmonised rules. In this case, these rules are intended to avoid fragmentation of the internal market on the essential elements concerning the requirements of products using automated algorithms to avoid legal uncertainty for both providers of such services and users of automated decision-making systems. Indeed, from the perspective of subsidiarity, if the principle of non-exclusive competence were to be strictly integrated, given that different big data sets may be incorporated in each product comprising automated systems, in this sense, a national-only approach is a harbinger of more significant, contradictory regulatory hurdles and uncertainties that would impede the circulation of goods and services, including those using automated decision-making systems.

In this sense, the AIA proposal aims to develop a legal framework adhering to the principle of proportionality that achieves its objectives by following a risk-based approach, imposing burdens only when artificial intelligence systems present high risks, i.e., outweighing the benefits, for the protection of fundamental rights and security. In order to verify such a risk and consider AI systems as not high risk, they must meet specific requirements: the data used must meet high quality, documentation, transparency, and traceability criteria.

In this regard, the instrument of the regulation was chosen because, under Article 288 TFEU, the direct applicability of regulation will reduce legal fragmentation and facilitate the development of a single market in legal, safe and reliable AI systems by introducing all EU member states a harmonised set of basic requirements for AI systems classified as high-risk and obligations for providers and users of such systems, improving the protection of fundamental rights and providing legal certainty for operators and consumers.

Regarding the protection of fundamental rights, the AIA proposal imposes certain restrictions on the freedom to conduct business and the freedom of art and science to ensure that overriding reasons of public interest, such as health, safety, consumer protection and the protection of other fundamental rights are respected when high-risk AI technology is developed and used. Such restrictions are proportionate and limited to the minimum necessary to prevent and mitigate severe risks and likely violations of fundamental rights. The use of AI with its specific characteristics (opacity, complexity, data dependency, autonomous behaviour) may adversely affect several fundamental rights enshrined in the EU Charter of Fundamental Rights. The obligation of ex-ante testing, risk management and human oversight will also facilitate the respect of other fundamental rights by minimising the risk of erroneous or biased AI-assisted decisions in critical areas such as education and training, employment, legal services, judiciary, health and welfare services.

It should be emphasised that, to the benefit of significant investments in funds, know-how and research, transparency obligations will not disproportionately affect the rights to protect intellectual property, know-how and trade secrets, and confidential information. However, this fact risks thwarting the objective of making the automated decision-making system transparent and trustworthy, as it would prevent, if not appropriately balanced, the disclosure of the way the data is processed and thus the source of possible discrimination, as happened in the litigation concerning the protection of crowd workers employed through job brokerage platforms.

As anticipated, the Proposal's approach is risk-based: it classifies artificial intelligence models as follows:

a) prohibited artificial intelligence practices insofar as they are oriented towards manipulating the conduct of individuals based on subliminal techniques or exploiting

the vulnerabilities of individuals on account of age or disability in order to influence their conduct. Also prohibited in principle are AI systems used by public authorities to establish the trustworthiness of individuals (i.e. "social scoring") (Maamar, 2018; Infantino et al., 2021) based on their social conduct and personal characteristics. However, this prohibition would seem to have taken on broad fears or apprehension from other legal systems, such as that of China, as such use of 'social scoring' models is already prohibited in Europe as violating dignity and equality.

Similarly, biometric recognition tools are prohibited, subject to a broad exception relating to their necessity for the targeted search of potential victims of criminal acts (e.g. missing children) or the prevention of a specific, substantial and imminent danger to a person's safety or of a terrorist attack, or for the detection, location, or prosecution of a person suspected of offences under Article 2(2) of Council Framework Decision 2002/584 for which the Member State concerned provides for a custodial sentence of three years or more. On this point, it is noted that these are offences for which a European arrest warrant will be issued. In any event, it is noted that the Proposal for regulation does not explicitly state anything about the possible use of such recognition systems by private entities.

b) high risk: these models' use may be permitted but must be subject to prior verification of precise requirements for protecting human dignity and respect for fundamental rights. The identification of this category is based on both the function attributed to the device and its overall purpose, including its specific purposes. For this assessment, a further evaluation of compliance with both the relevant legislation and the protection of fundamental rights is required. They cover an extensive range of tools (used in the areas specified in Annex III) and concern models used in job recruitment or diagnostic medical devices, models for biometric identification of individuals, for infrastructure management (both those used in so-called 'smart cities', such as intelligent traffic lights, and those used in the management of service supplies such as water, gas and electricity supply), education or staff training purposes, and so on.

The classification of an AI system as high-risk is based on the intended purposes of the AI system, in line with current product safety legislation. Thus, the high-risk classification depends not only on the function performed by the IA system but also on the specific purpose and manner in which the system is used. This classification has identified two categories of high-risk systems: (a) IA systems intended for use as safety components of products subject to ex-ante conformity assessment by third parties; (b) other stand-alone IA systems with implications primarily for fundamental rights that are explicitly listed in Annex II;

c) AI models with minimal or no risk are those that, although they maintain specific transparency requirements, are identified by subtraction from the previous categories, such as chatbots, whose use may be permitted but subject to information and transparency requirements on their nature, or if/then models.

Title I of the AIA proposal defines the subject matter and scope of the new rules governing the placing on the market, commissioning and use of AI systems. It also outlines the definitions used throughout the instrument, particularly under Art. 3(1) of the Proposal, an 'artificial intelligence system' (AI system) is defined as 'software developed using one or more of the techniques and approaches listed in Annex I, which can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations or decisions that influence the environments with which it interacts'. This definition of the AI system in the legal framework aims to be as technologically neutral and 'future-proof' as possible, considering the rapid technological and market developments related to AI.

In order to provide the necessary legal certainty, Title I is complemented by Annex I, which contains a detailed list of approaches and techniques for AI development to be adapted to the new technological scenario.

The key participants along the AI production and value chain are also clearly defined as suppliers and users of AI systems covering public and private operators to ensure a level playing field. They can be summarised as

(a) machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods, including deep learning;

(b) logic and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;

(c) statistical approaches, Bayesian estimation, search methods and optimisation.

Further, in light of Recital 6, an AI system should be clearly defined to ensure legal certainty while providing the flexibility to accommodate future technological developments. The definition should be based on the essential functional characteristics of software. In particular, the ability to generate outputs such as content, predictions, recommendations or decisions that influence the environment with which the system interacts. AI systems can be designed to operate with varying degrees of autonomy and be used on their own or as product components, regardless of whether the system is physically integrated into the product (embedded) or serves the product's functionality without being integrated into it (non-embedded). Furthermore, the definition of an AI system should be complemented by a list of specific techniques and approaches used in its development, which should be kept up-to-date in light of technological and market updates through the adoption of delegated acts by the Commission to amend this list.

So far, the impression received from the Artificial Intelligence Act Proposal is that it is an attempt at definitions and classifications, while what is worrying from an anti-discrimination protection perspective is the lack of remedial mechanisms to protect individuals subjected to discriminatory automated decisions especially in light of what should have been the premise of the (new) body of law, whose purpose was to put the person at the centre.

The exclusive reference to the discipline of Art. 22 GDPR is not sufficient since it does not cover the extensive areas of artificial intelligence, but focuses mainly on automated decisions productive of legal effects, which affect individual rights, unless such a decision, including profiling, is necessary for a contractual context or is authorised by a law of the European Union or a Member State. In this regard, it has been argued in the doctrine that decisions under Article 22 GDPR cannot be based on special categories of personal data, such as biometric data, t unless the data subject allows it (ex Art. 9(2)(a) GDPR) or if there are reasons of public interest (ex. Art. 9(2)(g)). Nevertheless, these exceptions must provide a clear legal framework for protecting fundamental rights (Martini, 2020), or concerning investments in this area, despite a possible link to Article 5 of the AIA proposal.

In this context, Article 22 GDPR manifests its inadequacy to make up for the absence of a procedure appropriate to the rationale and purposes of the AIA Proposal. This absence is even more pronounced when one considers that the AIA Proposal defines which AI systems are 'high-risk' but does not provide any remedy (nor does it strengthen Article 22 GDPR for this purpose) in order to make effective the protection of those affected against discrimination produced by a risk of harm to security, health or an adverse impact on fundamental rights.

One might wonder whether Article 13 of the Proposal (under the heading 'Transparency and provision of information to users') could be supplementary to Article 22 GDPR in the framework prepared by the AIA Proposal. The letter of Article 13 denies such a purpose, as the article would manifest a mere proclamation of good intentions. Indeed, it is unlikely that the average user will be able to interact with an algorithm or understand the mechanism for achieving its results, however transparent it may be, although risk AI system must be accompanied by 'instructions for use in an appropriate digital or non-digital format; though such must include concise, complete, correct and precise information that is relevant, accessible and understandable to users'[7].

Article 13 of the AI proposal explicitly provides for information's correctness (accuracy) but does not provide truthfulness (trustworthiness). The two terms are not perfect synonyms in Italian or English. The former refers to the accuracy of information details, which should coincide with their truthfulness, but not necessarily. Therefore, in areas where the two concepts do not adhere, it is possible for discrimination or inappropriate automated data processing to occur without the provision of sanctions or penalties.

The Draft text of the Recommendation takes up this concern on the Ethics of Artificial Intelligence proposed by UNESCO[8]. This draft, however, delegates the provision of remedies against discriminatory treatment to AI operators, who 'must make every reasonable effort

7 Art. 13 AI Proposal.

8 UNESCO. *Draft Recommendation on the Ethics of Artificial Intelligence*. UNESCO General Conference, Paris.

to minimise and avoid reinforcing or perpetuating discriminatory or biased applications and results throughout the life cycle of the AI system, to ensure the fairness of such systems'[9].

Effective remedies against discrimination and algorithmic determination biased by negative bias should be available, and the European legislator or each member state should arrange this task. Leaving the remedy instrument in the hands of the individual operator would result in fragmentation contrary to the minimisation of discrimination in the results produced by the algorithmic determinants.

The prescriptive nature of the contents and requirements relating to such information makes it possible to exclude that Article 13 of the AI Proposal can play a remedial role. It merely establishes content and information-related prescriptions.

The amendments made to the Artificial Intelligence Act Proposal are attractive because they manifest their contradictory nature due to the ambition to adopt 'universal' impact legislation that the Proposal (and its promoters) claim. Indeed, as with the GDPR, the Proposal aims to candidate itself as a model for AI regulation in other legal systems. However, one may wonder whether such a complex and disorienting model, a fact due to the coexistence of limitations and exceptions, can be adopted in systems where decision-making automation is used on the administrative and public side mainly to make bureaucracy and obedience to the order more efficient, as well as to strengthen defensive or offensive military apparatuses, while on the private side to improve cost-benefit analysis through automation of the production chain.

On the other hand, the amendments manifest a fundamental contradiction with the concept of AI present in the European Parliament and, by extension, in public opinion or the electorate. The Rapporteur to the European Parliament recalled that artificial intelligence systems are based on software using mathematical models of a probabilistic nature as well as algorithmic predictions for several specific purposes. On the contrary, 'Artificial Intelligence' is a generic term, covering a wide range of old and new technologies, techniques and approaches, better understood as 'artificial intelligence systems', which refer to any machine-based system and which often have little more in common than the fact that a particular set of human-defined goals guides them, that they have some, varying degrees of autonomy in their actions. They engage in predictions, recommendations or decisions based on available data. The development of such technologies is not homogeneous; some are widely used, while others are under development or consist only of speculation that has yet to find design and concreteness[10].

---

9 *Ibid*.

10 Voss, Axel. (2022, April 20). *Draft Report on Artificial Intelligence in a digital age (2020/2266(INI)) Compromise Amendments European Parliament; Draft European Parliament Legislative Resolution on the Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM2021/0206 – C9-0146/2021 – 2021/0106(COD))*.

The opposing political positions are confirmed by the Amendments proposed by the joint LIBE (European Parliament's Committee on Civil Liberties, Justice and Home Affairs) and IMCO (Internal Market and Consumer Protection) Committees[11] concerning the definition of Artificial Intelligence. In fact, on the one hand, a definition of AI is presented that is as broad as possible and overrides the technical classifications set out in Annex I of the Proposal, while on the other hand, insisting on precise definitions, including machine learning (defined as the ability to find patterns without being explicitly programmed for specific tasks).

A new criterion for classifying high-risk systems has been introduced, changing the automatic classification for systems in the list of areas mentioned in Annex III to a list of 'critical use cases'. Based on these uses, AI providers must self-assess whether their systems present significant risks to health, safety and fundamental rights. During the approval of these amendments, political forces clashed over the use of algorithms used to evaluate creditworthiness, health insurance processes, payments and media recommendation systems.

It seems relevant to recognise, outside of expressions of intent or ideological statements, that structural biases in society should be avoided or even increased through low-quality data sets. It is stated explicitly that algorithms learn to be as discriminatory as the data they work with. As a result of low-quality training data or biases and discriminations observed in society, they may suggest inherently discriminatory decisions, which exacerbates discrimination in society. It is noted, however, that AI biases can sometimes be corrected. Furthermore, it is necessary to apply technical means and establish different levels of control over the software of AI systems, the algorithms and the data they use and produce to minimise risk. Finally, it claims, probably deluding itself, that AI can and should be used to reduce prejudice and discrimination. In reality, AI tends to amplify discrimination in society.

The Artificial Intelligence Act would make it possible to go beyond the 'case studies' model for risk assessment algorithms, which instead relies mainly on the above-mentioned case studies, focusing on individual and not collective risks, while the human rights impact of risk assessment algorithms is crucial. Given the sectorial nature of each subject, it is an approach that has yet to be absorbed by those involved in creating or formulating the algorithms. However, it is a fundamental step from a purely cultural and empirical point of view because it would allow for a change of anti-discriminatory perspective in the use of data.

[11] Benifei, Brando, Tudorache, Ioan-Dragoş. *DRAFT REPORT on the Proposal for a regulation of the European Parliament and the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM2021/0206 – C9-0146/2021 – 2021/0106(COD))*. Committee on the Internal Market and Consumer Protection Committee on Civil Liberties, Justice and Home Affairs.

Equally relevant is the issue of the prohibition of the use of technology, the content of which is extensive. On this point, the Proposal for a regulation contains similarly broad exceptions: it is an indisputable fact of legal significance that the entire Proposal is built around the classification of types of algorithms, with the prohibition of the use of social scoring algorithms or particular biometric software at the heart of the framework. Nonetheless – especially for biometrics – such broad exceptions leave room for contradictions. However, the rationale conveyed by Proposal is that, at least in the European Union member states, there are limits related to the protection of dignity and fundamental rights that cannot be exceeded in using artificial intelligence algorithms.

In light of this, one can also ask oneself whether introducing so-called 'sandboxes' might be an appropriate solution to balance the speed of technological development with the need to protect the human rights, especially from an anti-discrimination perspective, of those subjected to algorithmic decision-making. The answer cannot be immediate and is probably not satisfactory. Indeed, a sandbox is a functional space limited in time to experiment with AI systems, especially the decisive ones, with a view to their release on the market. Their goal is to evaluate the impact of ADMs on individuals and the social fabric. However, their effect is only sometimes immediate but can only be seen after their medium- to long-term operation.

The filing of thousands of amendments manifests a profound criticism of the political forces in the European Parliament against the Commission that promoted such a proposal, which makes the approval process or even the final approval of this Proposal for a regulation complex.

Conclusion

Advance of technologies had led not only to a progress and improves our life. It also brought threats to human rights related to the violation of privacy as well as discrimination. Discriminatory cases often occur in the automated processing of personal data. Since this process is not neutral, nor can it be since it is choice-based, it contains potential discrimination. It is important today to investigate whether and how it is possible to adopt measures of different elements related to fairness of algorithms.

The results made in this paper can be used as basis for future research in the sphere of algorithmic discrimination and privacy protection. It can also be used in law-making process.

References

Abdollahpouri, H., Mansoury, M., Burke, R., & Mobasher, B. (2020). The connection between popularity bias, calibration, and fairness in recommendation. In *Proceedings of the 14th ACM Conference on Recommender Systems* (pp. 726–731). https://doi.org/10.1145/3383313.3418487

Ainis, M. (2015). *La piccola eguaglianza.* Einaudi.

Alpa, G. (2021). Quale modello normativo europeo per l'intelligenza artificiale? *Contratto e impresa*, *37*(4), 1003–1026.

Alpa, G., & Resta, G. (2006). *Trattato di diritto civile. Le persone e la famiglia: 1. Le persone fisiche ei diritti della personalità*. UTET giuridica.
Altenried, M. (2020). The platform as factory: Crowdwork and the hidden labour behind artificial intelligence. *Capital & Class*, *44*(2), 145–158. https://doi.org/10.1177/0309816819899410
Amodio, E. (1970). *L'obbligo costituzionale di motivare e l'istituto della giuria. Rivista di diritto processuale.*
Angiolini, C. S. A. (2020). *Lo statuto dei dati personali: uno studio a partire dalla nozione di bene*. Giappichelli.
Bao, M., Zhou, A., Zottola, S., Brubach, B., Desmarais, S., Horowitz, A., ... & Venkatasubramanian, S. (2021). It's complicated: The messy relationship between rai datasets and algorithmic fairness benchmarks. *arXiv preprint arXiv:2106.05498*
Bargi, A. (1997). *Sulla struttura normativa della motivazione e sul suo controllo in Cassazione.* Giur. it.
Battini, S. (2018). *Indipendenza e amministrazione fra diritto interno ed europeo*.
Bellamy, R. (2014). Citizenship: Historical development of. Citizenship: Historical Development of'. In J. Wright (Ed.), *International Encyclopaedia of Social and Behavioural Sciences, Elsevier*. https://doi.org/10.1016/b978-0-08-097086-8.62078-0
Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2021). Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, *50*(1), 3–44. https://doi.org/10.1177/0049124118782533
Brooks, R. (2017). Machine Learning Explained. *Robots, AI and other stuff*.
Bodei, R. (2019). *Dominio e sottomissione*. Bologna, Il Mulino.
Canetti, E. (1960). *Masse und Macht*. Hamburg, Claassen.
Casonato, C., & Marchetti, B. (2021). Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale. *BioLaw Journal-Rivista di BioDiritto*, *3*, 415–437.
Chizzini, A. (1998). *Sentenza nel diritto processuale civile*. Dig. disc. priv., Sez. civ.
Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, *5*(2), 153–163. https://doi.org/10.1089/big.2016.0047
Citino, Y. (2022). *Cittadinanza digitale a punti e social scoring: le pratiche scorrette nell'era dell'intelligenza artificiale*. Diritti comparati.
Claeys, G. (2018). *Marx and Marxism*. Nation Books, New York.
Cockburn, I. M., Henderson, R., & Stern, S. (2018). The impact of artificial intelligence on innovation: An exploratory analysis. In *The economics of artificial intelligence: An agenda*. University of Chicago Press.
Cossette-Lefebvre, H., & Maclure, J. (2022). AI's fairness problem: understanding wrongful discrimination in the context of automated decision-making. *AI and Ethics*, *5*, 1–15. https://doi.org/10.1007/s43681-022-00233-w
Crawford, K. (2021). Time to regulate AI that interprets human emotions. *Nature*, *592*(7853), 167. https://doi.org/10.1038/d41586-021-00868-5
Custers, B. (2022). AI in Criminal Law: An Overview of AI Applications in Substantive and Procedural Criminal Law. In B. H. M. Custers, & E. Fosch Villaronga (Eds.), *Law and Artificial Intelligence* (pp. 205–223). Heidelberg: Springer. http://dx.doi.org/10.2139/ssrn.4331759
De Gregorio, G. & Paolucci F. (2022). *Dati personali e AI Act. Media laws*.
Di Rosa, G. (2021). Quali regole per i sistemi automatizzati "intelligenti"?. *Rivista di diritto civile*, *67*(5), 823–853.
Epp, C. R. (1996). Do bills of rights matter? The Canadian Charter of Rights and Freedoms, *American Political Science Review*, *90*(4), 765–779.
Fanchiotti, V. (1995). *Processo penale nei paesi di Common Law.* Dig. Disc. Pen.
Freeman, C., Louçã, F., & Louçã, F. (2001). *As time goes by: from the industrial revolutions to the information revolution*. Oxford University Press.
Freeman, K. (2016). Algorithmic injustice: How the Wisconsin Supreme Court failed to protect due process rights in State v. Loomis. *North Carolina Journal of Law & Technology*, *18*(5), 75–90.
Fuchs, C. (2014). *Digital Labour and Karl Marx*. Routledge.
Gallese, C.(2022). *Legal aspects of the use of continuous-learning models in Telemedicine*. JURISIN.
Gallese, E. Falletti, M. S. Nobile, L. Ferrario, Schettini, F. & Foglia, E. (2020). Preventing litigation with a predictive model of COVID-19 ICUs occupancy. *2020 IEEE International Conference on Big Data (Big Data)*. (pp. 2111–2116). Atlanta, GA, USA. https://doi.org/10.1109/BigData50022.2020.9378295
Garg, P., Villasenor, J., & Foggo, V. (2020). Fairness metrics: A comparative analysis. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 3662–3666). IEEE. https://doi.org/10.1109/bigdata50022.2020.9378025
Gressel, S., Pauleen, D. J., & Taskin, N. (2020). *Management decision-making, big data and analytics*. Sage.
Guo, F., Li, F., Lv, W., Liu, L., & Duffy, V. G. (2020). Bibliometric analysis of affective computing researches during 1999–2018. *International Journal of Human-Computer Interaction*, *36*(9), 801–814. https://doi.org/10.1080/10447318.2019.1688985

Hildebrandt, M. (2021). The issue of bias. The framing powers of machine learning. In Pelillo, M., & Scantamburlo, T. (Eds.), *Machines We Trust: Perspectives on Dependable AI*. MIT Press. https://doi.org/10.7551/mitpress/12186.003.0009

Hoffrage, U., & Marewski, J. N. (2020). Social Scoring als Mensch-System-Interaktion. *Social Credit Rating: Reputation und Vertrauen beurteilen*, 305–329. https://doi.org/10.1007/978-3-658-29653-7_17

Iftene, A. (2018). *Who Is Worthy of Constitutional Protection?* A Commentary on Ewert v Canada.

Infantino, M., & Wang, W. (2021). Challenging Western Legal Orientalism: A Comparative Analysis of Chinese Municipal Social Credit Systems. *European Journal of Comparative Law and Governance*, *8*(1), 46–85. https://doi.org/10.1163/22134514-bja10011

Israni, E. (2017). *Algorithmic due process: mistaken accountability and attribution in State v. Loomis*.

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*.

Krawiec, A., Pawela, Ł., & Puchała, Z. (2023). Discrimination and certification of unknown quantum measurements. *arXiv preprint arXiv:2301.04948*.

Kubat, M., & Kubat, J. A. (2017). *An introduction to machine learning* (Vol. 2, pp. 321–329). Cham, Switzerland: Springer International Publishing.

Kuhn, Th. S. (1962). The structure of scientific revolutions. *International Encyclopedia of Unified Science*, *2*(2).

Lippert-Rasmussen, K. (2022). Algorithm-Based Sentencing and Discrimination, *Sentencing and Artificial Intelligence* (pp. 74–96). Oxford University Press.

Maamar, N. (2018). Social Scoring: Eine europäische Perspektive auf Verbraucher-Scores zwischen Big Data und Big Brother. *Computer und Recht*, *34*(12), 820–828. https://doi.org/10.9785/cr-2018-341212

Mannozzi, G. (1997). Sentencing. *Dig. Disc. Pen.*

Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust*. Vintage.

Martini, M. (2020). Regulating Algorithms – How to demystify the alchemy of code?. In *Algorithms and Law* (pp. 100–135). Cambridge University Press. https://doi.org/10.1017/9781108347846.004

Marx, K. (2016). Economic and philosophic manuscripts of 1844. In *Social Theory Re-Wired*. Routledge

Massa, M. (1990). *Motivazione della sentenza (diritto processuale penale)*. Enc. Giur.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Messinetti, R. (2019). La tutela della persona umana versus l'intelligenza artificiale. Potere decisionale dell'apparato tecnologico e diritto alla spiegazione della decisione automatizzata, *Contratto e impresa*, *3*, 861–894.

Mi, F., Kong, L., Lin, T., Yu, K., & Faltings, B. (2020). Generalised class incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 240–241). https://doi.org/10.1109/cvprw50498.2020.00128

Mitchell, T. M. (2007). *Machine learning* (Vol. 1). New York: McGraw-hill.

Nazir, A., Rao, Y., Wu, L., & Sun, L. (2020). Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. *IEEE Transactions on Affective Computing*, *13*(2), 845–863. https://doi.org/10.1109/taffc.2020.2970399

Oswald, M. (2018). Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2128), 20170359. https://doi.org/10.1098/rsta.2017.0359

Oswald, M., Grace, J., Urwin, S., & Barnes, G. C. (2018). Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental'proportionality. *Information & communications technology law*, *27*(2), 223–250.

Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. *Neural networks*, *113*, 54–71.

Parona, L. (2021). Government by algorithm": un contributo allo studio del ricorso all'intelligenza artificiale nell'esercizio di funzioni amministrative. *Giornale Dir. Amm*, *1*.

Pellecchia, E. (2018). Profilazione e decisioni automatizzate al tempo della black box society: qualità dei dati e leggibilità dell'algoritmo nella cornice della responsible research and innovation. *Nuove leg. civ. comm*, 1209–1235.

Pessach, D., & Shmueli, E. (2020). Algorithmic fairness. *arXiv preprint arXiv:2001.09784*.

Petronio, U. (2020*). Il precedente negli ordinamenti giuridici continentali di antico regime. Rivista di diritto civile*, *66*(5), 949–983.

Pleiss, G., Raghavan, M., Wu, F., Kleinberg, J., & Weinberger, K. Q. (2017). On fairness and calibration. *Advances in neural information processing systems*, *30*.

Poria, S., Hazarika, D., Majumder, N., & Mihalcea, R. (2020). Beneath the tip of the iceberg: Current challenges and new directions in sentiment analysis research, *IEEE Transactions on Affective Computing*. https://doi.org/10.1109/taffc.2020.3038167

Rebitschek, F. G., Gigerenzer, G., & Wagner, G. G. (2021). People underestimate the errors made by algorithms for credit scoring and recidivism prediction but accept even fewer errors. *Scientific reports*, *11*(1), 1–11.

Rodotà, S. (1995). *Tecnologie e diritti*, il Mulino. Bologna.

Rodotà, S. (2012). *Il diritto di avere diritti*. Gius. Laterza.

Rodotà, S. (2014). *Il mondo nella rete: Quali i diritti, quali i vincoli*. GLF Editori Laterza.

Russell, P. H. (1983). The political purposes of the Canadian Charter of Rights and Freedoms. *Can. B. Rev.*, *61*, 30–35.

Scassa, T. (2021). Administrative Law and the Governance of Automated Decision Making: A Critical Look at Canada's Directive on Automated Decision Making, *UBCL Rev*, *54*, 251–255. https://doi.org/10.2139/ssrn.3722192

Scherer, M. U. (2015). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies, *Harv. JL & Tech.*, *29*, 353–360. https://doi.org/10.2139/ssrn.2609777

Schiavone, A. (2019). *Eguaglianza.* Einaudi.

Starr, S. B. (2014). Evidence-based sentencing and the scientific rationalisation of discrimination. *Stanford Law Review*, *66*, 803–872.

Stuurman, K., & Lachaud, E. (2022). Regulating AI. A label to complete the proposed Act on Artificial Intelligence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3963890

Sunstein, C. R. (2019). Algorithms, correcting biases. *Social Research: An International Quarterly*, *86*(2), 499–511. https://doi.org/10.1353/sor.2019.0024

Tarrant, A., & Cowen, T. (2022). Big Tech Lobbying in the EU. *The Political Quarterly*, *93*(2), 218–226. https://doi.org/10.1111/1467-923x.13127

Taruffo, M. (1975). *La motivazione della sentenza civile*. Cedam, Padova.

Vale, D., El-Sharif, A., & Ali, M. (2022). Explainable artificial intelligence (XAI) post-hoc explainability methods: Risks and limitations in non-discrimination law. *AI and Ethics*, 1–12. https://doi.org/10.1007/s43681-022-00142-y

Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act-Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, *22*(4), 97–112. https://doi.org/10.31235/osf.io/38p5f

Vogel, P. A. (2020). "Right to explanation" for algorithmic decisions?, *Data-Driven Decision Making. Law, Ethics, Robotics, Health*, *49*, 1–12. https://doi.org/10.48550/arXiv.1606.08813

Von Tunzelmann, N. (2003). Historical coevolution of governance and technology in the industrial revolutions, *Structural Change and Economic Dynamics*, *14*(4), 365–384. https://doi.org/10.1016/s0954-349x(03)00029-8

Wang, C., Han, B., Patel, B., & Rudin, C. (2022). In pursuit of interpretable, fair and accurate machine learning for criminal recidivism prediction, *Journal of Quantitative Criminology*, *6*, 1–63. https://doi.org/10.1007/s10940-022-09545-w

Witt, A. C. (2022). Platform Regulation in Europe – Per Se Rules to the Rescue?, *Journal of Competition Law & Economics*, *18*(3), 670–708. https://doi.org/10.1093/joclec/nhac001

Woodcock, J. (2020). The algorithmic panopticon at Deliveroo: Measurement, precarity, and the illusion of control, *Ephemera: theory & politics in organisations*, *20*(3), 67–95.

York, J. C. (2022). *Silicon values: The future of free speech under surveillance capitalism*. Verso Books, London-New York.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile books, London.

Author information

Elena Faletti – PhD, Assistant Professor, Carlo Cattaneo University LIUC
Address: Corso Matteotti 22, Castellanza, 21053, Italy
E-mail: efalletti@liuc.it
ORCID ID: https://orcid.org/0000-0002-6121-6775
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=57040979500

Conflict of interest

The author declares no conflict of interest.

Financial disclosure

The research had no sponsorship.

Thematic rubrics

OECD: 5.05 / Law
PASJC: 3308 / Law
WoS: OM / Law

Article history

Date of receipt – February 24, 2023
Date of approval – April 13, 2023
Date of acceptance – June 16, 2023
Date of online placement – June 20, 2023

Научная статья
УДК 347.1:004.8
EDN: https://elibrary.ru/ktizpw
DOI: https://doi.org/10.21202/jdtl.2023.16

Алгоритмическая дискриминация и защита неприкосновенности частной жизни

Элена Фаллетти

Университет Карло Каттанео
г. Кастелланца, Итальянская Республика

Ключевые слова

Алгоритм,
дискриминация,
защита данных,
искусственный интеллект,
неприкосновенность,
персональные данные,
право,
регулирование,
цифровые технологии,
частная жизнь

Аннотация

Цель: появление цифровых технологий, таких как искусственный интеллект, стало вызовом для государств всего мира. Оно породило множество рисков, связанных с нарушением прав человека, включая права на неприкосновенность частной жизни и человеческое достоинство. Это определяет актуальность данного исследования. Цель статьи – проанализировать роль алгоритмов в случаях дискриминации и выяснить, каким образом алгоритмы могут способствовать предубежденности при принятии решений на основе персональных данных. Проведенный анализ помогает оценить проект закона об искусственном интеллекте, направленный на регулирование данной проблемы для предотвращения дискриминации при использовании алгоритмов.

Методы: в работе применялись методы эмпирического и сравнительного анализа. Сравнительный анализ позволил выявить сходства и различия существующего регулирования и положений проекта закона об искусственном интеллекте. С помощью эмпирического анализа рассмотрены реальные примеры алгоритмической дискриминации.

Результаты: результаты исследования показывают, что Закон об искусственном интеллекте нуждается в доработке, так как он остается на уровне дефиниций и недостаточно опирается на эмпирический материал. Автор выдвигает ряд предложений по совершенствованию данного законопроекта.

Научная новизна: заключается в мультидисциплинарности данной работы, рассматривающей вопросы дискриминации, защиты данных и влияния на эмпирическую реальность в сфере алгоритмической дискриминации и охраны неприкосновенности частной жизни.

Практическая значимость: состоит в привлечении внимания к тому факту, что алгоритмы выполняют инструкции, составленные на основе сообщенных им данных. Не имея возможностей для абдукции, алгоритмы действуют лишь как послушные исполнители приказов. Результаты работы могут использоваться в качестве основы для будущих исследований в данной области и в законотворческом процессе.

Для цитирования

Фаллетти, Э. (2023). Алгоритмическая дискриминация и защита неприкосновенности частной жизни. *Journal of Digital Technologies and Law*, *1*(2), 387–420. https://doi.org/10.21202/jdtl.2023.16

Список литературы

Abdollahpouri, H., Mansoury, M., Burke, R., & Mobasher, B. (2020). The connection between popularity bias, calibration, and fairness in recommendation. In *Proceedings of the 14th ACM Conference on Recommender Systems* (pp. 726–731). https://doi.org/10.1145/3383313.3418487

Ainis, M. (2015). *La piccola eguaglianza.* Einaudi.

Alpa, G. (2021). Quale modello normativo europeo per l'intelligenza artificiale? *Contratto e impresa*, *37*(4), 1003–1026.

Alpa, G., & Resta, G. (2006). *Trattato di diritto civile. Le persone e la famiglia: 1. Le persone fisiche ei diritti della personalità*. UTET giuridica.

Altenried, M. (2020). The platform as factory: Crowdwork and the hidden labour behind artificial intelligence. *Capital & Class*, *44*(2), 145–158. https://doi.org/10.1177/0309816819899410

Amodio, E. (1970). *L'obbligo costituzionale di motivare e l'istituto della giuria. Rivista di diritto processuale.*

Angiolini, C. S. A. (2020). *Lo statuto dei dati personali: uno studio a partire dalla nozione di bene*. Giappichelli.

Bao, M., Zhou, A., Zottola, S., Brubach, B., Desmarais, S., Horowitz, A., ... & Venkatasubramanian, S. (2021). It's complicated: The messy relationship between rai datasets and algorithmic fairness benchmarks. *arXiv preprint arXiv:2106.05498*

Bargi, A. (1997). *Sulla struttura normativa della motivazione e sul suo controllo in Cassazione.* Giur. it.

Battini, S. (2018). *Indipendenza e amministrazione fra diritto interno ed europeo*.

Bellamy, R. (2014). Citizenship: Historical development of. Citizenship: Historical Development of'. In J. Wright (Ed.), *International Encyclopaedia of Social and Behavioural Sciences, Elsevier*. https://doi.org/10.1016/b978-0-08-097086-8.62078-0

Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2021). Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, *50*(1), 3–44. https://doi.org/10.1177/0049124118782533

Brooks, R. (2017). Machine Learning Explained. *Robots, AI and other stuff*.

Bodei, R. (2019). *Dominio e sottomissione*. Bologna, Il Mulino.

Canetti, E. (1960). *Masse und Macht*. Hamburg, Claassen.

Casonato, C., & Marchetti, B. (2021). Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale. *BioLaw Journal-Rivista di BioDiritto*, *3*, 415–437.

Chizzini, A. (1998). *Sentenza nel diritto processuale civile*. Dig. disc. priv., Sez. civ.

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, *5*(2), 153–163. https://doi.org/10.1089/big.2016.0047

Citino, Y. (2022). *Cittadinanza digitale a punti e social scoring: le pratiche scorrette nell'era dell'intelligenza artificiale*. Diritti comparati.

Claeys, G. (2018). *Marx and Marxism*. Nation Books, New York.

Cockburn, I. M., Henderson, R., & Stern, S. (2018). The impact of artificial intelligence on innovation: An exploratory analysis. In *The economics of artificial intelligence: An agenda*. University of Chicago Press.

Cossette-Lefebvre, H., & Maclure, J. (2022). AI's fairness problem: understanding wrongful discrimination in the context of automated decision-making. *AI and Ethics*, *5*, 1–15. https://doi.org/10.1007/s43681-022-00233-w

Crawford, K. (2021). Time to regulate AI that interprets human emotions. *Nature*, *592*(7853), 167. https://doi.org/10.1038/d41586-021-00868-5

Custers, B. (2022). AI in Criminal Law: An Overview of AI Applications in Substantive and Procedural Criminal Law. In B. H. M. Custers, & E. Fosch Villaronga (Eds.), *Law and Artificial Intelligence* (pp. 205–223). Heidelberg: Springer. http://dx.doi.org/10.2139/ssrn.4331759

De Gregorio, G. & Paolucci F. (2022). *Dati personali e AI Act. Media laws*.

Di Rosa, G. (2021). Quali regole per i sistemi automatizzati "intelligenti"?. *Rivista di diritto civile*, *67*(5), 823–853.

Epp, C. R. (1996). Do bills of rights matter? The Canadian Charter of Rights and Freedoms, *American Political Science Review*, *90*(4), 765–779.

Fanchiotti, V. (1995). *Processo penale nei paesi di Common Law.* Dig. Disc. Pen.

Freeman, C., Louçã, F., & Louçã, F. (2001). *As time goes by: from the industrial revolutions to the information revolution*. Oxford University Press.

Freeman, K. (2016). Algorithmic injustice: How the Wisconsin Supreme Court failed to protect due process rights in State v. Loomis. *North Carolina Journal of Law & Technology*, *18*(5), 75–90.

Fuchs, C. (2014). *Digital Labour and Karl Marx*. Routledge.

Gallese, C.(2022). *Legal aspects of the use of continuous-learning models in Telemedicine*. JURISIN.

Gallese, E. Falletti, M. S. Nobile, L. Ferrario, Schettini, F. & Foglia, E. (2020). Preventing litigation with a predictive model of COVID-19 ICUs occupancy. *2020 IEEE International Conference on Big Data (Big Data)*. (pp. 2111–2116). Atlanta, GA, USA. https://doi.org/10.1109/BigData50022.2020.9378295

Garg, P., Villasenor, J., & Foggo, V. (2020). Fairness metrics: A comparative analysis. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 3662–3666). IEEE. https://doi.org/10.1109/bigdata50022.2020.9378025

Gressel, S., Pauleen, D. J., & Taskin, N. (2020). *Management decision-making, big data and analytics*. Sage.

Guo, F., Li, F., Lv, W., Liu, L., & Duffy, V. G. (2020). Bibliometric analysis of affective computing researches during 1999–2018. *International Journal of Human-Computer Interaction*, *36*(9), 801–814. https://doi.org/10.1080/10447318.2019.1688985

Hildebrandt, M. (2021). The issue of bias. The framing powers of machine learning. In Pelillo, M., & Scantamburlo, T. (Eds.), *Machines We Trust: Perspectives on Dependable AI*. MIT Press. https://doi.org/10.7551/mitpress/12186.003.0009

Hoffrage, U., & Marewski, J. N. (2020). Social Scoring als Mensch-System-Interaktion. *Social Credit Rating: Reputation und Vertrauen beurteilen*, 305–329. https://doi.org/10.1007/978-3-658-29653-7_17

Iftene, A. (2018). *Who Is Worthy of Constitutional Protection?* A Commentary on Ewert v Canada.

Infantino, M., & Wang, W. (2021). Challenging Western Legal Orientalism: A Comparative Analysis of Chinese Municipal Social Credit Systems. *European Journal of Comparative Law and Governance*, *8*(1), 46–85. https://doi.org/10.1163/22134514-bja10011

Israni, E. (2017). *Algorithmic due process: mistaken accountability and attribution in State v. Loomis*.

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*.

Krawiec, A., Pawela, Ł., & Puchała, Z. (2023). Discrimination and certification of unknown quantum measurements. *arXiv preprint arXiv:2301.04948*.

Kubat, M., & Kubat, J. A. (2017). *An introduction to machine learning* (Vol. 2, pp. 321–329). Cham, Switzerland: Springer International Publishing.

Kuhn, Th. S. (1962). The structure of scientific revolutions. *International Encyclopedia of Unified Science*, *2*(2).

Lippert-Rasmussen, K. (2022). Algorithm-Based Sentencing and Discrimination, *Sentencing and Artificial Intelligence* (pp. 74–96). Oxford University Press.

Maamar, N. (2018). Social Scoring: Eine europäische Perspektive auf Verbraucher-Scores zwischen Big Data und Big Brother. *Computer und Recht*, *34*(12), 820–828. https://doi.org/10.9785/cr-2018-341212

Mannozzi, G. (1997). Sentencing. *Dig. Disc. Pen.*

Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust*. Vintage.

Martini, M. (2020). Regulating Algorithms – How to demystify the alchemy of code?. In *Algorithms and Law* (pp. 100–135). Cambridge University Press. https://doi.org/10.1017/9781108347846.004

Marx, K. (2016). Economic and philosophic manuscripts of 1844. In *Social Theory Re-Wired*. Routledge

Massa, M. (1990). *Motivazione della sentenza (diritto processuale penale)*. Enc. Giur.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Messinetti, R. (2019). La tutela della persona umana versus l'intelligenza artificiale. Potere decisionale dell'apparato tecnologico e diritto alla spiegazione della decisione automatizzata, *Contratto e impresa*, *3*, 861–894.

Mi, F., Kong, L., Lin, T., Yu, K., & Faltings, B. (2020). Generalised class incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 240–241). https://doi.org/10.1109/cvprw50498.2020.00128

Mitchell, T. M. (2007). *Machine learning* (Vol. 1). New York: McGraw-hill.

Nazir, A., Rao, Y., Wu, L., & Sun, L. (2020). Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. *IEEE Transactions on Affective Computing*, *13*(2), 845–863. https://doi.org/10.1109/taffc.2020.2970399

Oswald, M. (2018). Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2128), 20170359. https://doi.org/10.1098/rsta.2017.0359

Oswald, M., Grace, J., Urwin, S., & Barnes, G. C. (2018). Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental'proportionality. *Information & communications technology law*, *27*(2), 223–250.

Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. *Neural networks*, *113*, 54–71.

Parona, L. (2021). Government by algorithm": un contributo allo studio del ricorso all'intelligenza artificiale nell'esercizio di funzioni amministrative. *Giornale Dir. Amm*, *1*.

Pellecchia, E. (2018). Profilazione e decisioni automatizzate al tempo della black box society: qualità dei dati e leggibilità dell'algoritmo nella cornice della responsible research and innovation. *Nuove leg. civ. comm*, 1209–1235.

Pessach, D., & Shmueli, E. (2020). Algorithmic fairness. *arXiv preprint arXiv:2001.09784*.

Petronio, U. (2020*). Il precedente negli ordinamenti giuridici continentali di antico regime. Rivista di diritto civile*, *66*(5), 949–983.

Pleiss, G., Raghavan, M., Wu, F., Kleinberg, J., & Weinberger, K. Q. (2017). On fairness and calibration. *Advances in neural information processing systems*, *30*.

Poria, S., Hazarika, D., Majumder, N., & Mihalcea, R. (2020). Beneath the tip of the iceberg: Current challenges and new directions in sentiment analysis research, *IEEE Transactions on Affective Computing*. https://doi.org/10.1109/taffc.2020.3038167

Rebitschek, F. G., Gigerenzer, G., & Wagner, G. G. (2021). People underestimate the errors made by algorithms for credit scoring and recidivism prediction but accept even fewer errors. *Scientific reports*, *11*(1), 1–11.

Rodotà, S. (1995). *Tecnologie e diritti*, il Mulino. Bologna.

Rodotà, S. (2012). *Il diritto di avere diritti*. Gius. Laterza.

Rodotà, S. (2014). *Il mondo nella rete: Quali i diritti, quali i vincoli*. GLF Editori Laterza.

Russell, P. H. (1983). The political purposes of the Canadian Charter of Rights and Freedoms. *Can. B. Rev.*, *61*, 30–35.

Scassa, T. (2021). Administrative Law and the Governance of Automated Decision Making: A Critical Look at Canada's Directive on Automated Decision Making, *UBCL Rev*, *54*, 251–255. https://doi.org/10.2139/ssrn.3722192

Scherer, M. U. (2015). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies, *Harv. JL & Tech.*, *29*, 353–360. https://doi.org/10.2139/ssrn.2609777

Schiavone, A. (2019). *Eguaglianza.* Einaudi.

Starr, S. B. (2014). Evidence-based sentencing and the scientific rationalisation of discrimination. *Stanford Law Review*, *66*, 803–872.

Stuurman, K., & Lachaud, E. (2022). Regulating AI. A label to complete the proposed Act on Artificial Intelligence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3963890

Sunstein, C. R. (2019). Algorithms, correcting biases. *Social Research: An International Quarterly*, *86*(2), 499–511. https://doi.org/10.1353/sor.2019.0024

Tarrant, A., & Cowen, T. (2022). Big Tech Lobbying in the EU. *The Political Quarterly*, *93*(2), 218–226. https://doi.org/10.1111/1467-923x.13127

Taruffo, M. (1975). *La motivazione della sentenza civile*. Cedam, Padova.

Vale, D., El-Sharif, A., & Ali, M. (2022). Explainable artificial intelligence (XAI) post-hoc explainability methods: Risks and limitations in non-discrimination law. *AI and Ethics*, 1–12. https://doi.org/10.1007/s43681-022-00142-y

Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act-Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, *22*(4), 97–112. https://doi.org/10.31235/osf.io/38p5f

Vogel, P. A. (2020). “Right to explanation” for algorithmic decisions?, *Data-Driven Decision Making. Law, Ethics, Robotics, Health*, *49*, 1–12. https://doi.org/10.48550/arXiv.1606.08813

Von Tunzelmann, N. (2003). Historical coevolution of governance and technology in the industrial revolutions, *Structural Change and Economic Dynamics*, *14*(4), 365–384. https://doi.org/10.1016/s0954-349x(03)00029-8

Wang, C., Han, B., Patel, B., & Rudin, C. (2022). In pursuit of interpretable, fair and accurate machine learning for criminal recidivism prediction, *Journal of Quantitative Criminology*, *6*, 1–63. https://doi.org/10.1007/s10940-022-09545-w

Witt, A. C. (2022). Platform Regulation in Europe – Per Se Rules to the Rescue?, *Journal of Competition Law & Economics*, *18*(3), 670–708. https://doi.org/10.1093/joclec/nhac001

Woodcock, J. (2020). The algorithmic panopticon at Deliveroo: Measurement, precarity, and the illusion of control, *Ephemera: theory & politics in organisations*, *20*(3), 67–95.

York, J. C. (2022). *Silicon values: The future of free speech under surveillance capitalism*. Verso Books, London-New York.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile books, London.

Сведения об авторе

Элена Фаллетти – доктор наук, доцент, Университет Карло Каттанео
Адрес: Корсо Маттеотти 22, Кастелланца, 21053, Италия
E-mail: efalletti@liuc.it
ORCID ID: https://orcid.org/0000-0002-6121-6775
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=57040979500

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law
Рубрика ASJC: 3308 / Law
Рубрика WoS: OM / Law
Рубрика ГРНТИ: 10.27.51 / Осуществление и защита гражданских прав
Специальность ВАК: 5.1.3 / Частно-правовые (цивилистические) науки

История статьи

Дата поступления – 24 февраля 2023 г.
Дата одобрения после рецензирования – 13 апреля 2023 г.
Дата принятия к опубликованию – 16 июня 2023 г.
Дата онлайн-размещения – 20 июня 2023 г.