Artificial intelligence, human oversight, and public policies: facial recognition systems in Brazilian cities

Rodrigo Brandão

C4AI USP-FAPESP-IBM

University of São Paulo, Brazil

1. Research problem

According to Berryhill et al. [1], Artificial Intelligence (AI) may be defined as "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments". Given these qualities, AI systems have been increasingly used for formulating and implementing public policies.

Australia has applied Machine Learning (ML) techniques to satellites so they may learn to recognize and distinguish between human-managed and natural areas. As such, the country hopes for an early identification of specific problems, such as the advance of agricultural land use in biodiversity areas. Belgium, in turn, has used AI to increase citizen participation in public policymaking. Based on ML and Natural Language Processing (NLP) techniques, the CitizenLab platform automatically compiles into clusters millions of public policy suggestions presented by the country's citizens through different participation platforms. The system also identifies the central theme and the geographic and demographic characteristics of each cluster. Information of this type can be useful for more assertive and efficient public sector actions. Portugal and Latvia are also using AI to bridge the gap between the public sector and citizens. Both countries have adopted a widespread use of chatbots in the provision of public services.

The positive examples above, however, clash with increasing reports of misuse of AI within public administrations. In China, this technology has allowed the government to compile citizen information from the Internet, thus classifying individuals as potential government opponents and identifying them in public demonstrations through the use of ML-based facial recognition systems. Evidence suggests that the US has used AI to target and surveil specific social groups. Eubanks [2], for instance, has mapped and analyzed several local governments and their use of emerging technologies such as AI to closely monitor the routine of poor people, seeking justifications to deny them social benefits or even imprison them.

On top of the risk of undue State surveillance, there is a second reason for different stakeholders to remain attentive and constantly scrutinize the public sector's use of AI systems. The positive contributions of AI for improving public policymaking are still recent or, as in most cases, not yet fully realized. These uncertainties raise the question: what are the necessary conditions for an effective use of AI by the public sector to improve people's lives?

2. Proposed solution and approach, and description of the progress to date

In an attempt to answer the above question, I have chosen to prepare case studies on the use of AI systems by the public sector. I have thus far focused solely on the Brazilian case, more specifically on the use of facial recognition (FR) systems by the country's major cities to identify fraud in free or discounted public transportation offered to specific social groups, such as students and the elderly.

Empirical analyses about the use of AI by the public sector are on the rise in Brazil, as revealed by Coelho and Burg [3] and Reis et al. [4]. Nonetheless, research remains scarce, and were even rarer when I began my doctoral research in July 2020. To situate myself in this research field, I initially sought to identify the public policy areas that Brazilian municipalities

* brandao-cs@usp.br

have used FR technologies and which – given the theoretical challenges described by Wirtz et al. [5] – seemed to be more critical. I therefore investigated – alongside other researchers – the electronic official diaries of 13 out of 17 municipalities with over one million inhabitants. The methodology used, as well as the results obtained, can be found in Brandão and Oliveira [6].

Broadly speaking, we identified the use of FR systems in six areas: security, transportation, health, social assistance, education, and assorted administrative areas. We opted to focus on public transportation because municipal prerogatives are more pronounced in this area than, for example, safety or health. In seven of the 13 municipalities, we identified laws, decrees, and resolutions referring to the use of FR systems in municipal public transport: Rio de Janeiro, Manaus, Belém, Porto Alegre, Campinas, São Luís, and São Gonçalo. We then performed a legal analysis of the compiled documents, dividing these seven municipalities into two large groups. In legal terms, the first group – comprising Campinas, Manaus, and Rio de Janeiro – seems to be better prepared than the second – comprising São Luís, Porto Alegre, São Gonçalo, and Belém – to address eventual FR systems failures in the control of social benefits in municipal public transport.

Two elements stand out among the laws, decrees, and resolutions in the second group. First, the legal provisions of these municipalities do not clarify if the passenger, once inside the bus, is allowed to continue the trip after the system informs that they are (allegedly) committing fraud, i.e., that they are not the owner of the discount card being used. Secondly, in São Luís, Porto Alegre, São Gonçalo, and Belém, the legal texts provide no indication of the process for blocking discount cards, giving rise to the following question: once the FR system determines a potential fraud, how do the people responsible for this information verify its consistency, converting this into a decision to suspend a citizen's prerogative for free or discounted public transportation?

Both issues above offer two distinct, albeit complementary, research paths. We chose to focus on the second path, i.e., the need for additional investigations into the human oversight of AI systems. We chose this topic because operators responsible for AI systems, especially biometric systems, are usually ill-prepared to accurately assess the quality or fairness of the outputs of these systems, which leads them towards biased or imprecise responses [7].

Otherwise stated, human oversight by FR systems operators may be insufficient to correct eventual errors in the technology, such as algorithmic race and gender biases as identified by Buolamwini and Gebru [8]. Within our research universe, for example, this reality translates into the possibility that FR system operators tend to confirm signs of fraud when the potential fraudster is Black and decide that the system failed when the potential fraudster is white – even if existing guidelines explicitly prohibit racist or sexist behaviors. In light of this possibility, a question was posed to us: has human oversight been able to correct eventual errors in the FR systems of public transportation in major Brazilian cities?

Pursuing empirical elements to answer this question, we assessed the existence of explicit human oversight protocols for counter-checking potential frauds identified by FR systems. To this end, we resorted to the Brazilian Access to Information Act, which allowed us to review the general information use and flow generated by FR systems, in an attempt to understand how technology outputs are converted into human decisions. More specifically, we sought to understand how a possible fraud notice is converted into the suspension of public transportation benefits. We conducted this phase of the research between August and November 2021. Our detailed methodology, as well as our findings, can be found in Brandão et al. [9].

For this paper, we requested information from 30 municipalities. This group included all state capitals and all municipalities with over one million inhabitants. One of the municipalities surveyed was currently implementing FR systems in the city's public transportation while four other municipalities do not use this technology in this specific area. Among the remaining 25 municipalities, 16 (64% of the total) either ignored our communications or were unable to inform us about human oversight for counter-checking

frauds detected by FR systems. We only managed to gather information on this topic from nine municipalities (36% of the total). Nevertheless, even in this case the information was insufficient to verify whether human oversight has been able to correct possible technology errors in public transportation, which could thus contribute to a public sector use of AI systems towards fairer and more accountable and transparent decisions.

3. Next steps

Throughout 2022, I will establish institutional partnerships between the research center in which I have been developing my research thesis and the secretariats of the nine municipalities that provided Brandão et al. [9] with research information, with two objectives:

(i) Access copies of fraud reports generated by FR systems in Brazilian municipalities in order to verify whether these materials contain information on race and gender. If so, I may investigate whether potential frauds identified by technology systems are usually associated with specific social groups, such as Black people;

(ii) Interview public officials of these nine cities who use FR systems to identify whether human counter-checking of fraud claims varies according to gender and/or race of potential fraudsters. The script of these interviews will be based on the contributions of Reisman et al. [10] and Leslie [11].

Acknowledgements

References

[1] J. Berryhill, et al. "Hello, World: Artificial intelligence and its use in the public sector. In: *OECD Working Papers on Public Governance*, n. 36, 2019.

[2] V. Eubanks. *Automating Inequality – How Hig-Tech Tools Profile, Police, And Punish the Poor*. New York: St. Martin's Press, 2018.

[3] J. Coelho and T. Burg. *Uso de inteligência artificial pelo poder público*. Rio de Janeiro: Transparência Brasil, 2020, Technical Report.

[4] C. Reis, et al. *Vigilância automatizada: uso de reconhecimento facial pela administração pública*. Brasília: Laboratório de Políticas Públicas e Internet (LAPIN), 2021, Technical Report.

[5] B. Wirtz, et al. "Artificial Intelligence and the Public Sector — Applications and Challenges". In: *International Journal of Public Administration*, vol.42, n.7, 2019, pp.596-615. DOI: 10.1080/01900692.2018.1498103.

[6] R. Brandão and J.L. Oliveira. "Reconhecimento facial e viés algorítmico em grandes municípios brasileiros". In: *Workshop sobre as implicações da computação na sociedade (WICS)*, 2., 2021. Online. Anais [...]: Sociedade Brasileira de Computação, pp. 122-127. ISSN 2763-8707. DOI: https://doi.org/10.5753/wics.2021.15970.

[7] A. Kak. "The State of Play and Open Questions for the Future". In: *Regulating Biometrics – Global Approaches and Urgent Questions*. New York: AI Now, 2020, p.16-43.

[8] J. Buolamwini and T. Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification". In: *Proceedings of the 1st conference on fairness, accountability and transparency*, PMLR 81 (2018), pp 77-91.

[9] R. Brandão, et al. "Reconhecimento facial, viés algorítmico e intervenção humana: o caso do transporte público em grandes municípios brasileiros". In: *Seminário Internacional Inteligência Artificial – Democracia e Impactos Sociais*, 1., 2021. Available at: https://www.youtube.com/watch?v=Z5hSQGAdRJQ

[10] D. Reisman, et al. *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*. New York: AI Now, 2018, Technical Report.

[11] D. Leslie. *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*. London: The Alan Turing Institute, 2019, Technical Report. DOI: https://doi.org/10.5281/zenodo.3240529

Cover Letter

Name: Rodrigo Brandão de Andrade e Silva

Institution:
University of São Paulo (USP)
C4AI USP-FAPESP-IBM

Supervisor's name and email address:
Prof. Dr. Glauco Arbix
garbix@usp.br
glaucoarbix@gmail.com

Program start date: 16 January 2020

Expected graduation date: 10 January 2025

2-5 keywords describing your work:
Artificial Intelligence; Algorithmic Bias; Facial Recognition; Public Sector; Ethics

References to any related refereed and non-refereed publications and presentations:
Peer-reviewed publications:
Brandão, R.; Oliveira, J. L. (2021). Reconhecimento facial e viés algorítmico em grandes municípios brasileiros. In: WORKSHOP SOBRE AS IMPLICAÇÕES DA COMPUTAÇÃO NA SOCIEDADE (WICS), 2. Online. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2021. p. 122-127. ISSN 2763-8707. DOI: https://doi.org/10.5753/wics.2021.15970

Brandão, R. (2021). O futuro do trabalho – Entre a automação e a integração entre humanos e máquinas. Revista EPTIC – Special Issue: Internet – Concentrarion and regulation, v. 23 n. 3. DOI: https://doi.org/10.54786/revista%20eptic.v23i3.15626 – Available at:
https://seer.ufs.br/index.php/eptic/article/view/15626

Brandão, R. (2020). Inteligência artificial, trabalho e produtividade. RAE-Revista de Administração de Empresas, [S. l.], v. 60, n. 5, p. 378–379. DOI: 10.1590/S0034-759020200508. Available at:
https://bibliotecadigital.fgv.br/ojs/index.php/rae/article/view/81791

Arbix, G., & Brandão, R. (2020). Vigilância, estágio superior do capitalismo. Será?. Estudos De Sociologia, 25(48). https://doi.org/10.52780/res.13404

Non-peer-reviewed publications:
Brandão, R.; Lazarou, A.; Oliveira, J.L.; Júnior, L. (2021). Reconhecimento facial e discriminação algorítmica nos municípios brasileiros. Migalhas, São Paulo, 7 May 2021. Migalhas de Peso. Available at:
https://www.migalhas.com.br/depeso/345092/reconhecimento-facial-e-discriminacao-algoritmica-nos-municipios

Arbix, G.; Candia, J.P.; Camargo, L.; Brandão, R. (2021). A estratégia do governo para investimentos em inteligência artificial é satisfatória? Não. Folha de S. Paulo, São Paulo, 7 May 2021. Tendências & Debates. Available at:
https://www1.folha.uol.com.br/opiniao/2021/05/a-estrategia-do-governo-para-investimentos-em-inteligencia-artificial-e-satisfatoria-nao.shtml.

Arbix, G.; Brandão, R.; Camargo, L. Pandemia, ciência conectada e IA. In: Vainzof, R.; Gutierrez, A. Inteligência Artificial – Sociedade, Economia e Estado.São Paulo: Thomson Reuters, 2021, p.357-373.

Presentations:
Brandão, R.; Oliveira, J. L.; Júnior, L. (2021). Reconhecimento facial, viés algorítmico e intervenção humana: o caso do transporte público em grandes municípios brasileiros. In: SEMINÁRIO INTERNACIONAL INTELIGÊNCIA ARTIFICIAL – DEMOCRACIA E IMPACTOS SOCIAIS, 1., São Paulo. Available at:
https://www.youtube.com/watch?v=Z5hSQGAdRJQ – This presentation was selected by the organizers of the event to be converted in a book chapter during 2022.

Brandão, R.; Vivas, P.; Camargo, L. (2021). Estratégias Nacionais de Inteligência Artificial: Uma Comparação das Políticas Públicas Voltadas ao Futuro do Trabalho. In: SEMINÁRIO INTERNACIONAL INTELIGÊNCIA ARTIFICIAL – DEMOCRACIA E IMPACTOS SOCIAIS, 1., São Paulo. Available at:
https://www.youtube.com/watch?v=LQfRFY96e2E – This presentation was selected by the organizers of the event to be converted in a book chapter during 2022.

Roundtables (organization and mediation):
The Regulation of AI in Brazil
Available at: https://www.youtube.com/watch?v=pW3OJ1r41tQ
27 October 2021
Eduardo Bismarck – Federal Deputy
Fabio Rua – IBM, Director, Government Relations

The Social Impacts of Facial Recognition: The Case of Public Security in Brazil
Available at:
http://www.iea.usp.br/eventos/reconhecimento-facial-seguranca-publica;
https://www.youtube.com/watch?v=8Co4qiwefr0
Prof. Dr. Lívia Ruback – Universidade Federal Rural do Rio de Janeiro (UFRRJ)
Prof. Dr. Pablo Nunes – Centro de Estudos de Segurança e Cidadania (CESeC)

Letter of support from thesis advisor:

Assessment of the current status thesis research: My student Rodrigo Brandão began his doctoral thesis research in July 2020. Initially, he focused on a literature review on two different research fronts: (i) the ethical challenges regarding the use of AI by the public sector; (ii) the operational challenges in the use of AI technologies in the formulation and implementation of public policies.

Attentive to the theoretical and empirical challenges that these literatures presented, between October 2020 and January 2021 Mr. Brandão conducted an initial empirical data collection on the use of AI by Brazil's public sector. Leading a team with two law scholars, he analyzed official city resolutions regarding the use of facial recognition technologies to prevent fraud in free or discounted public transportation offered to specific social groups, such as students and the elderly. Between February and May 2021, he analyzed the collected documents, and, in July 2021, presented this analysis at a workshop held by the Brazilian Computer Society.

Based on the critical feedback, he structured and implemented the second phase of his research. Once again at the head of a team of law scholars, he used a legal device in Brazil – the Access to Information Act – to access information on the use of facial recognition technologies in public transportation in major cities. The research took place between August

and November 2021, and he presented his findings in December 2021 during the international seminar "ARTIFICIAL INTELLIGENCE: DEMOCRACY AND SOCIAL CONSEQUENCES", promoted by the Humanities area of the C4AI USP-FAPESP-IBM. Based on the feedback from this academic event, Mr. Brandão has focused on: (i) concluding the case study on facial recognition systems in public transportation in major Brazilian cities; (ii) confronting his empirical research findings with the previously outlined theoretical challenges.

As laid out above, the candidate has shown commitment, maturity and, above all, interest in multidisciplinary research and in sharing his study findings with the academic community. Moreover, he shows confidence in conducting his doctoral thesis research.

Expected contributions and graduation date: Academic research on the use of AI by the public sector remain rare in Brazil. I am confident that Mr. Brandão's thesis will be vital to offset the Brazilian gap on empirical information regarding this subject and, most importantly, advance the theoretical debate on the ethical challenges in the use of AI by the public sector.

Mr. Brandão's consistent approach to constructing sociological questions and testing hypotheses stands out not only among other researchers under my supervision, but also among other students in my graduate courses. The articles he wrote as well as the debates he organized and attended demonstrate his progress in a short period of time (July 2020 to December 2021). I am confident he will expand into new empirical and theoretical ground until January 2025, when he is set to deliver his doctoral thesis.