Scientific Research Publishing

Journal of Service Science and Management, 2022, 15, 551-575
https://www.scirp.org/journal/jssm
ISSN Online: 1940-9907
ISSN Print: 1940-9893

Artificial Intelligence, Ethics and Public Policy—The Use of Facial Recognition Systems in Public Transport in the Largest Brazilian Cities

Rodrigo Brandão, Glauco Arbix

Department of Sociology, Faculty of Philosophy, Letters and Human Sciences, University of São Paulo, São Paulo, Brazil
Email: brandao-cs@usp.br, garbix@usp.br

How to cite this paper: Brandão, R., & Arbix, G. (2022). Artificial Intelligence, Ethics and Public Policy—The Use of Facial Recognition Systems in Public Transport in the Largest Brazilian Cities. *Journal of Service Science and Management, 15,* 551-575.
https://doi.org/10.4236/jssm.2022.155032

Received: September 15, 2022
Accepted: October 15, 2022
Published: October 18, 2022

Abstract

The literature on using Artificial Intelligence (AI) systems for elaborating and implementing public policies is under development. Therefore, little is known about how the public sector can and should use AI responsibly and in what situations it has been able to do so. To partially fill these gaps, we investigated the use of facial recognition (FR) systems to combat fraud in discounts and gratuities guaranteed by law to specific groups, such as students and the elderly, in the 30 largest Brazilian municipalities. Based on primary data gathered through the Access to Information Law, we found that 14 use the technology in question for the investigated purpose, and we prepared the Facial Recognition Responsible Use Index (FRRU-I) and the Facial Recognition Responsible Use Scale (FRRU-S). None of the cities studied reached the "Very high" level in the FRRU-S; five of them achieved "High" level scores; three were at the "Intermediary" level; two obtained a "Low" score, and the remaining four showed a "Very low" score. These results suggest that the largest cities in the country do not have the necessary administrative structures for the responsible use of FR systems or do not know how to mobilise the structures they have to promote such use. Our findings also indicate that the public sector's unfamiliarity with FR systems may extend to other AI applications.

Keywords

Artificial Intelligence, Facial Recognition, Ethics, Public Policy, Public Transport

1. Introduction

Artificial Intelligence (AI) is "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions in-

fluencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy" (Organisation for Economic Cooperation and Development [OECD], 2019). These qualities have encouraged many governments to adopt different applications of AI. In an extensive inventory of uses of this technology in the public sector, Berryhill, Heang, Clogher and McBride (2019) mention, among other examples, that Portugal has resorted to chatbots equipped with Natural Language Processing technologies to bring citizens and the public machine together during the provision of public services; that, concerned with preserving biodiversity, Australia has been using Machine Learning (ML) techniques to make its satellites learn to recognise and classify different land uses; and that Denmark plans to use AI systems to help public officials make decisions about granting social benefits. Kirwan and Fu (2020) also found different uses of AI, especially in smart city projects—both in traditional cities such as Beijing, London and New York and in newly built or under construction cities such as Songdo, Masdar and NEOM.

In addition to the intention to use AI to improve State actions, the examples above have a second common characteristic. In all of them, the AI uses are recent or experimental. This originality has given rise to positive and negative projections on the use of AI by the State. The first group comprises works such as Engin and Treleaven (2019)—which list how a series of technologies based on algorithms can contribute to the automation of public services, thus reducing the arbitrariness of public decisions, and Eggers, Schatsky and Viechnicki (2017)—who estimate, for the US case, that AI technologies can help automate administrative activities, saving public budgets up to US$ 41.1 billion annually. In the second group, studies discuss the possibility of algorithmic technologies making the State hyper-vigilant (Zuboff, 2019) or, thanks to opaque systems, entirely refractory for any accountability (Busuioc, 2020).

Amid social fears and hopes projected onto technology, the empirical literature on the use of AI by public agencies has found it challenging to move forward, often being summarised in discussions of pilot projects and speculations about the risks and benefits associated with it (Sun & Medaglia, 2019; Sharma, Yadav, & Chopra, 2020; Yigitcanlar, Corchado, Mehmood, Li, Mossberger, & Desouza, 2021). Because of this, we still lack concrete references to guide the responsible use of AI in public administration, which led us to ask: what factors are necessary for such use to be possible?

Guided by the above questioning, we structured a two-pronged research agenda. The first front aims to define the expression "responsible use of AI" in the light of theoretical debates about this technology and, based on this definition, to evaluate empirical cases of AI uses in the public sector. Informed by the literature on State capacities at the local level (Coelho, Guth, & Loureiro, 2020), the second part seeks to test the hypothesis that the greater the bureaucratic capabilities of the public machine, the more responsible the use of AI. This article presents the results of the first of these two work fronts and deepens theoretical and empirical discussions made in Brandão (2022), Brandão, Oliveira, Peres,

Júnior, Papp, Veiga et al. (2022), and Brandão, Oliveira and Júnior (2022).

As we will see, we investigated the use of facial recognition (FR) systems to combat fraud in discounts and gratuities guaranteed by law to specific groups in the largest Brazilian cities. More precisely, our investigation focused on a set of 30 towns that meet at least one of two criteria: they are state capitals and/or have at least one million inhabitants. We found that these municipalities face difficulties in using FR systems responsibly. This result suggests that the largest Brazilian cities do not have the necessary administrative structures to promote the responsible use of FR systems or do not know how to mobilise the structures they have to encourage such use.

In addition to this section, the article is composed of three others. In Section 2, we review the literature on the responsible use of AI by public departments, and based on this review, we justify the focuses of our investigation. Section 3 presents the methodology used to create the Facial Recognition Responsible Use Index (FRRU-I) and the Facial Recognition Responsible Use Scale (FRRU-S) and discusses our results, highlighting that future studies should not only deepen the investigation of State capacities and responsible uses of AI but also investigate whether our findings are valid for other AI technologies. Finally, in Section 4, we conclude the article

2. Literature Review

This section consists of three parts. In Subsection 1, we present the methodology used to select the papers that were read and evaluated. In Subsection 2, we position the expression "responsible use of AI" in the universe of selected texts. In Subsection 3, we emphasise the empirical works analysed to demonstrate concrete challenges to the responsible use of AI in the public sector.

2.1. From the Research Field to Theoretical Discussions on Ethics in AI

The expression "responsible use of AI" is used by different actors. These include the Canadian government and consultancy firm Oxford Insights—just to mention two examples. On its official website, the former has an "entry" dedicated exclusively to this topic, in which it gathers, among other contents, principles for the effective and ethical use of AI applications by federal agencies, lists of suppliers qualified to offer AI products to public entities, and standards to be observed by federal bureaucracy bodies when using systems capable of making automated decisions (Canada, 2021).

On the other hand, Oxford Insights periodically maps the level of readiness of different governments for adopting the technology in question. In the 2020 edition of the survey, Shearer, Stirling and Pasquarelli (2020) prepared the "Responsible AI" sub-index for a set of 34 countries, which possess four dimensions: inclusivity, accountability, transparency and privacy. As shown in **Table 1**, they work with nine indicators.

Table 1. Shearer et al.'s (2020) Responsible AI Index.

Dimension	Indicator	Source	Hypothesis
Inclusivity	Income Inequality	World Bank	AI and automation are most likely to exacerbate inequality in already unequal countries
	Automation Readiness Index	(Not Identified)	The most automation-ready countries will be best placed to ensure inclusive AI-driven growth
Accountability	Voice and Accountability	World Bank	In societies with higher levels of voice and accountability, citizens will be able to challenge irresponsible uses of AI by governments
	Freedom on the net	Freedom House	Countries that grant citizens free and unfettered access to the Internet will allow those citizens to access the relevant information to hold governments accountable
Transparency	Corruption Perception Index	Transparency International	Governments that are perceived as corrupt are less likely to implement AI in transparent and open ways
	Corporate Political Engagement Index	Transparency International	Powerful private companies can lobby governments for fewer regulations on AI and less scrutiny of their own irresponsible practices
Privacy	Rule of Law	World Bank	Countries with robust respect for rights and the rule of law will be more likely to implement AI in a way that respects privacy rights
	Surveillance Industry Index	(Not Identified)	Countries with extensive existing surveillance industries are at risk of implementing AI in a way that undermines privacy
	AI Surveillance Index	Carnegie Endowment	Countries that use AI more extensively for surveillance are at higher risk of abusing those capabilities

Source: Shearer et al. (2020)—Adapted.

Without offering an explicit definition for the expression "responsible use of AI", Shearer et al. (2020) make it possible to understand that it refers to AI uses that do not deepen deficits in inclusivity, accountability, transparency and privacy. This understanding is echoed in the concept "Artificial Intelligence for social good". Proposed by Luciano Floridi, one of the most prolific AI ethicists, it carries the idea that the design, development and employment of AI systems should "(i) prevent, mitigate or resolve problems adversely affecting human life and/or the wellbeing of the natural world, and/or (ii) enable socially preferable and/or environmentally sustainable development" (Floridi, Cowls, King, & Taddeo, 2020: pp. 1773-1774).

As can be seen, practical uses of the expression "responsible use of AI" make it possible to associate it with the term "ethics". Concerned in investigating whether academic documents support this association, we researched three combinations of terms in the Scopus database: (i) artificial intelligence AND ethics; (ii) artificial intelligence AND public policy OR public sector OR public administration OR government; (iii) artificial intelligence AND ethics AND public policy OR public sector OR public administration OR government. In the three cases, we used five filters: (i) Open Access: "All Open Access"; (ii) Year: "2022"; "2021"; "2020"; "2019"; "2018"; (iii) Subject area: "Computer Science"; "Social Sciences AND Arts and Humanities"; (v) Document type: "Article";

"Conference Paper".

As we combined three combinations of terms and two subject areas, we generated three pairs of lists, which were ordered from the most cited document to the least cited document, considering the 20 most cited documents. The selection of documents to be read was based on two criteria: (i) in each of the three pairs, we read all the documents common to both lists; (ii) excluding these texts, we read the two most cited from the "Computer Science" lists and the two most cited from the "Social Sciences AND Arts and Humanities" lists. (In the second case, we made some exceptions, such as reading more than two documents or replacing reading a much-cited document for another less-cited document).

Finally, we carried out similar searches in the journals Public Administration Review (PAR) and Regulation & Governance and the Brazilian journals Revista da Administração Pública and Revista do Serviço Público. In these last two, we did not find any articles on AI. In Regulation & Governance, we came across articles related to this technology, but—after reading their abstracts—we judged them irrelevant to our investigation. Finally, in PAR, we identified—by reading the abstracts—four articles published between 2018 and 2022 that were adherent to our research.

The searches above culminated in the selection of 42 documents. They are listed in Methodological Annex 1. (It can be found in the digital annex that integrates this paper. Its link is available below in "Methodological Annexes".) We added others to this set of documents on AI and public policy. The analysis of this bibliographic universe led us to conclude that the use of AI by public offices should not be evaluated only by its consequences—as Floridi et al. (2020) and Shearer et al. (2020) may lead one to think—but also by the existence of means that make possible the responsible use of AI. In other words, a government that resorts to AI applications without preparation can be considered irresponsible from the moment it deploys them.

2.2. Theoretical Discussions on Ethics in AI: A Brief Overview

Among the documents reviewed, some are concerned with mapping the field of studies on the relationship between ethics and AI. This is the case, for instance, of Yu, Shen, Miao, Leung, Lesser and Yang (2018) and Winfield, Michael, Pitt and Evers (2019). However, most of them are dedicated to reviewing public and private treaties on ethics in AI—as Morley, Floridi, Kinsey and Elhalal (2020) and Hagendorff (2020)—or to discussing particular ethical principles, such as "privacy", "explainability" or "accountability". The works of the first of these two groups make it clear not only that the term "ethics" is polysemic but also that the treaties on the subject lack enforcement mechanisms, giving rise to a risk that Hasselbach (2019) (as cited in Aitken, Toreini, Carmichael, Coopamootoo, Elliott, & Moorsel, 2020) calls "ethics washing": the practice of public and private institutions publishing AI ethics treaties just to appear concerned about the topic. A similar danger is that codes of ethics have no impact on re-

searchers and developers of AI applications. In this sense, Vakkuri, Kemell and Abrahamsson (2020) point out that:

> McNamara et al. (2018) studied the impact the ACM [Association for Computing Machinery] Code of Ethics had had on practice in the area [of computing machinery], finding little to none. This also seems to be the case in AI ethics: in a recent paper (2020), we studied the current state of practice in AI ethics and found that the principles present in the literature are not actively tackled out on the field. (p. 196)

Among the works that discuss ethical principles, the term "ethics" ends up being defined by examples or being equated with a specific principle, without making it clear the definition of "ethics" used by the authors and for what reasons the observance of the principle under discussion is instrumental to the existence of ethics in AI. This approach can be found in Stahl and Wright (2018), Martin (2018), Felzmann, Villaronga, Lutz and Tamò-Larrieux (2019), Milano, Taddeo and Floridi (2020) and Robinson (2020), among others.

Floridi, Cowls, Beltrametti, Chatila, Chazerand, Dignum et al. (2018) are the only authors to offer a complete theoretical framework. According to them, the ethical uses of AI are those that promote human dignity, and the means to achieve this end is respecting five principles: "beneficence" ("do only good"), "non-maleficence" ("do no harm"), "autonomy", "justice" and "explicability". For the authors, human dignity consists of "[…] *who we can become* (autonomous self-realisation), *what we can do* (human agency), *what we can achieve* (individual and societal capabilities), and *how we can interact with each other and the world* (societal cohesion)" (p. 690). The good uses of AI, the authors go on, are those that contribute to the flourishing of these four elements. In contrast, harmful uses damage them – either because the technology was misused or used in situations in which it was not needed. The authors also envision opportunity risks linked to AI, which occur in cases where AI could promote human dignity. Still, the technology is discharged for fear or ignorance of the actors involved.

Going further, Floridi et al. (2018) observe that, except for the principle of explicability, the above principles are the most frequent in AI ethics treaties, such as the Asilomar AI Principles and the Montreal Declaration for a Responsible Development of Artificial Intelligence—both published in 2017. The authors also claim that these principles apply both to those who develop AI systems and to the systems themselves. In other words, we can understand from Floridi et al. (2018) that the imperative "do no harm", for example, must be used to assess both the behaviour of the developers of the technology and the functioning of the technology in non-laboratory contexts.

The first principle ("beneficence") relates to promoting well-being, sharing social benefits, and advancing the common good. "Non-maleficence" usually refers to privacy protection in the treaties reviewed by the authors. "Autonomy" usually rules that human beings must be capable of choosing in which situations

they want to delegate decision powers to intelligent systems and in which cases they want to revoke this decision. The principle of "justice", in turn, refers to the use of AI to promote prosperity and preserve solidarity. Finally, the authors point out that "explainability" is a principle that appears diffusely among the documents they analysed. Floridi et al. (2018) understand it as a synthesis of "intelligibility" ("how do AI systems work?") and accountability ("who is responsible for the way AI systems work?").

In our bibliographic universe, the most incisive criticism of the theoretical framework above is presented by Robbins (2019), which refers to the principle of "explainability". For him, the way Floridi et al. (2018) address this principle suggests that AI system operators should be able to understand the technical intricacies that convert inputs into outputs and, based on this understanding, assess whether the latter should be accepted or rejected. The author opposes this understanding arguing that "[i]t is the context and the potential harm resulting from decisions that drive the moral need for explicability—not the process by which decisions are reached" (p. 495). For this reason, according to Robbins (2019), no decision that requires explanation should be made by an AI algorithm since what is at stake are the premises that base the decision. Furthermore, if a human being needs to be able to know and explain all the assumptions that underlie an AI system and how they relate to each other, such systems are not necessary.

Two hypothetical examples discussed by the author are especially effective in illustrating his point. For him, we may want to understand how an ML algorithm that proved efficient in predicting weather conditions works. This desire, argues Robbins (2019), is based on our intention to know more deeply about climatic conditions and not on the perception that it would be wrong to use pieces of information that we do not know how they were generated. On the other hand, if we understand that a predictive policing or credit rating system is based on racial information, we disapprove (or should disapprove) of using these systems—regardless of their accuracy rates. As can be seen, for Robbins (2019), algorithmic opacity is not a problem in itself—what would demand attention is the context in which it occurs.

Robbins' (2019) critique of Floridi et al. (2018) seems unfounded to us since the theoretical framework of the latter allows us to state that the uses of opaque AI systems in morally sensitive situations characterise a situation in which the technology must not be used since it can damage human dignity. However, Robbins' (2019) position on the limitations of human supervision is pertinent, especially when considering recent research findings by Green (2021).

The author reviewed 40 documents with guidelines on human oversight of algorithmic systems—such as AI systems—used by governments. This set includes, for instance, the European Union General Data Protection Regulation (GDPR) and the Brazilian General Data Protection Law (Lei Geral de Proteção de Dados Pessoais—LGPD, in the Portuguese acronym). Green (2021) concluded that many of the guidelines investigated fail to make human supervision

an instrument capable of preventing or correcting errors in algorithmic systems, but that, in what one could call a clear example of ethics washing, the mere existence of these guidelines legitimises the use of these systems, even when they are flawed or present low levels of accountability. Based on these findings, the author proposes two courses of action:

> First, rather than crafting blanket rules that enable governments to use algorithms as long as a human provides oversight, policymakers must place far greater scrutiny on whether an algorithm is even appropriate to use in a given context. Second, rather than assuming that people can oversee algorithms effectively, policymakers must empirically evaluate whether the proposed forms of human oversight can actually function as desired. A central principle guiding this proposal is to increase the burden placed on agencies to affirmatively justify both their decisions to adopt algorithms and their proposed mechanisms for governing those algorithms. (p. 23)

Green's (2021) considerations touch on two challenges. The first one is to understand whether human intervention can prevent or correct AI errors and thus ensure that the use of AI systems improves State actions. Secondly, Green's (2021) findings indicate it remains to be investigated the capacity of public offices to assess whether or not an AI application is necessary and, if so, their ability to adopt governance structures regarding the use of this application.

2.3. Back to the Research Field: A Landing in Empirical Studies

2.3.1. Types of AI Uses in the Public Sector

As we saw in the introductory section of this article, there are a growing number of reports of AI use by public offices. However, a lack of systematisation has prevented us from knowing in which areas the technology has been most used, the main benefits obtained through it, and, in the case of errors, which are the most frequent. The works by Engstrom, Ho, Sharkey and Cuéllar (2020) and Coelho and Burg (2020) do not fill all these gaps, but they present an overview of the use of AI in the public sector. The first one refers to the US, and the second to Brazil.

Of the various research findings by Engstrom et al. (2020), we highlight that, of the 142 US federal agencies surveyed, 45% use or have already tested AI tools, especially ML ones, both in administrative activities and in actions that directly impact citizens. In the first group are tools to contain cyberattacks, optimise the management of internal resources and analyse large volumes of data during the policymaking process; in the second, some software helps public managers to make decisions about social benefit granting and about people to be monitored by public agencies.

Coelho and Burg (2020), in turn, identified 44 AI tools used in the Brazilian federal public administration. Out of these, 20 are used within public agencies and are responsible for generating results that support human decisions. The Bem-te-vi tool, for example, is used by the Superior Labour Court to assist its

servers in classifying cases. Another eight tools involve decision-making but are used in the State's relationship with citizens. The chatbots present in some public hospitals are an example of this type of use, as they are responsible for the triage of patients. Finally, the remaining 16 tools are used within public offices but do not involve decision-making—either by the technology itself or by human beings who use them in work processes. They are, for example, software that helps public managers to analyse large volumes of data, such as contributions received in public hearings.

As can be seen, the typology of Coelho and Burg (2020) comprises four uses of AI: (i) internal, with decision-making; (ii) internal, without decision-making; (iii) external, with decision-making; and (iv) external, without decision-making. Throughout this section, we will discuss the challenges posed to the public sector in uses that involve decision-making or, more precisely, the generation of outputs that support public managers in making decisions.

Finally, it is worth noting—still in the Brazilian case—that the use of FR systems by public departments of states and municipalities is becoming more frequent. According to the think-and-do tank Instituto Igarapé (2019), the number of cases in which this technology has been used to operationalise public policies jumped from one in 2011 to 47 in 2019. In 21 of these cases, the technology has been used to combat fraud in public transport. Pieces of information collected by Brandão and Oliveira (2021) reinforce these findings. Focusing only on the 17 Brazilian municipalities with at least one million inhabitants, the authors searched for occurrences of the term "facial recognition" in editions of official electronic journals published between January 2010 and December 2020. They identified uses already completed, in progress or under discussion in ten locations. In nine of them, the use of FR systems was associated with the fight against fraud in discounts and gratuities in public transport; in four, with goals in public security; in two, with the operationalisation of health and education policies; and, in one, with public actions in social assistance.

2.3.2. (In)voluntary State Actions and Algorithmic Accountability

Among the uses of AI applications in the public sector, the case of the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) system is the most notorious. The system in question calculates the risk that a detainee commits a new crime if a judge grants him parole. After analysing the results of this system in Broward County, Florida, Angwin, Larson, Mattu and Kirchner (2016) found that it presents marked racial biases. The risk score assigned to white people is generally lower than that calculated for black people. In addition, when making mistakes, COMPAS makes mistakes favourable to white people: "[…] blacks are almost twice as likely as whites to be labelled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labelled lower risk but go on to commit other crimes".

Still, in public safety, FR software has also made dangerous mistakes. There

are records in the US press that, as of January 2021, there were at least three proven cases in the country of people who were wrongly arrested because of mistakes made by this technology: all of them were black (Johnson, 2022). In Brazil, there is also initial evidence of biased results associated with using FR systems in public safety. Based on newspaper articles, Nunes (2019) identified arrests made using this technology in four Brazilian states: Bahia, Rio de Janeiro, Santa Catarina and Paraíba. The author identified 151 cases of this nature, and in 42 of them, he was also able to observe the race of the detainees: 90.5% of them were black.

Eubanks (2018), in turn, identified errors in granting social benefits. The author states that based on the outputs of algorithmic systems, the state of Indiana denied more than one million requests for financial, food and health assistance between 2006 and 2008. The author found that numerous denials contained errors. Calo and Citron (2020) also mention errors in granting social benefits in the US. The authors point out that, in 2016, the Arkansas Department of Human Services turned to intelligent systems to decide which home health care requests to accept. They report controversial and offensive results, such as the case of a person who, having had a foot amputated, had his request denied because he could not have any foot problem. The mistakes made by the technology led the Department of Human Services to deny requests for home medical care to those who would be entitled to receive it, which, according to Calo and Citron (2020), motivated numerous lawsuits against it.

Among the examples above, the case of Arkansas makes it clear that driven by the unintelligibility of technology, the State can involuntarily damage citizens' rights. How to avoid situations of this kind, in which the results of technology harm citizens and the Public Power? By the theoretical framework of Floridi et al. (2018), the answer lies in the principle of "explainability"—the synthesis of "intelligibility" ("how do AI systems work?") and accountability ("who is responsible for how AI systems work?"). However, the weight of these two elements is not the same for citizens and public managers.

Following Robbins (2019), we can say that, among citizens, some may be interested in "intelligibility", while others may not; but that all of them, when suffering some damage caused by the malfunction of the technology, will understand that the State is, in whole or in part, responsible for this result. The responsibility under discussion, it should be noted, is of an ethical rather than legal nature. After all, "[c]ompliance with the law is merely necessary (it is the least that is required), but significantly insufficient (it is not the most that can and should be done)" (Floridi et al., 2018: p. 694).

As can be seen, when employing AI systems, especially in the implementation phase of public policies, the public administration must also be prepared to adopt mechanisms that guarantee citizens that, if they suffer any undue damage, it will be repaired, and the developers and/or public managers will be responsible for the adverse outcome. Without these mechanisms, the principle of "non-maleficence" is in jeopardy. This accountability strategy is called by Ada

Lovelace Institute, AI Now Institute and Open Government Partnership (2021) "algorithmic accountability". When researching around 20 national and local governments in the Global North, the researchers found different mechanisms concerned with promoting algorithmic accountability, such as publishing principles and guidelines on the use of algorithmic systems, carrying out algorithmic impact assessments, carrying out regulatory audits and inspections, creating external/independent oversight bodies, and establishing conditions in technology procurement.

Algorithmic accountability mechanisms are necessary for a second reason. If it is true that, without control over technology, the State can be led to harm citizens unintentionally, it is equally valid that it can do so intentionally. In this case, the technology errors are not at stake but the behaviour of public institutions when using it. Eubanks (2018) investigated this issue in depth by carrying out three case studies in the US on using algorithmic technologies to operationalise social policies.

In addition to the Indiana case, already mentioned in this subsection, Eubanks (2018) analysed the use of space allocation systems devoted to homeless people in Los Angeles and the use of the Allegheny Family Screening Tool in Pennsylvania. Based on the information gathered, the author concluded that the US government had used cutting-edge technologies to monitor the daily lives of the poorest, always attempting to punish them. The author also argues that this action is usually based on the discourse that technology would help to improve social policies and that it is only possible because the data of the poorest are more readily available to the State since, in comparison with people from middle and upper classes, they rely more often on public services. In the analytical scheme of Eubanks (2018), algorithmic systems would be, therefore, a valuable instrument for penalising the poorest for their "dependency" on the State, in a clear signal that, for many stakeholders, social policies should not be seen as a legal right, but as a benevolent act of the State. In such a reality, algorithmic accountability mechanisms are essential for citizens to know which of their personal data have been used by the State and for what purposes.

Despite being central to protecting citizens against voluntary or involuntary abuses by the State, the effectiveness of algorithmic accountability mechanisms has not yet been tested. One of the few exceptions is the study by Grimmelikhuijsen (2022), who conducted two survey-based experiments to test whether algorithmic transparency increases citizens' trust in public decisions made by or with the aid of algorithms. The author found that access to data and algorithmic code matters, but the decisive element in increasing trust is access to understandable explanations about the results.

2.3.3. Technological Complexity and Human Limitations

In addition to being attentive to the promotion of mechanisms that make them accountable to citizens, public managers must be able to understand—and, consequently, explain to other people—how the AI systems they have developed or

acquired in the market work. This understanding is critical for two reasons. Firstly, without it, the State can be led to damage itself and citizens' rights—as discussed above. In the second place, public managers can see their autonomy diminished by AI systems. Kuziemski and Misuraca (2020) found an eloquent example in Poland.

In recent years, the country's Ministry of Labor and Social Policies has used AI tools to classify unemployed people and to dedicate a greater volume of resources and support in professional relocation to those with a lower chance of getting a new job. The technology's outputs should support the decisions of those responsible for operating it. However, this was not the observed result. The authors note that:

> […] less than 1 in 100 decisions made by the algorithm have been questioned by the responsible clerks. Unless for the outstanding precision of the algorithm (which is to be considered unlikely), the reasons for not challenging automated decisions include lack of time to ponder its details; fear of repercussions from the supervisors; and a belief in the objectivity of the process - all in all, rendering what was supposed to be an advisory mechanism the ultimate "automated" decision-maker. (p. 08)

Fussey and Murray (2020) found similar results in the use of FR software by UK police. Green and Chen (2019), in turn, carried out a controlled experimental study concerned with verifying whether the study participants would be able to interpret the outputs of risk assessment systems and improve the assessments generated by them. In general, the participants failed to do both.

The studies above suggest that the human ability to oversee AI systems is limited, even when there are clear guidelines on how it should proceed. Aware of this limitation, public managers must be prepared to assess whether "intelligibility" is minimally guaranteed before and during the use of AI applications. If it is not, it is mandatory—mainly due to the principle of "non-maleficence"—not to use AI systems or to suspend their use if it has already begun. Making such assessments, however, is not trivial, as Sun and Medaglia (2019) demonstrate. The authors investigated the adoption of IBM's Watson system in a public hospital in China. More precisely, they analysed the challenges perceived by three actors involved in this process: public policymakers, doctors and hospital managers, and information technology managers. One of the authors' main conclusions was that the challenges pointed out by the three groups were a little coincident and, in some cases, even divergent. This finding indicates that aligning stakeholders' expectations (and their actions upon them) can be challenging during the adoption of AI applications, making risk assessment extremely difficult, especially when evaluating the technicalities that drive their intelligibility.

2.3.4. Conclusion

The works above make it clear that the use of AI technologies in the public sector is a notoriously delicate process, not allowing for the possibility of adverse

results to be considered complete surprises. For this reason, the "Responsible AI" index prepared by Shearer et al. (2020) does not seem appropriate since the nine indicators used by the authors allow evaluating only the results of the uses of AI applications and not the processes and factors that underpin them. Furthermore, the indicators mobilised are not suitable for studying subnational governments.

Cognizant of the limitations of Shearer et al.'s (2020) framework, we created a responsibility index with information on the entire process of using the technology, including steps before its start. Presented in Section 3, our index refers only to FR systems and to the public transportation system. These research focuses were guided by Instituto Igarapé's (2019) and Brandão and Oliveira's (2021) research findings. As we have seen, these two works identified that using FR technologies in public transport is an expressive phenomenon in the Brazilian public sector. The scientific community should turn to it for two reasons.

Firstly, Brazilian municipalities generally have fewer institutional resources than the Union and states to deal with Information and Communications Technology challenges. Secondly, FR technologies are a high-risk technology, as many systems in the market correctly recognise white people's faces but not black people's (Hao, 2019; Silva, 2020). Buolamwini and Gebru (2018) "show that darker-skinned females are the most misclassified group (with error rates of up to 34.7%). The maximum error rate for lighter-skinned males is 0.8%" (p. 01). Outputs like these can lead the State to become—voluntarily or involuntarily—a reproducer of social wounds, such as racism and misogyny. After all, "unfavourable stereotypes of blacks, for example, are reinforced every time a black student is wrongly taken as a possible fraudster of a free student pass to which s/he is entitled to or when a young black man is confused by technology, leading the police to approach him" (Brandão, Oliveira, & Júnior, 2022).

3. Methodology, Results and Discussion

3.1. Data Collection Strategy and General Results

Between August and October 2021, we requested information from 30 municipalities about using FR systems in public transport to prevent fraud in discounts and gratuities guaranteed by law to specific audiences, such as students and the elderly. The information requests were addressed to the public transportation offices of all state capitals and all municipalities with at least one million inhabitants. They were anchored in the Access to Information Law (Lei de Acesso à Informação—LAI, in the Portuguese acronym). This legislation gives any citizen the right to request information from the State about its activities. For this paper, we just highlight that this request can be made through different channels, such as in-person, telephone, and electronic platforms. In our research, we used the latter channel to request and receive the requested information.

Going on, it must be said that, in Brazil, some public policy areas are operated directly by the public administration. This is the case with public safety, for in-

stance. In other areas, such as municipal public transport, the Public Power is the Granting Power as it grants public or private companies the legal right to provide public services. For this reason, we could have addressed our request for information to the companies that offer transportation services. We preferred to approach the municipal bodies responsible for the transportation sector for two reasons.

Firstly, we understand these bodies integrate the Granting Authority and, therefore, must be aware of the operations of the concessionary companies, including the technical aspects of their operations. Our understanding is based on Article 30, Item V, of the Federal Constitution and on Laws No. 8,987/1995 and 11,079/2004. In the second place, based on Floridi et al. (2018), one can assert that merely ensuring that private and public companies comply with their legal obligations is not the most the State can and should do to ensure the ethical and responsible use of AI for elaborating and implementing public policies. In other words, even when private companies participate in the public policy cycle, the State is ethically accountable for AI outcomes.

By December 2021, four municipalities completely ignored our request for information (Belém, Florianópolis, Macapá, and Natal). In six towns, technical problems linked to the electronic platforms through which LAI is operationalised made it impossible to obtain answers (Cuiabá, São Gonçalo, Fortaleza, Recife, São Luís, and Teresina). The other 20 municipalities responded fully or partially to the questionnaire. Out of them, four stated they do not use the technology in question (Aracaju, Belo Horizonte, Boa Vista, Vitória), one indicated that it was implementing it (Curitiba), and another one did not make it clear whether or not it uses FR systems in public transport (Goiânia). The other 14 municipalities use FR tools to avoid fraud in discounts and gratuities: Brasília, Campinas, Campo Grande, Guarulhos, João Pessoa, Maceió, Manaus, Palmas, Porto Alegre, Porto Velho, Rio Branco, Rio de Janeiro, Salvador, and São Paulo. (In Campinas, João Pessoa and Rio Branco, the use of FR systems in public transport was suspended during the Covid-19 pandemic, as the use of masks negatively interferes with the functioning of the technology.)

In general, all of them described a similar five-step process related to the technical functioning of FR systems: (i) people entitled to gratuity or reduced fares in public transport have an electronic card that attests to their legal entitlement—when they go inside a bus, they must have this card with them; (ii) inside the bus, this card must be used on an electronic turnstile that divides the bus entrance area from the passenger area; (iii) the electronic turnstile is equipped with a camera that captures the beneficiary picture; (iv) the FR system compares this picture with a picture of reference of the beneficiary that is stored in a database of all holders of gratuity/reduced fares; (v) if incompatibility is detected, the images must be submitted to human visual inspection.

3.2. Data Collection Instrument, Analysis Strategy and Detailed Results

Our questionnaire comprised about 40 questions (including primary and condi-

tional questions) and was based on guides for the responsible use of AI in the public sector, such as the works of Reisman, Schultz, Crawford and Whittaker (2018) and Leslie (2019). The questions were divided into six sections: (i) general information about the use of FR systems; (ii) general characteristics of the FR system; (iii) measures adopted before the employment of the system; (iv) measures adopted to make the use of FR aligned with the purposes of LGPD (the Brazilian General Data Protection Law); (v) how fraud-checking works; (vi) what happens when possible fraud is detected and the number of frauds identified by the system. Methodological Annex 2 contains the entire questionnaire.

We initially aggregated the responses around 23 indicators and assigned points to assess them. In general, we attributed [−1] to the lack of response and to responses that mention conducts contrary to the responsible use of FR; we reserved [0] for neutral replies or responses that, due to insufficient details, we were unable to assess; and, finally, we assigned [+1] to responses that indicate careful behaviour about the use of FR. Methodological Annex 3 provides the details of our analysis strategy.

After attributing points to each of the 23 indicators, we had a first indicator ranging from [−23] to [+23] points. We then gave weight [1] to 12 indicators, weight [1.5] to eight indicators, and weight [2] to three indicators. In the first group are indicators that refer to general characteristics of FR systems. The second group comprises indicators regarding the LGPD (the Brazilian General Data Protection Law) and the existence of specific regulations ruling the technology use. Finally, the three indicators with weight 2 refer to three types of constraints that may be caused by FR technology malfunction: moral, physical and legal.

Based on the weight assignment strategy above, we created the Facial Recognition Responsible Use Index (FRRU-I), which ranges from [−30] to [+30] points. This universe consists of 61 points (since 0 is part of the set), divided into five intervals – four of them with 12 points and one of them (the intermediary one) with 13 points. This point division resulted in the creation of the Facial Recognition Responsible Use Scale (FRRU-S). **Table 2** exhibits the 23 FRRU-I indicators and the five intervals of FRRU-S. **Table 3**, in turn, shows the municipalities' scores in the FRRU-I and their position in the FRRU-S.

As shown in **Table 3**, none of the 14 cities reached the "Very high" level in the FRRU-S; five of them achieved "High" level scores; three were at the "Intermediary" level; two obtained a "Low" score, and the remaining four showed a "Very low" score. In this context, it is essential to remember that, out of the 30 municipalities investigated, only four reported not using the technology, and one reported that it was implementing it. Assuming that the use of FR systems may be in progress in the locations that did not take part in our investigation, we have a universe of 25 municipalities in which only 20% seem to take significant care about the responsible use of FR technologies in public transport.

FR responsible use has shown to be challenging in different aspects, as seen in **Table 4**. (In order not to bias the analysis, we disregarded in the construction of

Table 2. Facial recognition responsible use index and facial recognition responsible use scale.

Facial Recognition Responsible Use Index (FRRU-I)

Section	Topic	N°	Indicator	Weight	Assumption supporting positive score
1	General	1	Period of use	1	Respondent provided a detailed answer
		2	Distribution	1	Respondent provided a detailed answer
		3	Regulation	1.5	Local law governs the use of FR systems
2	Software	4	Legal ownership	1	Compliance with public procurement rules (if the concessionary company is a public company)
		5	Name of the owner	1	Respondent provided a detailed answer
		6	Name of the software	1	Respondent provided a detailed answer
		7	Annual cost	1	Respondent provided a detailed answer
3	Implementation	8	Technical studies	1	Before contracting and implementing the technology, technical studies were demanded
		9	Training	1	Public/private servers involved in the implementation and daily usage of the technology received training
4	Data Protection	10	Adequacy	1.5	At least one measure was taken to bring the transportation service into compliance with LGPD
		11	Impact Reports	1.5	Impact reports were produced
		12	Responsible for data processing	1.5	Respondent cited the name of the responsible for data processing
		13	Security measures	1.5	Personal data protection measures exist
		14	Collection of consent	1.5	Consent is collected
		15	Accessing data/ Revoking consent	1.5	There is at least one mechanism that allows personal data subjects to access their data and/or to revoke their consent to data processing
		16	Personal data sharing	1.5	No sharing or it happens only under judicial request
5	Fraud Checking	17	Human intervention	1	Respondent detailed how human intervention works
		18	Moral constraint	2	The warning of possible fraud is done discreetly
		19	Reviewing outputs	1	When fraud is contested, the second check is different from the first one
		20	References	1	Respondent provided details on how the process was structured
6	Relationship with users	21	Restriction of rights	2	Even when the system detects possible fraud, users can conclude their journey
		22	Physical constraint	2	Contestation for possible fraud can be presented either in-person or online
		23	N° of frauds and contestation	1	Respondent provided detailed quantitative information

Score-Range: From [−30] to [+30]

Facial Recognition Responsible Use Scale (FRRU-S) - Levels

Very Low	Low	Intermediary	High	Very High
From [−30] to [−19]	From [−18] to [−7]	From [−6] to [+6]	From [+7] to [+18]	From [+19] to [+30]

Source: Authors' compilation.

Table 3. Municipalities' score in the FRRU-I and position in the FRRU-S.

Municipalities	FRRU-S				
	Very Low [-30] to [−19]	Low [-18] to [−7]	Intermediary [-6] to [+6]	High [+7] to [+18]	Very High [+19] to [+30]
Brasília				+8.5	
Campinas				+7	
Campo Grande		−13.5			
Guarulhos				+18	
João Pessoa	−22				
Maceió[a]	−18.5				
Manaus	−26				
Palmas		−15.5			
Porto Alegre				+7.5	
Porto Velho			−1.5		
Rio Branco			+1		
Rio de Janeiro[b]			+6.5		
Salvador	−29				
São Paulo				+13	

[a]The municipality's score is between "Low" and "Very Low". We chose to place it in the worst position. [b]The municipality's score is between "Intermediary" and "High". We chose to place it in the worst position. Source: Authors' compilation.

Table 4. Number of municipalities with a positive score *per* indicator[a].

Section	Topic	N°	Indicator	Weight	N° of municipalities with a positive score	
					Observed number	Possible number
1	General	1	Period of use	1	10	10
		2	Distribution	1	9	10
		3	Regulation	1.5	1	10
2	Software	4	Legal ownership	1	9	10
		5	Name of the owner	1	6	10
		6	Name of the software	1	8	10
		7	Annual cost	1	3	10
3	Implementation	8	Technical studies	1	4	10
		9	Training	1	1	10
4	Data Protection	10	Adequacy	1.5	4	10
		11	Impact Reports	1.5	2	10
		12	Responsible for data processing	1.5	4	10
		13	Security measures	1.5	1	10

Continued

		14	Collection of consent	1.5	4	10
		15	Accessing data/Revoking consent	1.5	5	10
		16	Personal data sharing	1.5	1	10
5	Fraud Checking	17	Human intervention	1	3	10
		18	Moral constraint	2	3	10
		19	Reviewing outputs	1	2	10
		20	References	1	6	10
6	Relationship with users	21	Restriction of rights	2	9	10
		22	Physical constraint	2	1	10
		23	N° of frauds and contestation	1	5	10

[a]Does not include João Pessoa, Maceio, Manaus and Salvador. Source: Authors' compilation.

Table 4 the four municipalities with a "Very low" score in Table 3.) Out of the 23 FRRU-I indicators, in only seven of them, at least six cities presented a positive score: "Period of use"; "Distribution"; "Legal ownership"; "Name of the owner"; "Name of the software"; "References"; and "Restriction of rights". Among these indicators, only "Restriction of rights" is weighted 2. The others are weighted 1. In other words, out of the 11 indicators weighted 1.5 or 2, in only one, most of the "Low", "Intermediary" or "High"-scored municipalities presented a positive score.

The result indicates that, in general, the municipalities possess basic information on the use of FR technologies by the concessionary companies but that they ignore whether and how the latter deal with this technology's social threats, especially with its potential not to recognise black people's faces correctly, thus submitting them to embarrassing public situations. In this regard, it must be mentioned that, in only two municipalities, the warning of possible fraud is done discreetly, as in Campinas, where letters are sent to the beneficiary holders' homes informing them about misconduct. In the remaining cities, the technology's output is communicated publicly—for instance, by a warning sound or a message on the turnstile's screen. If the technology makes a mistake, everybody surrounding the subject will witness his/her alleged misbehaviour. Considering Buolamwini and Gebru's (2018) research findings, we cannot discharge the possibility that black people have been more exposed than white people to this risk.

As if that was not enough, if a black person is wrongly taken as a possible fraudster by the technology, s/he must face difficulties to challenge this result, as in general, the municipalities give citizens the chance to do that only in-person or only online. In only one city, the two options are available. This is worrisome because people may not have access to electronic devices or the internet, making online contestation exclusionary. On the other hand, demanding public beneficiary holders the obligation to dispute the technology's output in-person may

impose the burden of spending money and time to resolve a problem that may not have been caused by them.

Table 4 also reveals that municipalities have faced difficulties dealing with the legal aspects of FR responsible use. Only one city has specific legislation governing the use of the technology. Some of the other nine municipalities have ordinances and decrees, but, according to Reis, Almeida, Dourado and Silva (2021), instruments like these are insufficient to contain FR technologies' invasive potential. Therefore, we do not consider their existence as an indicator of responsible use.

Besides that, municipalities have little control over the compliance of concessionary companies with LGPD and, thus, over their conduct to promote data protection. Among the cities included in Table 4, Campo Grande, Palmas, Porto Velho and Rio Branco could not provide any information on this topic, stating that we should contact the concessionary companies to get it. (Excluded from Table 4, João Pessoa, Maceió, Manaus and Salvador gave us the same orientation). In other cases, such as Campinas, Guarulhos and Porto Alegre, we were also referred to them, but the municipalities could give us important pieces of information.

As Brandão, Oliveira, Peres, Júnior, Papp, Veiga et al. (2022) point out, the data protection results suggest that "coordination problems among essential actors for the construction and operation of algorithmic transparency may have been ongoing in the municipal public transportation system, which indicates that future studies must investigate whether, or not, concession contracts can be a hindrance to the promotion of algorithmic transparency" (p. 13). The authors pay special attention to the low number of municipalities in which impact reports are produced, and consent collection happens. According to them, when using FR technologies to operationalise public policies, stakeholders are not legally obliged to adopt these two mechanisms, but they should do so. The production of impact reports, Brandão, Oliveira, Peres, Júnior, Papp, Veiga et al. (2022) go on, gives stakeholders the chance to make it clear how they try to prevent or mitigate errors commonly made by FR technologies, like the ones identified by Buolamwini and Gebru (2018). The collection of consent, in turn, makes clear that the data subjects know for what purpose their personal pieces of information (in this case, a picture of their faces) will be used.

Among the indicators weighted 1, we highlight that a reduced number of municipalities have information on how the human intervention works in fraud-checking. As we have seen in Section 2, it is not clear whether, or not, this kind of supervision can be effective in correcting eventual errors of AI systems. It does not mean that public administration should give up on evaluating "intelligibility" or creating protocols that improve the interaction between public servants with these systems. In this regard, we highlight the cases of Campinas and Porto Velho. In the former, a 'biometric engine' does the first comparison automatically (without human intervention). Human agents check images with a distortion rate higher than 80%. Porto Velho, in turn, pointed out that two dif-

ferent employees must check possible fraud identified by the FR system.

As a final note on our results, it is vital to comment that we identified that FR errors might not immediately restrict citizens' rights since beneficiary holders can conclude their journey when the system identifies possible fraud. We also identified that, in general, the public transportation departments know which software is used in the city. Future studies should investigate technical information about the systems adopted, such as their accuracy rate.

The multi-faceted challenges faced by municipalities in using FR technologies responsibly reveal that, at least in public transport, they are not well prepared to promote either algorithmic accountability or intelligibility, which drives us to three research questions:

1) Do Brazilian largest cities not have the necessary bureaucratic structures for dealing with FR systems' risks, or due to the novelty of the technology, do they still not know how to mobilise the administrative structures they have to promote FR responsible uses?

2) Should these challenges be attributed to the technology itself? In other words, do public departments face diminished difficulties when using other AI applications that support decision-making?

3) The 30 municipalities we investigated are among the most prominent in Brazil. If they face difficulties getting concessionaires to use FR systems responsibly, what will be the reality in medium and small municipalities that often have low administrative capacities?

Future studies should answer these questions, trying to understand whether institutional variables—such as the level of qualification of employees of municipal transport secretariats, the financial resources available to these bodies, and the quality of planning and management instruments—positively impact the responsible use of FR systems and other AI decision-making technologies in different public policy areas. Without this understanding, all the efforts that different institutions have made in publishing guidelines on the responsible use of AI in the public sector (Desouza, 2018; Langevin & Fassio, 2022) can be as limited as the impact of the ACM Code of Ethics in the area of computing machinery, as observed by McNamara, Smith and Murphy-Hill (2018), simply because we do not know which are the critical factors for public offices absorbing practical orientations in using AI responsibly. In such a scenario, this technology will hardly be a reliable and effective instrument for promoting human agency and societal cohesion.

4. Conclusion

The literature on AI systems for designing and implementing public policies is under development. Therefore, little is known about how the public sector can and should use AI responsibly and in what situations it has been able to do so. Concerned with partially filling these gaps, we reviewed a set of works dealing with ethics in AI and public policies. We concluded that the responsible use of

AI technologies in the public sector is independent of its results and can be considered inattentive or even contrary to human dignity when it happens without safeguards that protect citizens' rights against voluntary or involuntary State actions guided by the technology. Upon this understanding, we then investigated how FR systems have been used in the most prominent Brazilian cities to prevent fraud in discounts and gratuities guaranteed to specific groups. We revealed a reality in which the level of responsibility in the technology's use cannot be considered high and the challenges to such use are multi-faceted. Based on these findings, we concluded that future studies should investigate the institutional capacities of local governments to responsibly use different AI applications that involve decision-making.

Acknowledgements

The authors of this work would like to thank the C4AI-USP and the support from the São Paulo Research Foundation (FAPESP grant #2019/07665-4) and from the IBM Corporation. The authors also thank Professors Cristina Godoy Bernardo de Oliveira, Sarajane Marques Peres and João Paulo Candia Veiga and the researchers Leôncio da Silva Júnior and Marco Papp for their invaluable comments on algorithmic transparency and on previous versions of the Facial Recognition Responsible Use Index (FRRU-I), and finally João Lucas Oliveira for his support in data gathering.

Methodological Annexes

Mentioned throughout the article, the Methodological Annexes 1, 2 and 3 can be found in the following link: https://drive.google.com/drive/folders/18kZzmectVeV13y3-vDYmV53D1KRDn5zT?usp=sharing.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Ada Lovelace Institute, AI Now Institute, & Open Government Partnership (2021). *Algorithmic Accountability for the Public Sector—Learning from the First Wave of Policy Implementation*. Ada Lovelace Institute, AI Now Institute and Open Government Partnership.

Aitken, M., Toreini, E., Carmichael, P., Coopamootoo, K., Elliott, K., & Moorsel, A. (2020). Establishing a Social License for Financial Technology: Reflections on the Role of the Private Sector in Pursuing Ethical Data Practices. *Big Data & Society, 7,* 1-15. https://doi.org/10.1177/2053951720908892

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine Bias—There's Software Used across the Country to Predict Future Criminals. And It's Biased against Blacks. *ProPublica*, Online Edition.

Berryhill, J., Heang, K., Clogher, R., & McBride, K. (2019). Hello, World: Artificial Intel-

ligence and Its Use in the Public Sector. *OECD Working Papers on Public Governance, 36,* 1-185.

Brandão, R. (2022). Artificial Intelligence, Human Oversight, and Public Policies: Facial Recognition Systems in Brazilian Cities. In I. Kiringa, S. Gambs, & K. H. Kalala (Eds.), *Proceedings of the 35th Canadian Conference on Artificial Intelligence* (Article ID: 2022G6). Canadian Artificial Intelligence Association. https://doi.org/10.21428/594757db.16fe1022

Brandão, R., & Oliveira, J. L. (2021). Reconhecimento Facial e Viés Algorítmico em Grandes Municípios Brasileiros. In *II Workshop sobre as Implicações da Computação na Sociedade* (pp. 121-127). Sociedade Brasileira de Computação. https://doi.org/10.5753/wics.2021.15970

Brandão, R., Oliveira, C., Peres, S., Júnior, L., Papp, M., Veiga, J. P., Beçak, R., & Camargo, L. (2022). Artificial Intelligence, Algorithmic Transparency and Public Policies: The Case of Facial Recognition Technologies in the Public Transportation System of Large Brazilian Municipalities. In Springer, *11th Brazilian Conference on Intelligent Systems – BRACIS* (On Press). Brazilian Institute of Data Science.

Brandão, R., Oliveira, J. L., & Júnior, L. (2022). *Reconhecimento Facial, Viés Algorítmico e Intervenção Humana no Transporte Municipal*. (Unpublished Manuscript/Under Peer-Review).

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (PMLR 81)* (pp. 77-91). https://proceedings.mlr.press/v81/buolamwini18a.html

Busuioc, M. (2020). Accountable Artificial Intelligence: Holding Algorithms to Account. *Public Administration Review, 81,* 825-836. https://doi.org/10.1111/puar.13293

Calo, R., & Citron, D. (2020). The Automated Administrative State: A Crisis of Legitimacy. *Emory Law Journal,* Forthcoming. https://ssrn.com/abstract=3553590

Canada Government Official Webpage (2021). *Digital Government Innovations*. https://www.canada.ca/en/government/system/digital-government/digital-government-innovations.html

Coelho, J., & Burg, T. (2020). *Uso de Inteligência Artificial Pelo Poder Público*. Departament of Research, Transparência Brasil.

Coelho, R., Guth, F., & Loureiro, M. (2020). Capacidades governamentais municipais e desenvolvimento humano local no Brasil. *Revista Do Serviço Público, 71,* 778-808. https://doi.org/10.21874/rsp.v71i4.4524

Desouza, K. (2018). *Delivering Artificial Intelligence in Government: Challenges and Opportunities*. IBM Center for the Business of Government.

Eggers, W., Schatsky, D., & Viechnicki, P. (2017). *AI-Augmented Government: Using Cognitive Technologies to Redesign Public Sector Work*. Deloitte Center for Government Insights, Deloitte University Press.

Engin, Z., & Treleaven, P. (2019). Algorithmic Government: Automating Public Services and Supporting Civil Servants in Using Data Science Technologies. *The Computer Journal, 62,* 448-460. https://doi.org/10.1093/comjnl/bxy082

Engstrom, D., Ho, D., Sharkey, C., & Cuéllar, M.-F. (2020). *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies* (Public Law Research Paper No. 20-54). NYU School of Law. https://doi.org/10.2139/ssrn.3551505

Eubanks, V. (2018). *Automating Inequality—How Hig-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.

Felzmann, H., Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2019). Transparency You Can Trust: Transparency Requirements for Artificial Intelligence between Legal Norms and Contextual Concerns. *Big Data & Society, 6,* No. 1. https://doi.org/10.1177/2053951719860542

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V. et al. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines,* 28, 689-707. https://doi.org/10.1007/s11023-018-9482-5

Floridi, L., Cowls, J., King, T., & Taddeo, M. (2020). How to Design AI for Social Good: Seven Essential Factors. *Science and Engineering Ethics, 26,* 1771-1796. https://doi.org/10.1007/s11948-020-00213-5

Fussey, P., & Murray, D. (2020). Policing Uses of Live Facial Recognition in the United Kingdom. In A. Kak (Ed.), *Regulating Biometrics—Global Approaches and Urgent Questions* (pp. 78-85). AI Now.

Green, B. (2021). The Flaws of Policies Requiring Human Oversight of Government Algorithms. *Computer Law & Security Review, 45,* Article ID: 105681. https://doi.org/10.1016/j.clsr.2022.105681

Green, B., & Chen, Y. (2019). Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (ACM FAT'19)* (pp. 90-99). Association for Computing Machinery. https://doi.org/10.1145/3287560.3287563

Grimmelikhuijsen, S. (2022). Explaining Why the Computer Says No: Algorithmic Transparency Affects the Perceived Trustworthiness of Automated Decision-Making. *Public Administration Review,* 1-22. https://doi.org/10.1111/puar.13483

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines, 30,* 99-120. https://doi.org/10.1007/s11023-020-09517-8

Hao, K. (2019, December 20). A US Government Study Confirms Most Face Recognition Systems Are Racist. *MIT Technology Review*, Online Edition.

Instituto Igarapé. (2019). *Infográfico – Reconhecimento Facial no Brasil*. Instituto Igarapé.

Johnson, K. (2022, March 7). How Wrongful Arrests Based on AI Derailed 3 Men's Lives. *Wired*, Online Edition.

Kirwan, C., & Fu, Z. Y. (2020). *Smart Cities and Artificial Intelligence: Convergent Systems for Planning, Design, and Operations*. Elsevier.

Kuziemski, M., & Misuraca, G. (2020). AI Governance in the Public Sector: Three Tales from the Frontiers of Automated Decision-Making in Democratic Settings. *Telecommunications Policy, 44,* Article ID: 101976. https://doi.org/10.1016/j.telpol.2020.101976

Langevin, C., & Fassio, R. (2022). *Guia de Contratações Públicas de Inteligência Artificial*. Centro para a Quarta Revolução Industrial, World Economic Forum.

Leslie, D. (2019). *Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector*. Public Policy Programme, Alan Turing Institute. https://doi.org/10.2139/ssrn.3403301

Martin, K. (2018). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics, 160,* 835-850. https://doi.org/10.1007/s10551-018-3921-3

McNamara, A., Smith, J., & Murphy-Hill, E. (2018). Does ACM's Code of Ethics Change Ethical Decision Making in Software Development? In *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2018)* (pp. 729-733). Associa-

tion for Computing Machinery. https://doi.org/10.1145/3236024.3264833

Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender Systems and their Ethical Challenges. *AI & Society, 35,* 957-967. https://doi.org/10.1007/s00146-020-00950-y

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics, 26,* 2141-2168. https://doi.org/10.1007/s11948-019-00165-5

Nunes, P. (2019). Novas Ferramentas, Velhas Práticas: Reconhecimento Facial e Policiamento no Brasil. In CeSEC [Centro de Estudos de Segurança e Cidadania] (Ed.), *Retratos da Violência: Cinco meses de monitoramento, análises e descobertas – Junho a Outubro* 2019 (pp. 67-70). CeSEC – Centro de Estudos de Segurança e Cidadania.

Organisation for Economic Cooperation and Development (2019). *Legal Instruments*. https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449

Reis, C., Almeida, E., Dourado, F., & Silva, R. (2021). *Relatório sobre o uso de tecnologias de reconhecimento facial e câmeras de vigilância pela administração pública no Brasil*. Department of Research, LAPIN [Laboratório de Políticas Públicas e Internet].

Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*. AI Now.

Robbins, S. (2019). A Misdirected Principle with a Catch: Explicability for AI. *Minds and Machines, 29,* 495-514. https://doi.org/10.1007/s11023-019-09509-3

Robinson, S. (2020). Trust, Transparency, and Openness: How Inclusion of Cultural Values Shapes Nordic National Public Policy Strategies for Artificial Intelligence (AI). *Technology in Society, 63,* Article ID: 101421. https://doi.org/10.1016/j.techsoc.2020.101421

Sharma, G., Yadav, A., & Chopra, R. (2020). Artificial Intelligence and Effective Governance: A Review, Critique and Research Agenda. *Sustainable Futures, 2,* Article ID: 100004. https://doi.org/10.1016/j.sftr.2019.100004

Shearer, E., Stirling, R., & Pasquarelli, W. (2020). *Government AI Readiness Index 2020*. IDRC & Oxford Insights.

Silva, T. (2020). Visão Computacional e Racismo Algorítmico: Branquitude e Opacidade no Aprendizado de Máquina. *Revista da ABPN, 12,* 428-448. https://doi.org/10.31418/2177-2770.2020.v12.n.31.p428-448

Stahl, B., & Wright, D. (2018). Ethics and Privacy in AI and Big Data: Implementing Responsible Research and Innovation. *IEEE Security & Privacy, 16,* 26-33. https://doi.org/10.1109/MSP.2018.2701164

Sun, T., & Medaglia, R. (2019). Mapping the Challenges of Artificial Intelligence in the Public Sector: Evidence from Public Healthcare. *Government Information Quarterly, 36,* 368-383. https://doi.org/10.1016/j.giq.2018.09.008

Vakkuri, V., Kemell, K.-K., & Abrahamsson, P. (2020). ECCOLA—A Method for Implementing Ethically Aligned AI Systems. In *46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)* (pp. 195-204). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/SEAA51224.2020.00043

Winfield, A., Michael, K., Pitt, J., & Evers, V. (2019). Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems. *Proceedings of the IEEE, 107,* 509-517. https://doi.org/10.1109/JPROC.2019.2900622

Yigitcanlar, T., Corchado, J. M., Mehmood, R., Li, R. Y. M., Mossberger, K., & Desouza, K. (2021). Responsible Urban Innovation with Local Government Artificial Intelligence (AI): A Conceptual Framework and Research Agenda. *Journal of Open Innovation:*

Technology, Market, and Complexity, 7, Article No. 71.
https://doi.org/10.3390/joitmc7010071

Yu, H., Shen, Z., Miao, C., Leung, C., Lesser, V., & Yang, Q. (2018). Building Ethics into Artificial Intelligence. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence* (pp. 5527-5533). International Joint Conferences on Artificial Intelligence Organization. https://doi.org/10.24963/ijcai.2018/779

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Public Affairs.