


## CONTROVERSIES ON ALGORITHMIC HARMS: corporate discourses on coded discrimination

CONTROVÉRSIAS SOBRE DANOS ALGORÍTMICOS: discursos corporativos sobre discriminação codificada  
CONTROVERSIAS SOBRE DAÑO ALGORÍTMICO: discursos corporativos sobre discriminación codificada

### Sergio Amadeu da Silveira

PhD and Master in Political Science from the University of São Paulo (USP). Professor at the Federal University of ABC (UFABC). Researcher at CNPq / Research Productivity - 2. [sergio.amadeu@ufabc.edu.br](mailto:sergio.amadeu@ufabc.edu.br).

 0000-0003-1029-9133

### Tarcizio Roberto da Silva

PhD Student at the Social Sciences and Humanities program at Federal University of ABC and Master in Communication (Federal University of Bahia). [eu@tarciziosilva.com.br](mailto:eu@tarciziosilva.com.br).

 0000-0002-7094-8708

Mailing address: Universidade Federal do ABC (UFABC). Avenida dos Estados, 5001, 09210-580 - Bangú, Santo André, SP – Brasil

Received: 04.03.2020.  
Accepted: 05.27.2020.  
Published: 07.01.2020.

### ABSTRACT:

Discriminatory impacts and the damages due to algorithmic systems have opened discussions regarding the scope of responsibility of communication technology and artificial intelligence companies. The article presents public controversies triggered by eight public cases of harm and algorithmic discrimination that generated public responses from technology companies, addressing the efforts made by them in framing the debate about responsibility in the course of planning, training and implementation of systems. Following that, it discusses how the opacity of systems is defended by the commercial companies that develop them, alleging prerogatives such as “industry secrets” and algorithmic inscrutability.

**KEYWORDS:** Algorithms; Algorithmic Auditing; Explainability; Technology Journalism; Platforms.

## Introduction

### Algorithmic systems, explainability and responsibility

Algorithms never act in isolation (SEEVER, 2019; SILVEIRA, 2019). Defined, in general, as a collection of instructions or rules to solve a problem or perform a task, they need to be in contact with a data structure in order to act. Algorithms are part of a network of actors (LATOUR, 2005). Their connections with incoming data, with feedback, with the effects of their own decisions and other system components that implement them must be considered. As such, we use the expression “algorithmic systems” in this text.

These systems may be designed to follow rules for how to perform their actions based on the information they receive. They can be created to learn from the data they receive in the performance of their prescribed objectives. They can also have the purpose of finding strong correlations in the data they receive. Ultimately, they may create their operations based on data rather than fixed rules.

So-called machine learning systems utilize innumerable computational models, among which, the model of neural networks that has obtained great success in various areas, such as robotics, medical diagnosis, voice recognition, biometrics, data mining, automatic target recognition, among many other applications. As with other systems that learn from data, artificial neural networks act where rule-based programming has not performed well. Inspired by the central nervous system, they seek to simulate the action of neurons.

The success of so-called Artificial Intelligence (AI), which involves models of artificial neural networks, deep learning, graphical probabilistic models, as well as others, is due to its high performance (GUNNING; AHA, 2019). This performance in the treatment of data, detecting patterns and making predictions has been useful in the advance of competitiveness in a scenario dominated by neoliberal economic doctrine. As such, the business models based on collection, storage and analysis of consumer data for the purpose of predictive uses has incentivized and increased the utilization of deep-learning algorithmic systems.

Frank Pasquale (2015) demonstrated that this informational process occurs opaquely. As he alerted us in the book *The Black Box Society*, the opacity of algorithmic systems is defended as indispensable to protect business secrets and the intellectual property of code, and to avoid users defeating the purpose of those systems. As such, transparency is seen as an obstacle by large corporations; therefore, companies, consultants and technology platforms consider it fundamental for the “improvement of the experience” of users, clients and consumers. This being the case, people are convinced that their personal data will be in good hands if given to private companies.

The life of each individual is being converted into an immense flow of data, as the statistical prediction models and their algorithms demand a great quantity and variety of data to extract patterns and create forecasts. The logic of competition is an energizer which turns the data market into an expanding ecosystem, aggregating new data-generating devices into its network of actors. Jose Van Dijck warned about the double alienation that society will bring about in this process. First, the belief that data is natural and expresses reality. Second, the belief that data platforms are, like the data itself, neutral (VAN DIJCK, 2014).

The fact is that even with the existence of open AI frameworks, the majority of algorithmic systems of great public relevance (GILLESPIE, 2015) are closed, opaque, with no transparency at all. It is sufficient to remember the most-used search engine on the

planet, Google Search. It is a closed algorithmic system. The same applies to the algorithmic system of Facebook and all other digital intermediary platforms.

There is a relationship between the lack of transparency of algorithmic systems and processes discriminatory to people and population segments when submitted to governance practiced by algorithms. This is why there are movements for the transparency of code and the recognition that algorithmic systems possess bias, with pre-existing definitions embedded in their models (DIAKOPOULOS, 2014). It is common to hear claims that deviations and biases are not in the algorithms but in the databases, or better, in the collected data. This does not appear consistent with information found on the development of various algorithmic systems that have clear objectives of seeking differentiation in physical features, in certain behaviors, in residential areas, in schools attended, etc.

In 2019, the city council of San Francisco, California prohibited the use of facial-recognition technologies by the police and other public agencies (FRANCE PRESSE, 2019, electronic text). The principal argument is that the risks to civil rights and liberties outweighed the possible benefits. Besides that, the decision of the council argued that facial recognition could “exacerbate racial injustice and threaten our ability to live free of continuous government monitoring.” (FRANCE PRESSE, 2019, electronic text). The debate turned on the risks of algorithmic systems to persecute and discriminate against minorities and socially marginalized groups.

The transparency of algorithmic systems may not solve the problem of explaining how they arrive at certain results, many of them prejudiced, racist and discriminatory. In some AI models, of deep learning, for example, such as artificial neural networks, the form in which the algorithm acts does not permit the explanation of its processes, the steps it takes which result in a given decision. They are algorithmic models, considered inscrutable, unfathomable or incomprehensible.

The U.S. Department of Defense faced the problem of explainability and the understanding of how a deep-learning system offers a certain action plan for using intelligent systems in actions of national defense. This was the main reason that DARPA (Defense Advanced Research Projects Agency) created the XAI program—Explainable Artificial Intelligence. The objective of XAI is to create a collection of machine-learning techniques that produce explainable models, maintaining a high level of learning performance, as well as permitting people to understand, trust and effectively administer these algorithmic systems (GUNNING, 2016).

This question takes on a significant sociotechnical or technopolitical dimension when we recognize that there are algorithmic models and systems that can find solutions or propose decisions of great social relevance without their administrators or even developers knowing exactly what procedures or calculations were performed to achieve that result. Even being commercial solutions acquired by private corporations, in general, liberal democracies typically have consumer defense laws that demand explanations and responsibility for the decisions adopted by companies.

The European General Data Protection Regulation provides the right of explanation and human review of automated decisions, principally to avoid a possible business or governmental allegation that an algorithmic system does not permit knowledge of the reasons for certain actions. It is likely that to confront socially and democratically unacceptable racism and discrimination, it is necessary that algorithmic systems be transparent, explainable and supervised by those responsible for quickly reconfiguring them. It seems that with the advance of algorithmic systems, the risks of segregation, exclusion and marginalization may increase with the argument of a certain systemic neutrality and objectivity that hides decisions embedded in code, or biases included in databases.

### **Algorithmic harm as controversy: public hearings and civil engagement**

Part of the civil mobilization on possible algorithmic damage has been carried out in the civil society through expedients such as public audits and articles based on investigative journalism or reports offered by system users. In this article we cite, as empirical evidence, eight notes of cases that involved public repercussion and declaration of the organizations involved through resources such as press releases or public statements, listed in Table 1. Before considering corporate reactions in the following section, we will present in this section the concept of algorithmic auditing and some of the public repercussion cases analyzed later.

Sandvig and collaborators (2014) propose a methodology inspired by the Auditing Studies to propose a set of five possible approaches to auditing algorithmic systems: non-invasive user auditing; sock-puppet auditing; crowdsourced auditing; code auditing; and scraping. *Noninvasive Auditing* is, basically, the adaptation of classical social science methods such as deep interviewing, surveys or non-participative observation of normal user interactions to investigate the behaviors, dynamics and perceptions of users within the studied systems. Being a "non-invasive selection of information about users' normal interactions with a platform" (SANDVIG *et al.*, 2014, p.11), journalistic reporting based on

the consultation of users approaches the model. This is the case of recurring problems with YouTube's recommendation algorithms when analyzed regarding children's videos. Reports from the New York Times<sup>1</sup> and Wired<sup>2</sup> discovered in 2017 and 2019, respectively (see notes 4 and 5), that disturbing videos of violent, scatological cartoons simulated juvenile content in order to be viewed by children, dodging the platform's automatic filters, and that a network of pedophiles used the platform's recommendations to access videos of semi-nude, dancing children.

Very similarly, a second approach can involve the construction of *Crowdsourced* or *Collaborative* systems to evaluate some points through usage, reporting or distributed code. Technically and financially more complex, one example is the *FeedVis* project, developed by Eslami and collaborators (2015). Through the development of an application for Facebook that analyzes, with consent of participants, timeline data that researchers could use to compare the interference of the algorithmic interference of Facebook in the interpersonal interactions on the platform, it was found that participants were "attributing the algorithm's actions to be the intent of their own friends and family. Users incorrectly concluded that they held unpopular views or were being given the cold shoulder" (ESLAMI *et al.*, 2015, p. 9), which reinforces the thesis of the influence of the platform on interpersonal distancing.

A third proposed approach is called the *Sock-Puppet Audit* (Sandvig *et al.*, 2014) and involves the simulation of users with variables controlled by the parameters of the study or even employing bot users. In one of the publicly documented cases, based on user reports of racial discrimination in the lodging booking platform Airbnb, the California Department of Fair Employment and Housing audited the platform using simulated accounts with a variety of demographic characteristics<sup>3</sup>.

Regarding the analysis of system aspects seen as strictly technical, a *Scraping Audit* encompasses the collection of system data, including data scraping techniques, access through APIs, screen captures and so on. When dealing with systems focused on communication (such as social media platforms and search engines) or with self-administration user interfaces (such as recruitment forms, credit-score tools and so on), this approach is often used to allow evaluation of the results and requests offered to the users. The tactic is to collect and analyze platform data through usage simulations or

---

<sup>1</sup> Link: <https://www.nytimes.com/2017/11/04/business/media/youtube-kids-paw-patrol.html>.

<sup>2</sup> Link: <https://www.wired.co.uk/article/youtube-pedophile-videos-advertising>.

<sup>3</sup> Link: <https://www.usatoday.com/story/tech/news/2016/06/06/airbnb-openair-diversity-racism-airbnb-connect/85490536/>.

interactions at scale, in a way distinct from earlier methods in “accessing the platform directly via an API or they may be making queries that it is unlikely a user would ever make (or at least at a frequency a user is unlikely to ever make)” (SANDVIG *et al.*, 2014, p. 12). Recent investigations of the methods by which characteristics of the YouTube interface and algorithms promote extremist channels, especially from the right, follow this path through the analysis of recommendation networks among videos and channels (RIBEIRO *et al.*, 2019; RIEDER *et al.*, 2018).

*Code Audit*, through which code effectively incorporates decision chains, methodological choices, datasets, packages and programming modules, is typically the most recommended. It is the most difficult to apply due to institutional hurdles (most platforms have closed code for commercial and competitive reasons) as well as technical limitations (the myriad of technologies far exceeds the capacity of lone researchers). As such, “even given the specific details of an algorithm, at the normal level of complexity at which these systems operate an algorithm cannot be interpreted just by reading it (SANDVIG *et al.*, 2014, p. 10), but a deep knowledge of the technical processes involved in a given system permits researchers to attack the roots of problems through the same computational logic, but with sensitivity to the possible algorithmic harm according to demographic variability and diversity of uses.

Among the uses that mix scraping and code auditing techniques with high-impact investigation, the series of studies in the *Gender Shades* project of the *Algorithmic Justice League* deserves special mention. The researchers analyze the precision of gender and age identification resources in facial recognition in three of the principal technologies on the market, from the companies IBM, Microsoft and Face++. An intersectional inequality was discovered: the systems fail more often with dark skinned people, resulting in enormous error rates in photos of dark skinned women, having wide impact in applications of social media and police surveillance. Besides identifying the root of the problem—above all the uncritical use of biased training data—the researchers identified that the “intersectional phenotypic and demographic error analysis can help inform methods to improve dataset composition, feature selection, and neural network architectures” (BUOLAMWINI; GEBRU, 2018, p.12). Beyond the scientific merit of the academic text, the publication of an interactive site<sup>4</sup> has been essential to the project, as well as public presentations of the data, generating media coverage and public interest in the findings.

---

<sup>4</sup> <http://gendershades.org/>.

Turning the question from a “matter of fact” into a “matter of concern” (LATOURE, 2004), the public impact of the discoveries forced the companies involved to publicly pronounce, through public postings and commitments to improving the systems. In subsequent work (RAJI; BUOLAMWINI, 2019), the researchers of the *Gender Shades* project reviewed the error rates in the analyzed systems, identifying effective improvements, and compared them with two other providers, Amazon and Kairos. In the project trajectory described by the authors, after the problem identification phase, the involved companies are offered advance knowledge of the study and a period to respond, before the public release of the results. After this period, the results are shared at scientific conferences, with the press, and in the case of Gender Shades, on an interactive website—which later included the corporate responses. Upon returning to the data and identifying the reduction in error rates, the authors proposed the concept of “actionable public auditing” as “one mechanism to incentivize corporations to address the algorithmic bias present in data-centric technologies” (RAJI; BUOLAMWINI, 2019, p. 1).

Factors such as explainability and responsibility for algorithmic harm are still controversial and dependent on considerable networks of regulatory and legislative power flow, so proposals like that of Pasquale of considering specialized, mandatory intermediaries as a possibility of acting toward algorithmic governance through prior regulation of systems with the force of a “fourth law of robotics” (PASQUALE, 2017) seem distant still. Data from NeurIPS, the largest artificial intelligence and neural network event in the world, shows that the number of papers with new proposals exceeds by ten times the number of papers analyzing existing models, demonstrating a knowledge gap regarding algorithmic systems (EPSTEIN *et al.*, 2018).

Somewhat systematic public reports or audits about algorithmic harm can force corporations to come forward regarding their responsibilities, through the power of public pressure and the press. Bucher alerts that the performative character of algorithms, as well as discussion about them constructed upon the interfaces between their use, public opinion, the press and civic engagement about them create an “algorithmic imaginary” which would be the “ways of thinking about what algorithms are, what they should be, how they function and what these imaginations in turn make possible” (BUCHER, 2016a, p. 39-40).

The public space for communication and journalism, beyond the exchanges of experts, is therefore a fertile source of investigation of the discursive strategies in search of corporate context in cases of publication of algorithmic data. In the following section,

we will review documented cases of harm in which the involved corporations reacted publicly to their discovered errors.

### **Corporate reactions: evading responsibility**

In the panorama of relations resulting from the dissemination of algorithmic systems in the social spheres, the paradigm of the invisibility of the functioning of the systems is the result of their integration into everyday life. In moments of publication of algorithmic damages, the systems become "matters of interest" (LATOIR, 2004) opening the controversy about the neutrality or objectivity of technology already incorporated in the daily life. The public and press releases that we will analyze below are part of the organized effort, through public relations techniques and organizational communication management, to present the case so as to minimize damage to the perception of positive values of the organization or its technologies.

Bucher criticizes the concept of the algorithmic "black box" when it is articulated only as a question of investigation of the inputs and outputs of a system, once those systems, increasingly constructed to adapt calculations and procedures through machine learning reconfigure the status of input and output (2016b). It is possible to expand the scope of observation of algorithms beyond their immediate, apparent functions, in search of social networks and power relationships, materialized or intermediated in the flows of performance (INTRONA, 2015).

We can approximate Bucher's concept of "technography," created to approach the manner by which software intersects with social behavior, "the norms and values that have been delegated to and materialized in technology" (2016b, p.86) with the approach proposed by Brock of Critical Technoculture Discourse Analysis. For Brock, part of the principles of the analysis of discourse: that it shows recurring patterns; that it involves choices of the sender; and that the discourse mediated by computers, as with other media and formats, can be molded and adapted to environmental characteristics (BROCK, 2016).

To understand the corporate efforts in the framing of cases of algorithmic harm, we have selected eight releases and statements from corporations and their representatives, as seen in Table 1.



**Table 1** Press Releases / Public Statements Analyzed

Number	Public Release / Reporting	Company Involved	Year
1	MICROSOFT JANUARY 2018 STATEMENT to lead author of "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification" <sup>5</sup>	Microsoft	2018
2	IBM Response to "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification" <sup>6</sup>	IBM	2018
3	FaceApp apologises for 'racist' filter that lightens users' skintone <sup>7</sup>	Faceapp	2017
4	On YouTube, a network of paedophiles is hiding in plain sight <sup>8</sup>	Youtube	2019
5	On YouTube Kids, Startling Videos Slip Past Filter <sup>9</sup>	Youtube	2017
6	YouTube won't stop recommending videos with children, despite pedophilia problem <sup>10</sup>	Youtube	2019
7	Google tweaks algorithm to show less porn when searching for 'lesbian' content <sup>11</sup>	Google	2019
8	Search for 'tranças bonitas' ('beautiful braids') and 'tranças feias' ('ugly braids') on Google: a case of algorithmic racism <sup>12</sup>	Google	2019

The texts listed in Table 1 were analyzed below within their sociotechnical context, "as a communicative process, by unpacking what a specific ICT artifact is based upon, what it is designed to 'do,' and critically, how users articulate themselves in and about the artifact" (BROCK, 2016, p. 15). Under the light of concepts of explainability and inscrutability, we explore this group of statements to propose the categories below.

### *Continual Process of Optimization*

The idea of the *perpetual beta* came into being in communication technologies in the form of platforms and software as devices and bandwidth advanced in technical quality and efficiency, allowing the paradigm of "software as a service" (SaaS) to dominate the offering of products in recent years (ROMANI & KUKLINSKI, 2007), culminating in the emergence of mobile applications and platformization (SRNICEK, 2017). Besides being a manner of approaching software development, it has also become a commercial tactic: by presenting informational products as *betas*, the corporations at the same time

<sup>5</sup> <http://gendershades.org/docs/ibm.pdf>.

<sup>6</sup> <http://gendershades.org/docs/lbm.pdf>.

<sup>7</sup> <https://www.mirror.co.uk/tech/faceapp-apologises-hot-selfie-filter-10293590>.

<sup>8</sup> <https://www.wired.co.uk/article/youtube-pedophile-videos-advertising>.

<sup>9</sup> <https://www.nytimes.com/2017/11/04/business/media/youtube-kids-paw-patrol.html>.

<sup>10</sup> <https://www.theverge.com/2019/6/3/18650318/youtube-child-predator-pedophilia-family-vlogging-comments-recommendation-algorithm>.

<sup>11</sup> <https://thenextweb.com/tech/2019/08/07/google-tweaks-algorithm-to-show-less-porn-when-searching-for-lesbian-content/>.

<sup>12</sup> <https://blogs.oglobo.globo.com/ancelmo/post/pesquisa-trancas-bonitas-e-trancas-feias-no-google-um-caso-de-racismo-algoritmico.html>.

cultivated an aura of creative innovation while managing expectations regarding *bugs*, or system failures.

Currently the main suppliers of technology have abandoned the term beta as a qualifier of their monetizable products, but resuscitate the principle as needed. Two cases reported about YouTube regarding the recommendation of problematic content were approached via the tactic of evoking continual improvements or the perpetual beta. As a response to the case of recommendations of videos of children to pedophiles, cited in the previous section, YouTube declared in an official statement in its blog that “over the last 2+ years, we’ve been making regular improvements to the machine learning classifier that helps us protect minors and families. We rolled out our most recent improvement earlier this month” (Note 6). Continuous optimization as a game of cat and mouse, resulting from the excessive complexity of web content production, was also mentioned by Malik Ducard, responsible for supervising family and educational content on the platform. In a case from 2017, Ducard indicated machine learning as a solution, as the continuous monitoring process is “multilayered and uses a lot of machine learning” (note 5), reinforcing the technical as the focus of solutions.

### *Reproduction of Society*

As a corporation especially strategically positioned in the access and organization of information, Google is interested in maintaining its image of neutrality. With the declared mission to “organize the world’s information and make it universally accessible and useful”<sup>13</sup>, the corporation encompasses various information services and products under the umbrella of Alphabet, but possesses a search engine as one of its principal, indispensable assets, serving also as the base for other services, such as cloud computing.

Two recent cases of press coverage in Brazil about offensive biases in the presentation of results received comments from the corporation. In July of 2019, a comparison of the results of searches for “tranças feias” (“ugly braids”) and “tranças bonitas” (“beautiful braids”) went viral in social media, which was covered in news outlets such as *O Globo*, via the column *O Blog do Ancelmo*. In this space, Google came forward and announced that the results merely reproduced existing “stereotypes”: “As our systems find and organize information available on the web, eventually, a search may reflect existing stereotypes on the internet and in the real world according to the manner in which some authors create and label their content” (Note 8).

---

<sup>13</sup> <https://about.google/>.

In November 2019 another case came about: users of the search engine identified that terms searched such as “mulher negra dando aula” (“black woman giving class”) returned basically pornographic results. Consulted about the report published on the site *Universa* from *UOL*, the company recognized that “the collection of results for the term in question is not compatible with this principle” (Note 9) and alleged that they would “seek a solution to refine the results not only for these terms, but also for other searches that may present similar challenges” (Note 9) and suggested that users also take an extra step: activate the SafeSearch option, originally created to hide pornographic content from minors.

As in prior cases, the framing of the question as a surprise to the corporations is directly at odds with the academic and specialist bibliography which has dealt with, in the past 10 years (NOBLE, 2011; EDELMAN, 2011; SWEENEY, 2013), the potential algorithmic harm of search engines. In *Algorithms of Oppression*, Safiya Noble reveals the means by which search engines perform representations “decontextualized in one specific type of information-retrieval process, particularly for groups whose images, identities, and social histories are framed through forms of systemic domination” (NOBLE, 2018, pos.2467).

### *Distinct Reactions*

The posture of Google in the two cases cited above contrast directly with the case of the mobilization of the French organization SEOLesbienne. The project, connected to the organization combatting sexual and gender-motivated violence *Nous Totes*, pressured the search engine to change its algorithm to reverse the hyper-sexualization of search results for terms such as “lesbian” and “lesbianism” on the platform, to prioritize informative, newsworthy and cultural results about lesbian identities, in opposition to the misogynist content often found on pornographic sites.

In reporting from *The Next Web* with information from the French portal *Numerama*, Google’s Vice President of Search, Pandu Nayak, recognizes that there are problems like these in various languages and searches and explains the decision to take measures “in cases where, when there is a reason for the word to be interpreted in a non-pornographic way, that interpretation is put forward” (Note 7).

The contrast can be seen in the attribution of judgement of the results. While Nayak clearly expresses “I find that these [search] results are terrible, there is no doubt about it” (Note 7), the Brazilian declarations apologize to “those who felt impacted or offended” (Note 8), shifting the perception of offense to the affected public while minimizing the power relationships over marginalized groups by saying that “people of all races, genders

and groups can be affected" (Note 8) and resurrect the ideology of race-blindness by emphasizing that they will supposedly do the same "also for other searches that may present similar challenges" (Note 9).

### *Denial of Scope of Responsibility*

The denial of responsibility beyond the explicit objective of algorithmic applications and systems is accomplished via the appeal to the complexity of the information products in question. One of the most emblematic cases of this type of strategy is the repeated controversies of the selfie posting and editing app *FaceApp*. In April 2017, during one of the first waves of the app's popularity, users noticed that one of the "beautifying" filters consistently whitened users from groups with darker skin. Questioned by *Techcrunch*, the application's CEO, Yaroslav Goncharov claimed that the problem was "an unfortunate side-effect of the underlying neural network caused by the training set bias, not intended behavior" (Note 3) and that they would work on a "complete adjustment" to deliver soon. However, in August of the same year, the application launched a problematic racial-simulation feature and in 2019 it was seen that the new aging-simulation filter also whitened users in terms of skin color and facial features.

In spite of trying to avoid responsibility by placing the bias within the training database, Goncharov admitted that the company used its own created data. The contradiction is pointed out in the reporting of the technology portal (Note 3) and is of parallel relevance to the case of the scores (Notes 1 and 2) in reaction to the *Gender Shades* project cited earlier.

In an extensive response, IBM took the opportunity of the case of the algorithmic audit to conduct an experiment reproducing part of the methodology of the initial effort, claiming lower error rates than the competition (Note 2), information contradicted by the later study of researchers (RAJI & BUOLAMWINI, 2019). The corporation claimed, without further detail, that it "now uses different training data and different recognition capabilities than the service evaluated in this study" (Note 2) and that it seeks to support "projects to address dataset biases" (Note 2). However, it is worth noting the absence of any mention of one of the principal discoveries of the original work, the fact that the two open visual databases most used by the segment are extremely biased.

### **Final considerations**

As seen above, we can recognize that corporations seek to simplify the debate on algorithmic harm in the public sphere through various expedients. The fight against

algorithmic harm in contemporary societies characterized by oligopolies of digital platforms necessarily comes to question the notion that algorithms are inscrutable black boxes, as this guarantees “a special place in the world of unknowns that perhaps is not fully deserved” (BUCHER, 2016b, p. 85-86).

On the contrary, the scope of responsibility in implementing algorithmic systems in commercial or public systems involves dealing with the controversies regarding their limits and the means by which the avoidance of responsibility and agency (RUBEL *et al*, 2019) is put into practice through public statements in the press or corporate communication channels.

We have identified, in cases of algorithmic harm with wide impact, three methods by which companies or corporations react to criticism and algorithmic audits: evoking a continuous process of optimization as a characteristic of contemporary digital technology; claiming that the systems only reproduce the inequalities and problematic stereotypes already present in society, therefore restorative actions are always optional or even unfair; and framing of the scope of responsibility in supposedly strictly technical minutiae, assigning the prior training of systems and impacts on society to externalities. Therefore, the comparison between statements about cases in the centers of power in relation to the Global South show distinct reactions depending on nationality, race and class, undermining the common arguments of neutrality in technology. In the current scenario of media confusion and crises of authority of the traditional journalistic outlets, the critical engagement of the public in the algorithmic controversies that moderate the access or restriction to egalitarian use, and the security of the means of communication are proven to be essential.

## References

- BROCK, Andre. Análise Crítica Tecnocultural do Discurso. In: SILVA, T. Comunidades, Algoritmos e Ativismos Digitais: olhares afrodiaspóricos. São Paulo, LiteraRUA, 2020.
- BUCHER, Taina. The algorithmic imaginary: exploring the ordinary effects of Facebook algorithms. *Information, Communication & Society*, v. 20, n. 1, p. 30-44, 2016a.
- BUCHER, Taina. Neither black nor box: ways of knowing algorithms. In: KUBITSCHKO, S. & KAUN, A. (orgs.) *Innovative methods in media and communication research*. Palgrave Macmillan, Cham, 2016b. p. 81-98.

- BUOLAMWINI, Joy; GEBRU, Timnit. Gender shades: Intersectional accuracy disparities in commercial gender classification. In: Proceedings of Conference on fairness, accountability and transparency, 2018. pp. 77-91.
- DIAKOPOULOS, Nicholas. Accountability in algorithmic decision making. *Communications of the ACM*, v. 59, n. 2, p. 56-62, 2016.
- EPSTEIN, Ziv *et al.* Closing the AI Knowledge Gap. arXiv preprint arXiv:1803.07233, 2018.
- ESLAMI, Motahhare *et al.* I always assumed that I wasn't really that close to [her]: Reasoning about Invisible Algorithms in News Feeds. In: Proceedings of the 33rd annual ACM conference on human factors in computing systems. ACM, 2015. p. 153-162
- FRANCE PRESSE. (2019). San Francisco proíbe a polícia de usar reconhecimento facial Oito dos nove conselheiros municipais são contrários à tecnologia. G1, 16/05/2019, online. Disponível em: <https://g1.globo.com/pop-arte/noticia/2019/05/16/san-francisco-proibe-a-policia-de-usar-reconhecimento-facial.ghtml> Accessed 22/04/2020.
- GILLESPIE, Tarleton. A relevância dos algoritmos. *Paragraph*, 6(1), 2018, pp. 95-121.
- GUNNING, D. Broad Agency Announcement Explainable Artificial Intelligence (XAI). Technical report, 2016.
- GUNNING, David. Explainable artificial intelligence (xai) Program. *AI Magazine*, v. 40, n. 2, 2019. pp.44-58.
- LATOUR, Bruno. Why has critique run out of steam? From matters of fact to matters of concern. *Critical inquiry*, v. 30, n. 2, p. 225-248, 2004.
- LATOUR, Bruno. *Reassembling the Social: An Introduction to Actor-Network-Theory*. New York: Oxford University Press, 2005.
- NOBLE, Safiya Umoja. *Searching for Black Girls: Ranking Race and Gender in Commercial Search Engines*. Doctoral Thesis defended at Urbana-Champaign: University of Illinois at Urbana-Champaign, 2011.
- NOBLE, Safiya Umoja. *Algorithms of oppression: How search engines reinforce racism*. New York: NYU Press, 2018.
- PASQUALE, Frank. *The black box society*. Harvard University Press, 2015.
- PASQUALE, Frank. Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society. *Ohio St. LJ*, v. 78, p. 1243, 2017.

- RAJI, Inioluwa Deborah; BUOLAMWINI, Joy. Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial ai products. In: AAAI/ACM Conf. on AI Ethics and Society, 2019.
- RIBEIRO, Manoel Horta *et al.* Auditing radicalization pathways on youtube. arXiv preprint arXiv:1908.08313, 2019.
- RIEDER, Bernhard; MATAMOROS-FERNÁNDEZ, Ariadna; COROMINA, Òscar. From ranking algorithms to 'ranking cultures' Investigating the modulation of visibility in YouTube search results. *Convergence*, v. 24, n. 1, p. 50-68, 2018.
- ROMANI, Cristóbal C.; KUKLINSKI, Hugo P. *Planeta Web 2.0: Inteligencia colectiva o medios fast food*. Barcelona: Grup de Recerca d'Interaccions Digitals, Universitat de Vic. Flacso, 2007.
- RUBEL, Alan; PHAM, Adam; CASTRO, Clinton. Agency Laundering and Algorithmic Decision Systems. In: *International Conference on Information*. Springer, Cham, 2019. p. 590-598.
- SANDVIG, Christian *et al.* Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry*, v. 22, 2014.
- SEAVER, N. *Knowing Algorithms*. In: VERTESI, J.; RIBES, D. (orgs.) *digitalSTS: A Field Guide for Science & Technology Studies*. Princeton University Press, 2019. pp.412-422.
- SILVEIRA, S. A. *Democracia e os códigos invisíveis: como os algoritmos estão modulando comportamentos e escolhas políticas*. São Paulo: Edições SESC-SP, 2019.
- SRNICEK, Nick. *Platform capitalism*. John Wiley & Sons, 2017.
- SWEENEY, Latanya. Discrimination in online ad delivery. arXiv preprint arXiv:1301.6822, 2013.
- VAN DIJCK, José. Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 2014. pp. 197-208.

**RESUMO:**

Impactos discriminatórios e danos de sistemas algorítmicos têm gerado discussões sobre o escopo da responsabilidade de empresas de tecnologia da comunicação e inteligência artificial. O artigo apresenta controvérsias públicas engatilhadas por 8 casos públicos de danos e discriminação algorítmica que geraram respostas públicas de empresas de tecnologia, abordando o esforço realizado pelas empresas de tecnologia em enquadrar o debate sobre responsabilidades no fluxo de planejamento, treinamento e implementação dos sistemas. Em seguida, discute como a opacidade dos sistemas é defendida pelas empresas comerciais que os desenvolvem, alegando prerrogativas como “segredo de negócio” e inescrutabilidade algorítmica.

**PALAVRAS-CHAVE:** Algoritmos; Auditoria algorítmica; Explicabilidade; Jornalismo de Tecnologia; Plataformas.

**RESUMEN:**

Los impactos y daños discriminatorios por sistemas algorítmicos han abierto discusiones sobre el alcance de responsabilidad de las empresas de tecnología de la comunicación e inteligencia artificial. El artículo presenta controversias públicas desencadenadas por ocho casos públicos de daño y discriminación algorítmica que generaron respuestas públicas por parte de las empresas, abordando los esfuerzos realizados por ellas en enmarcar el debate sobre la responsabilidad en el transcurso de la planeamiento, alimentación con datos e implementación de sistemas. A continuación, se analiza cómo la opacidad de los sistemas es defendida por las empresas comerciales que los desarrollan, alegando prerrogativas como los “secretos de la industria” y la inescrutabilidad algorítmica.

**PALABRAS-CLAVES:** Algoritmos; Auditoría algorítmica; Explicabilidad; Periodismo Tecnológico; Plataformas.