

<https://doi.org/10.17803/lexgen-2022-1-1-63-74>

Original article



Algorithmic Bias and Non-Discrimination in Argentina

Favio Farinella*

Mar del Plata National University, Buenos Aires, Argentina

Abstract

One of the major research problems related to artificial intelligence (AI) models at present is algorithmic bias. When an automated system “makes a decision” based on its training data, it can reveal biases similar to those inherent in the humans who provided the training data. Much of the data used to train the models comes from vector representations of words obtained from text corpuses, which can transmit stereotypes and social prejudices. AI system design focused on optimising processes and improving prediction accuracy ignores the need for new standards for compensating the negative impact of AI on the most vulnerable categories of peoples. An improved understanding of the relationship between algorithms, bias, and non-discrimination not only precedes any eventual solution, but also helps us to recognize how discrimination is created, maintained, and disseminated in the AI era, as well as how it could be projected into the future using various neurotechnologies. The opacity of the algorithmic decision-making process should be replaced by transparency in AI processes and models. The present work aims to reconcile the use of AI with algorithmic decision processes that respect the basic human rights of the individual, especially the principles of non-discrimination and positive discrimination. The Argentine legislation serves as the legal basis of this work.

Keywords: algorithmic bias, discrimination, artificial intelligence, explainable AI, transparency, human rights

To cite this article: Farinella, F. (2022). Algorithmic bias and non-discrimination in Argentina. *Lex Genetica*, 2022;1(1):62–74. <https://doi.org/10.17803/lexgen-2022-1-1-63-74>

Received 15.11.2022

Review completed 16.12.2022

Passed for printing 03.01.2023

* Email: faviofarinella@hotmail.com

Introduction

The development of technologies aimed at understanding the functioning of the brain paves the way to intervene directly in its processes, and consequently, to manipulate human brain activity. While such technologies may be claimed to be neutral, they may have both positive and negative consequences depending on how they are used. The claimed possibility of deciphering the neural code implies ethical challenges in terms of novel medical and technological applications that could be realised on the basis of such an informational infrastructure. Thus, the need to recognise new rights related to neuroscientific technologies is a topic of current discussion. While the various algorithms forming an artificial intelligence (AI) program may be transparent and specific, any AI or neurotechnological intervention in the brain's activities suffers from the same deficiencies as the humans who design them or intervene in their creative processes in terms of prejudices or biases. Banjali and Greenwald (2013) remind us that even the most ostensibly well-intentioned people may possess unconscious or implicit biases against other groups.

The present work discusses the problem of biases applying in AI and neurotechnology. In a postmodern age, such algorithmic biases can be used to surreptitiously perpetuate individual and social discrimination. Even unconscious personal biases can be easily transferred to products that make use of AI to generate discriminatory results. Reliance on data generated by AI does not give algorithms a presumption of veracity. This is so because in addition to the prejudices that may exist at the time of software design, we must consider that the data on which it is based may also be biased. If these data express a present reality that discriminates

against certain groups, their use by AI is consequently likely to reinforce existing patterns of discrimination (Mayson, 2018). Beyond the speed and efficiency with which a problem can be solved, neither AI nor neuroscience guarantee the justice of the decisions made through their use. The realisation of justice as a human value (even through the use of technology) requires some degree of human intervention.

In early 2016, Microsoft launched Tay, an AI chatbot that was supposed to mimic the behaviour of a curious teenage girl engaging in discussions with Twitter users. While the stated intention was to demonstrate the potential of conversational interfaces powered by AI, in less than a day, innocent Tay had apparently become a racist, misogynist Holocaust denier (Metz, 2018). Thus, the myth of algorithmic impartiality failure was debunked, sparking discussions concerning potential solutions to the identified biases. Reasons behind biases in software can be found in machine learning and deep learning processes underpinning AI. Deep learning algorithms rely on very large quantities of data. The more tagged data an algorithm records, the better the result. However, as data are recognised, deep learning algorithms develop blind spots based on missing data or excess data relative to what is trained. This will be the beginning of its bias. Tay, for instance, did not have the opportunity to interact with respectful, non-discriminatory, open-minded or empathic people.

Non-Discrimination and Bias

The prohibition of discrimination is a vital pillar of international human rights law and can already be considered part of *jus cogens*. The right to equality before the law and the principle of non-discrimination are

provided in the Argentine Constitution (Argentine Const. art. 16, 19, 22, 23, 37, 75) and the American Convention on Human Rights, as well as in various international instruments that employ a constitutional hierarchy¹.

The Supreme Court of Justice of Argentina has ruled on numerous occasions on the scope of Article 16 of the National Constitution confirming that equality before the law involves the duty on the part of the State to 'treat equally those people who are in identical circumstances' (Supreme Court of Justice, Argentina, Sentences 16:118) and furthermore, that 'equality before the law (...) is nothing other than denying the creation of exceptions or privileges that exclude some people from what is granted to others under the same conditions' (Supreme Court of Justice, Argentina, Sentences 153:67).

Discrimination consists not only in making a distinction or difference, but implies the unfavourable treatment of a person in a particular circumstance as prohibited by law. Certain differentiated treatments are even legal. In this sense, when determining the scope of the Discriminatory Acts Law (Law No. 23, 592), the Supreme Court of Argentina argued that the Law:

(...) does not sanction all discrimination, but exclusively that which arbitrarily restricts in some way or undermines the full exercise on equal bases of the fundamental rights and guarantees recognised in the National Constitution' (Supreme Court of Justice, Argentina, Sentences 314:1531 and ss).

In the words of the Inter-American Court of Human Rights (IACHR):

'Not all different legal treatment is properly discriminatory, because not every distinction in treatment can be considered offensive to human dignity. There are certain inequalities in fact that can be translated into justi-

fied inequalities of legal treatment, which express a proportionate relationship between the objective differences and the aims of the norm' (*Inter American Court of Human Rights, Advisory Opinion OC-4/84 of 01/19/1984, § 56-58*).

In this context, both private companies and state institutions increasingly rely on the automated decisions of algorithm-based systems, all of which could potentially involve the discriminative use of AI models and algorithms. As stated by the Inter-American Court, obligations in matters of equality and non-discrimination fall under the powers of the State, as well as applying to individuals, since the obligation:

'(...) extends as much with respect to those cases in which the discriminatory situation is the result of the actions and omissions of the public powers as when it is the result of the behavior of individuals' (*Inter-American Court of Human Rights, Consultative Opinion 18/03, § 4*).

A study conducted by the Institute for Technology Assessment and Systems Analysis in Karlsruhe (Germany) on behalf of the German Federal Agency Against Discrimination, found that, although AI increases efficiency to save time and money, it also carries risks of discrimination against individuals or vulnerable population groups.

In an increasing number of categories, such as granting a loan, hiring new staff members, or making legal decisions, algorithms are applied to decision-making or helping human decision makers to come to a final decision. In both cases, this circumstance affects the lives of other individuals. Carsten Orwat, who works at the Institute for Technology Assessment and Systems Analysis, states that "situations become particularly critical when algorithms operate on inaccurate data and

are based on criteria that must be protected, such as age, gender, ethnicity, religion, sexual orientation and disabilities" (Karlsruhe Institute of Technology, 2019). Biased data samples can teach machines that women shop and cook, while men go out to work. This type of problem occurs when the training data provided by scientists reflects their own prejudices (Mullane, 2018).

The study of algorithmic bias focuses on algorithms that reflect some type of systematic and unfair discrimination and has only recently begun to be considered for the purposes of its legal regulation, for example with the General Data Protection Regulation of the European Union (2018).

Areas of protection against discrimination and biased algorithms

When talking about illegal discrimination, it is necessary also to note that the law seeks to protect vulnerable groups, which, depending on their characteristics, tend to be usual victims of discrimination. This circumstance gives rise to a list of "suspicious" or "prohibited" categories. Generally speaking, such categories include race, gender, religion, political opinions, national or social origin, economic status, as well as certain physical characteristics.

The standardisation of suspicious categories is useful in order to distinguish the work to be performed by the justice as well as to know the distribution of the burden of proof in a legal process. When differences in treatment are based on such "suspicious" categories, a rigorous test of reasonableness is required. On occasions, the norm or practice is analysed by means of a "standard scrutiny", which attempts to maintain a balance between the parties regarding the burden of proof, and where the applicant must prove

that the differential treatment to which he or she was allegedly subjected is in violation of the principle of non-discrimination, as unconstitutionality is not presumed. In other situations, a "strict scrutiny" approach should be used: the contested rule or practice is presumed unconstitutional, and it is for the defendant to prove that it pursues a legitimate, relevant and imperative purpose, and that the means he/she chose is suitable, essential and constitutes the less harmful alternative in terms of the rights of those people affected.

European Union regulations distinguish specific areas of protection such as employment, welfare and social security, education, access to the supply of goods and services including housing, access to justice, private and family life, adoption, domicile and marriage, political participation, freedom of expression, assembly, association, free elections and criminal matters. It also protects against discrimination on various grounds, specially including: sex, gender identity, sexual orientation, disability, age, race, ethnic origin, colour and pertaining to a national minority, religion or beliefs, social origin, birth or property, language, and political or other opinions.

Article 14 of the European Convention on Human Rights and Fundamental Freedoms (ECHR) applies in relation to the enjoyment of the substantive rights recognised therein; while Protocol 12 to the ECHR protects all rights recognised at the national level, even those not protected by the ECHR. In contrast to this, the prohibition of discrimination emanating from the EU Directives applies only in three areas: (i) employment, (ii) the social welfare system and (iii) goods and services. The Racial Equality Directive also applies to such areas only. As regards the Directive, when it refers to equal treatment in employment, it is only applied

to labour matters, even though its extension to the other aforementioned areas is currently being debated. The Directive on equal treatment between men and women and the Directive on equal treatment between men and women in terms of access to goods and services only apply to the aforementioned contexts, but not in relation to access to the social welfare system (*de Europa & de Derechos Humanos*, 2019).

In Argentina, the law penalises discriminatory acts, paying particular attention to 'discriminatory acts or omissions determined for reasons such as race, religion, nationality, ideology, political or union opinion, sex, economic position, social condition or physical characteristics' (Anti-Discrimination Law, 1988). The Office of the Public Prosecutor of the Nation distinguishes several types of discrimination: (i) against women; (ii) based on sexual orientation; (iii) based on disability; (iii) religious reasons; and (iv) other reasons (MPF Argentina, 2012–2017). The regulations of the Autonomous City of Buenos Aires distinguish between discrimination in facts and in law, the latter being able to manifest itself directly or indirectly (Law No. 5261, 2015).

There are some well-known examples of discrimination by algorithms. First, in the case of women, international obligations on non-discrimination require the State to adopt positive action measures to counteract gender segregation and reverse the sociocultural patterns that explain it (*Assembly*, 1979, art. 2, 4). The Committee of the Convention emphasized that such measures are intended to accelerate the participation of women in the political, economic, social, cultural and civil spheres under conditions of equality. These measures may consist of outreach and support programs, reallo-

cation of resources, preferential treatment, determination of hiring and promotion goals, and quota systems (*Assembly*, 1979, *General Recommendation* 25, § 22).

The Committee for the Elimination of all Forms of Discrimination against Women warned that the States parties to the Convention must guarantee through the competent courts and the imposition of sanctions or other forms of reparation, the protection of women against discrimination committed both by public authorities and by organisations, companies and individuals (*Assembly*, 1979, art. 4). It also recommended that States should make greater use of temporary special measures in matters of employment aimed at achieving equality (*Assembly*, 1979, *General Recommendation* 5, § 18).

Stereotypes and practices that devalue the feminine are found not only in real life but also in virtual life (*Consejo Nacional para prevenir la Discriminación*, 2016). The reinforcement of stereotypes and consequent deepening of silent discrimination against women is accentuated by the use of AI. For example, a algorithm-led search for cooking-related activity produces 33% more women than men in a normal internet search. If we add to the same activity, the training of the program from a continuous search, the figure grows from 33% to 68%. Researchers from Cornell University set out to correct this type of algorithm, but only with the hope of maintaining the deviation from the initial stage, since according to the current state of the art, it is not possible to correct it. To achieve this end, the authors designed an algorithm based on Lagrangean relaxation for collective inference (Zhao, Wang, Yatskar, Ordonez, & Chang, 2017).

Second, the State and individuals are obliged to adopt positive action measures

to counteract gender segregation, as well as to reverse the socio-cultural patterns that structure segregation. Relevant human rights treaties expressly prohibit discrimination based on gender, economic position and origin, or any other social condition². O'Neil (2018) refers to the setback suffered by Amazon when trying to hire staff through a learning machine. After testing the program, Amazon found that it only repeated the male bias of the technology industry to the detriment of women and other dissidences.

Third, a person's sexual orientation cannot constitute sufficient ground for restricting a right (*Inter-American Court of Human Rights, Judgment of February 24, 2012*).

In 2009, Amazon removed 57,310 books from its ranking of best sellers, after an algorithmic change flagged as "adult content" books that dealt with sexuality issues (basically gay and lesbian issues). These titles disappeared from the site until the reason was known, upon which blame was apportioned to algorithms and the titles returned to their former ranking (*Kafka, 2009*).

Fourth, people with disabilities shall enjoy the guarantee of the effective enjoyment of rights on equal terms with any other. For this to be possible, certain "reasonable adjustments" should be carried out to software programs. The new social model of disability implies making such adjustments and providing technical support, so that people with disabilities can fully exercise their rights. The so-called 'reasonable accommodations', according to the language used by the Convention on the Rights of Persons with Dis-

abilities (*The United Nations, 2006, art. 2*), are those necessary and adequate adaptations that do not impose a disproportionate or undue burden, when required in a particular case, in order to guarantee people with disability the enjoyment or exercise of all fundamental rights on an equal basis with the others.

Notwithstanding what has been said, software can also passively discriminate against certain people. For example, the algorithms of autonomous cars are trained to know what pedestrians look like so as not to run over them. If the training dataset does not include people in wheelchairs, the technology could become a life-threatening hazard. Algorithmic fairness to people with disabilities is a different problem than fairness to other vulnerable groups such as those based on race or gender. Many systems consider race or gender as simple variables with a small number of possible values. But as regards disability, there are several forms and grades of severity. Some are permanent, others are temporary. Thus, it is a dynamic group. Privacy of information and sensitive data are interrelated here. The first thing that comes into one's mind is that if the program does not know the user's disability, it will not discriminate against him/her. But in relation to disabilities, this is not the case. Information on disability should be provided as it often involves necessary information. Let's take the example of a person with visual impairment, who needs to use a screen reader to access the internet and takes an online test to access a job. If the test program is not

² Among the main documents we find articles 1.1, American Convention on Human Rights; 2, 3 and 26, International Covenant on Civil and Political Rights; 2 and 3, International Covenant on Economic, Social and Cultural Rights; 2, American Declaration of the Rights and Duties of Man; 2, Universal Declaration of Human Rights; and Convention on the Elimination of All Forms of Discrimination against Women.

well designed and is not accessible to the applicant, it will take longer for the applicant to navigate the page and answer the questions. Thus, people with a similar disability will face a systemic disadvantage (Hao, 2018).

Fifth, the right to freedom of religion and conscience encompasses, among other aspects, the right not to be discriminated against for one's religious beliefs. The Supreme Court of Justice of Argentina stated that:

'[freedom of religion and conscience is] (...) a particularly valuable right that includes respect for those who hold religious beliefs and for those who do not hold them' (Supreme Court of Justice of Argentina, Sentences 312:496).

According to the Argentine Constitution and other relevant international documents ratified by Argentina, freedom of religion and conscience has different aspects: the freedom to possess/not possess beliefs of one's own choice without suffering external interference, the right not to be discriminated against for religious beliefs, and the freedom to be educated according to one's own convictions.

In 2019, Facebook faced legal proceedings initiated by the US government for allowing advertisers to deliberately target advertising based on religion, race, and gender. Using this strategy, companies excluded people of a certain race, age, or gender from viewing housing advertisements, in violation of the Fair Housing Act. In another case, a group calling themselves "enlightened souls" that publish content related to spirituality, ancient practices, the worship of goddesses, etc., became a victim of biased Facebook ads. This occurred when the social network, which uses targeting algorithms, removed an ad that contained images of the goddess 'Kali' along with other goddesses, erroneously labelled as sexual content (*E-Hacking News*, 2020).

Finally, with regard to discrimination for reasons different than those expressed, when the existence of a discriminatory circumstance is alleged, it is up to the defendant to prove that the allegedly discriminatory act was caused by an objective and reasonable motive, unrelated to any discrimination:

'...In cases in which Law 23,592 is applicable, and the existence of a discriminatory motive is disputed (...), it will be sufficient, for the party that affirms said motive, the accreditation of facts, *prima facie* evaluated, that result suitable to induce their existence, in which case the defendant who is accused of committing the contested treatment will be the proof that it was caused by an objective and reasonable motive unrelated to any discrimination (...)' (Supreme Court of Justice of Argentina, Sentences 334:1387).

Digital discrimination and responsible algorithms

We have seen that in certain cases programmers transfer their biases (even involuntarily) to the algorithms of the programs they create. The automatic training tools of a computer system expose it to the assimilation of a large and relevant amount of data, so that the program learns to make judgments or predictions about the information it processes based on the observed patterns. In a simple example, if someone wants to train a computer system to recognise whether an object is a book based on certain factors (e.g., texture, weight), such factors are provided to the system, and the software is programmed in order to recognise in which case the objects are books and in which they are not. After multiple tests, the system is supposed to learn what a book is and be able to predict without human help wheth-

er a certain object is/not a book, depending on the data received.

It has been demonstrated that when scientific or technological decisions are based on a limited set of systemic, structural or social concepts and norms, the created technology can privilege certain social groups and harm others. AI models are determined by different biases that reproduce and sometimes amplify the power relations that underlie reality. The examples cited above exemplify this statement.

We should also discuss so-called “sexist” or “racist” algorithms. In terms of everyday applications to complex algorithms, Ruha Benjamin (2019) describes how emerging technologies can reinforce “white supremacy” to deepen social inequity, arguing that automation has the potential to conceal, accelerate, and deepen discrimination. A similar issue arises in the application of AI in areas of criminal justice. In 2016, an investigation on judicial software conducted by a non-governmental organisation called ProPublica revealed that algorithms used by US law enforcement agencies erroneously predict that black defendants are more likely to commit repeat offenses than white defendants with similar criminal records (Angwin, Larson, Mattu, & Kirchner, 2016).

Noble (2019) presents the idea that search engines like Google offer an ideal playing field for discrimination of ideas, identities and activities. Considering data discrimination as a genuine social problem, Noble (2019) argues that the combination of private interests in promoting certain sites, together with the status of a quasi monopoly enjoyed by a relatively small number of Internet search engines, leads to a skewed set of search algorithms that privilege “white” people and discriminate against people of colour, especially

black women. Through an analysis of textual and media searches, as well as extensive research on paid online advertising, Noble (2019) exposes a culture of racism and sexism that is present in the way online search ability is built.

Fair or at least non-discriminatory search tools are scarce today. Whether searching for a job, applying from a university course, or predicting inmate recidivism, clear examples of discrimination through algorithms can be identified. In these situations, the AI used by search engines favours discriminatory patterns, generated by algorithms that are not programmed to compensate or correct human prejudices (Gomez Abajo, 2017) and consequently end up reinforcing them.

Amid discussions of algorithmic biases, companies using AI claim to be taking steps to use more representative training data, while regularly auditing their systems for unwanted biases and the eventual negative impact against certain groups. Harvard researcher Lily Hu says this is no guarantee that their systems will work fairly in the future (Heilweil, 2020). While there is not a wide range of studies on the demographics of AI, currently the sector tends to be male-dominated, and the high-tech sector tends to over-represent whites, according to the US Equal Opportunity in Employment Commission.

At this point, we should consider the relation between non-discrimination and affirmative action. Private companies and other users of algorithms are not legally obliged to take affirmative actions in the benefit of vulnerable groups, unless there are laws which bind them, even though they may stated it to be a desirable goal. The principle of non-discrimination is consistent with affirmative action (Lawrence III, 2001). However, since they are designed for the specific inter-

ests of the user, algorithms may have different aims. This is where the legislator or the industry itself through must intervene in terms of applying the law to impose the necessary trade-offs to balance society's varying goals. As an example, some authors comment on the case of a university admissions procedure that priorities the admission of the most talented students but, at the same time, aims to represent a degree of diversity according to the composition of society. In this case, algorithms would help to obtain and compensate both goals at the same time, uniting private and public interests (Kleinberg, Ludwig, Mullainathan, & Sunstein, 2018).

When we talk about users of AI, another important concept that is worth mentioning is explainable AI. When it comes to explain any algorithmic decision-making process, it is not enough to observe the classification result given one instance. To fully understand an automated decision, the whole process needs to be taken into account. If not, the explanation might not be representative and might not reflect all input factors and parameters which led to a particular decision (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). A possible definition for explanation in the context of ADM could be:

'A formal and unambiguous descriptive representation of the output of a classifier based on the current and prior input represented as parameters' (Waltl & Vogl, 2018).

Explainable AI looks for methods useful for analysing and/or complementing AI models with the aim to make the internal logic and output of algorithms transparent and easy to monitor and, eventually, correct, making these processes humanly understandable and meaningful. Gunning and Aha (2019) provide a basic set of questions in order to help assessing algorithmic decision mak-

ing. The question concerns how to correct errors. They can be considered as guidelines which provide a stronger structure to the development of this kind of decision-making and improve their intrinsic explainability. On this basis, algorithms can be built to provide explanations for why specific instances or entire classes were classified in the way they were. This would greatly help to satisfy the need for more algorithmic transparency.

Conclusion

As search engines and their related businesses grow in importance, operating as information sources, social communities, and – especially in a time of pandemic – acting as vehicles of learning at all levels, the increasing threat requires understanding to reverse discriminatory practices. Discrimination does not only reveal itself in the form of violation of norms that prohibit some practices, but, more importantly, as obligations to take action in order to improve the situation of those vulnerable groups which are inherently unequal (so-called positive discrimination). No amount of increased accuracy or efficiency that a particular AI may add to the final result compensates for a model that is unfair or unethical. For example, the Chinese government uses AI to track its Uighur Muslim minority in Xinjiang, of whom around one million are believed to be living in re-education camps.

Of the various options that exist to counteract discrimination through algorithms, preventive measures appear to be the most reasonable. Businesses can seek assistance from anti-discrimination agencies to educate their staff and IT experts and raise awareness. These people will then use data sets that do not reflect discriminatory practices or unequal treatment. The goal is to make future

algorithms “free from discrimination by design”. This circumstance implies that the programs are constantly verified along their initial development and then are continuously monitored. In any case, many authors affirm that defining justice in a mathematically rigorous way is very difficult, if not impossible (Angwin *et al.*, 2016).

Nevertheless, if we cannot make mathematics fair, at least we can make it less opaque. Therefore, we affirm that, in relation to AI models and algorithmic decision-making, transparency is paramount. If these processes become more transparent, their explainability would increase. Any algorithmic decision-making process should be able to explain its inputs, outputs and results, as well as be prepared to correct undesired er-

rors. Moreover, it will also help to optimise these systems and understand the boundaries of AI and to assign responsibilities when we get an unwanted result.

Ultimately, this article deals with something more important than algorithmic bias. It is about the protection of supreme values of societies that respect human rights, such as equality and free development of the personality. Considering the rapid developments of big data and AI, it is necessary to urgently improve anti-discrimination legislation and data protection. These actions will help to eliminate or at least minimise algorithmic bias. Additional research should be carried out on the programmers themselves, including procedures for their selection and / or training.

REFERENCES

1. Angwin, J., Larson, J., Mattu, S. & Kirchner, L. (2016, May 23). Machine Bias. *ProPublica*. Available at: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
2. Anti-Discrimination Law, No. 23, 592, art. I (1988).
3. Argentine Const., art. 16, 19, 22, 23, 37, 75.
4. United Nations General Assembly (1979). *Convention on the elimination of all forms of discrimination against women*. Available at: <https://www.un.org/womenwatch/daw/cedaw/cedaw.htm>
5. Banaji, M. R. & Greenwald, A. G. (2013). *Hidden Biases of Good People*. New York: Bantam.
6. Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Wiley, Princeton University, USA.
7. Consejo Nacional para prevenir la Discriminación. (2016). *Ficha temática Mujeres, México*. Available at: https://www.conapred.org.mx/userfiles/files/FichaTematica_Mujeres.pdf
8. Agencia de los Derechos Fundamentales de la Unión Europea y Consejo de Europa. (2019). *Manual de legislación europea contra la discriminación: Edición de 2018*. Available at: http://repositori.uji.es/xmlui/bitstream/handle/10234/187708/manual_agencia_2019.pdf?sequence=1
9. Hacking and Cyber Security News. (2020, May 31). Religion Biased Algorithms Continue to Depict How Facebook Doesn't Believe in Free Speech. Available at: <https://hackingncysecnews.blogspot.com/2020/05/religion-biased-algorithms-continue-to.html>
10. Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–37. <https://doi.org/10.1609/aimag.v17i3.1230>

11. Gomez Abajo, C. (2017, August 28). La inteligencia artificial tiene prejuicios, pero se pueden corregir. *El Pais*. Available at: https://elpais.com/retina/2017/08/25/tendencias/1503671184_739399.html
12. Gunning, D. & Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI Magazine*, 40(2), 44–58. <https://doi.org/10.1609/aimag.v40i2.2850>
13. Hao, K. (2018, November 28). Can you make an AI that isn't ableist? *MIT Technology Review*. Available at: <https://www.technologyreview.com/2018/11/28/1797/can-you-make-an-ai-that-isnt-ableist/>
14. Heilweil, R. (2020, February 18). Why algorithms can be racist and sexist. *Vox*. Available at: <https://www.vox.com/recode/2020/2/18/21121286/algorithms-bias-discrimination-facial-recognition-transparency>
15. Inter American Court of Human Rights, Proposal to amend the Political Constitution of Costa Rica related to naturalization, Advisory Opinion OC-4/84 of 01/19/1984, Series A, No.4, Chapter IV, § 56-58.
16. Inter-American Court of Human Rights, Consultative Opinion 18/03, § 4.
17. Inter-American Court of Human Rights. *Case of Atala Riffo and daughters v. Chile. Judgment of February 24, 2012*. (Merits, Reparations and Costs). Available at: https://www.humanrightstrust.org/wp-content/uploads/resources/Atala_Riffo_and_Daughters_v_Chile_24_February_2012_Series_C_No._239.pdf
18. Kafka, P. (2009, April 13). Amazon Apologizes for 'Ham-fisted Cataloging Error'. *All Things D*. Available at: <https://allthingsd.com/20090413/amazon-apologizes-for-ham-fisted-cataloging-error/>
19. Karlsruhe Institute of Technology. (2019, November 13). The risk of discrimination by algorithm. *Tech Xplore*. Available at: <https://techxplore.com/news/2019-11-discrimination-algorithm.html>
20. Kleinberg, J., Ludwig, J., Mullainathan, S. & Sunstein, C. R. (2018). Discrimination in the Age of Algorithms. *Journal of Legal Analysis*, 10, 113–174. <https://doi.org/10.1093/jla/laz001>
21. Law No. 5261, Ciudad Autónoma de Buenos Aires, *Buenos Aires* (2015).
22. Lawrence III, C. R. (2001). Two View of the River: A Critique of the Liberal Defense of Affirmative Action. *Columbia Law Review*, 101, 928. Available at: <https://scholarship.law.georgetown.edu/cgi/viewcontent.cgi?article=1339&context=facpub>
23. Mayson, S. G. (2018). Bias in, bias out. *The Yale Law Journal*, 128, 2218–2230. Available at: <https://ssrn.com/abstract=3257004> (accessed 23 October 2021).
24. Metz, R. (2018, March 27). Microsoft's neo-Nazi sexbot was a great lesson for makers of AI assistants. *MIT Technology Review*. Available at: <https://www.technologyreview.com/2018/03/27/144290/microsofts-neo-nazi-sexbot-was-a-great-lesson-for-makers-of-ai-assistants/>
25. MPF Argentina, Igualdad y no Discriminación, Dictámenes del Ministerio Público Fiscal ante la Corte Suprema de Justicia de la Nación (2012-2016). Available at: <https://www.mpf.gob.ar/dgdh/files/2016/06/Cuadernillo-2-Igualdad-y-no-Discriminaci%C3%B3n.pdf> (Spanish).
26. Mullane, M. (2018, June 15). Eliminating bias from the data used to train algorithms is a key challenge for the future of machine learning. *E-tech*. Available at: <https://etech.iec.ch/issue/2018-06/eliminating-bias-from-algorithms>
27. Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press. <https://doi.org/10.2307/j.ctt1pwt9w5>
28. O'Neil, C. (2018). Amazon's gender-biased algorithm is not alone. *Bloomberg*. Available at: <https://www.bloomberg.com/opinion/articles/2018-10-16/amazon-s-gender-biased-algorithm-is-not-alone>

29. Supreme Court of Justice of Argentina, Sentences 312:496, *Portillo case*, § 8, 9, 10.
30. Supreme Court of Justice of Argentina, Sentences 334:1387, *Pellicori case*.
31. Supreme Court of Justice, Argentina, Sentences 153:67.
32. Supreme Court of Justice, Argentina, Sentences 16:118.
33. Supreme Court of Justice, Argentina, Sentences 314:1531 and ss.
34. The United Nations. (2008). Convention on the Rights of Persons with Disabilities. New York, 13 December 2006. *Treaty Series*, 2515, 3. Available at: <https://treaties.un.org/doc/publication/unts/volume%202515/v2515.pdf>
35. Wärtl, B. & Vogl, R. (2018). Explainable Artificial Intelligence: The New Frontier in Legal Informatics. *Jusletter IT*, 4, 1–10.
36. Zhao, J., Wang, T., Yatskar, M., Ordonez, V. & Chang, K. W. (2017). Men also like shopping: Reducing gender bias amplification using corpus-level constraints. *arXiv* [Preprint]. Available at: <https://arxiv.org/abs/1707.09457>. <https://doi.org/10.48550/arXiv.1707.09457>

AUTHOR

Favio Farinella, Director of the Research Centre of International Law at Mar del Plata National University, Buenos Aires, Argentina

ORCID: <https://orcid.org/0000-0002-6610-3235>