Reconhecimento facial e viés algorítmico em grandes municípios brasileiros

Rodrigo Brandão[1], João Lucas Oliveira[2]

[1]Departamento de Sociologia, Universidade de São Paulo (USP)
Av. Professor Luciano Gualberto, 315, Butantã, São Paulo, SP, 05.508-010

[2]Departamento de Filosofia do Direito, Universidade de São Paulo (USP)
Largo São Francisco, 95, Centro, São Paulo, SP, 01.005-010

brandao-cs@usp.br, joao_lso10@hotmail.com

Abstract. *Are Brazilian governments aware of the social risks of facial recognition applications when they use them? To gather elements that allow us to answer this question, we investigated the digital official diaries of 13 of the 17 Brazilian cities with more than one million inhabitants. Based on the collected material, we preliminary analysed the use of facial recognition in the sector of public transport. We found that some local governments seem to be better prepared than others to deal with the risks in question: in the face of alleged cases of fraud, they allow the user to follow his/her journey, and their legal documents bring minimum guidelines on the role of human work in the process of fraud revision.*

Resumo. *Ao utilizar tecnologias de reconhecimento facial, o setor público brasileiro mostra-se atento aos riscos sociais dessa tecnologia? A fim de reunir elementos que nos permitam responder a essa pergunta, investigamos os diários oficiais digitais de 13 dos 17 municípios brasileiros com mais de um milhão de habitantes. Analisamos preliminarmente o uso do reconhecimento facial no transporte público. Constatamos que alguns municípios parecem mais bem preparados do que outros para lidar com os ricos em questão: diante de supostos casos de fraude, eles permitem ao usuário(a) seguir viagem e, em seus textos legais, apresentam orientações mínimas sobre o papel do trabalho humano no processo de revisão das fraudes.*

1. Introdução

A inteligência artificial (IA) pode ser definida como "um sistema baseado em máquina que pode, para um dado conjunto de objetivos definidos por humanos, realizar predições, recomendações ou decisões que influenciam ambientes reais ou virtuais. [...] Além disso, são 'máquinas que performam funções cognitivas como as humanas'" (Berryhill et al., 2019, p. 12, tradução nossa). Os sistemas de reconhecimento facial (RF), por sua vez, podem ser entendidos como um exemplo de aplicação de IA, à medida que podem ser treinados para reconhecerem atributos físicos diversos e, assim, cumprirem um objetivo específico: captar a imagem de um rosto humano e estimar a porcentagem de semelhança entre esta imagem e a imagem de um banco de dados de referência, como, por exemplo, a de um banco de fotos de estudantes com direito à gratuidade no transporte público.

Os sistemas em questão vêm sendo alvo de críticas, pois parecem não funcionar adequadamente quando utilizados com membros de grupos sociais específicos. Ao

analisarem alguns dos principais programas de reconhecimento facial do mercado estadunidense, Buolamwini & Gebru (2018) descobriram que os algoritmos de análise facial baseados em *machine learning* costumam ser treinados em bases de dados desbalanceadas em termos de raça e de gênero. Isso acaba por se traduzir em taxas máximas de erro fortemente desproporcionais entre os diferentes grupos sociais: se, entre os homens de pele mais clara, elas giram em torno de 0,8%, elas saltam para até 34,7% entre as mulheres com tonalidades mais escuras de pele. Achados de pesquisa do *National Institute of Science and Technology* – o equivalente, nos EUA, ao INMETRO brasileiro – vão na mesma direção (MIT Technology Review, 2019).

Ainda assim, tais sistemas vêm sendo usados pelo setor público brasileiro (Transparência Brasil, 2020). Diante disso, cabe indagar: ao se fazer valer de aplicações de RF, a administração pública do país mostra-se preparada para lidar com os riscos sociais de vieses algorítmicos (como os identificados por Buolamwini & Gebru (2018) e pelas análises do NIST)? Nossa hipótese é que ela não se mostra preparada para isso. A fim de reunir elementos que nos permitissem confirmar ou refutar tal hipótese, realizamos uma empreitada de natureza empírica junto a diários oficiais eletrônicos de municípios brasileiros. Antes de descrevermos tal empreitada, bem como os resultados preliminares aos quais chegamos, faz-se necessário situar nossa pesquisa no universo ao qual ela pertence, qual seja, os estudos sobre as regulações necessárias para que a IA, em geral, e os sistemas de RF, em particular, possam ser usados de modo responsável pelo setor público para a formulação, implementação e/ou avaliação de políticas públicas.

Vale registrar ainda que, baseado em Shearer, Stirling & Pasquarelli (2020), o presente trabalho entende a expressão "uso responsável de aplicações de IA pelo setor público" como um uso que, entre outros elementos, se pauta pela inclusividade – ou seja, pela tentativa de não acentuar desigualdades de diferentes tipos previamente existentes – e pelo respeito à privacidade dos titulares dos dados com os quais opera.

2. Aplicações de IA e de RF no setor público: breve revisão da literatura

A literatura sobre usos de aplicações de IA, em geral, e de sistemas de RF, em particular, pelo setor público pode ser dividida em três grandes grupos. O primeiro deles preocupa-se em investigar como essas tecnologias vêm sendo usadas contra grupos sociais específicos. Eubanks (2018), por exemplo, procura explicitar como, mesmo nos Estados Unidos, um país incontestavelmente democrático, a tecnologia pode ser usada para controlar a vida dos mais pobres de um modo que (ainda) não pode ser utilizada para vigiar e, eventualmente, punir os mais ricos. No polo oposto, investigadores como Berryhill et al. (2019) apontam quão estratégica a IA pode ser para a melhoria de serviços públicos.

Um terceiro setor da literatura preocupa-se em mapear e sistematizar os usos feitos pelas administrações públicas de diferentes países. No caso brasileiro, Coelho & Burg (2020), por exemplo, identificaram que o poder público federal, isso é, os poderes executivo, legislativo e judiciário tomados em conjunto, utilizam 44 ferramentas que se baseiam em sistemas de IA, sendo que 28 delas (64%) ajudam os membros dos três poderes a tomarem decisões, e, portanto, impactam – direta ou indiretamente – os diferentes atores da sociedade civil durante a implementação de políticas públicas diversas.

Levantamentos como o realizado por Coelho & Burg (2020) são especialmente importantes para o caso brasileiro, pois dispomos de poucos registros sobre os usos que o setor público nacional faz de aplicações de IA. Sem evidências empíricas como essas,

a literatura brasileira sobre IA vem encontrando dificuldades para desenvolver acúmulos teóricos consistentes como os produzidos por Eubanks (2018) e por Berryhill et al. (2019), e, em decorrência disso, para avaliar quais são as condições, entre nós, para que o setor público se faça valer da IA de um modo responsável. Para suprir em parte essa lacuna, optamos por levantar evidências dos usos feitos pelos municípios brasileiros de sistemas de RF, preocupados em reunir elementos que nos permitissem investigar – ainda que preliminarmente – se esses usos vêm se dando, ou não, de modo responsável.

3. Coleta de dados: material inicial

Para nossa investigação, elegemos como fonte de dados os diários oficiais eletrônicos – uma fonte oficial de informações. Investigamos 13 dos 17 municípios com mais de um milhão de habitantes: São Paulo, Rio de Janeiro, Brasília, Fortaleza, Belo Horizonte, Manaus, Recife, Goiânia, Belém, Porto Alegre, Campinas, São Luís e São Gonçalo. Neles, buscamos as ocorrências da expressão "reconhecimento facial" ou – quando o buscador digital não permitia a utilização de expressões – da palavra "facial". Essa busca foi orientada por uma lógica indutiva de pesquisa. Inicialmente, não buscamos informações específicas que indicassem usos potencialmente (ir)responsáveis de sistemas de RF pelas administrações locais. Optamos por apenas registrar os textos legais de acordo com a área de política pública em que ocorrem, para que, em seguida, conforme discutido na próxima seção, pudéssemos utilizar as informações neles encontradas para aprimorarmos a hipótese de trabalho enunciada na seção "1. Introdução". Em outros termos, deixamos, em um primeiro momento, que o campo de pesquisa nos revelasse as variáveis que parecem mais relevantes para analisá-lo.

A varredura nos diários oficiais eletrônicos ocorreu entre outubro de 2020 e janeiro de 2021 e abarcou um período de quase 11 anos: de janeiro de 2010 a dezembro de 2020 – exceção feita aos casos em que o diário oficial digital teve início após janeiro de 2010 e à cidade do Rio de Janeiro, onde retroagimos até 2006, pelo fato de uma ocorrência de 2010 remeter àquele ano. Em quatro cidades com mais de um milhão de habitantes (Salvador, Curitiba, Guarulhos e Maceió), não foi possível realizar o levantamento, pois o diário oficial digital não permite buscar nem palavra, nem expressão.

A tabela 1 apresenta um panorama geral das informações encontradas em cada uma das cidades pesquisadas. A coluna "Nº de documentos com 'reconhecimento facial' / 'facial'" traz o número de documentos elencados pelos buscadores digitais para o termo e/ou para a palavra em questão. De modo geral, encontramos três tipos de documentos: (i) projetos de lei; (ii) licitações, chamamentos públicos e contratos; e (iii) leis, decretos e resoluções. A coluna seguinte ("Nº de ocorrências relevantes") foi composta a partir de dois tipos de descarte: (i) excluímos os registros em que a palavra "facial" remete a ocorrências não relacionadas a nossas preocupações – como a compra de máscara facial; (ii) desconsideramos também as ocorrências em que a expressão "reconhecimento facial" aparece sem qualquer contexto ou como parte do nome de empresas.

Na sequência, fizemos um movimento de aglutinação de ocorrências, registrado na coluna "Nº de ocorrências relevantes agrupadas". Explicamos as informações ali reunidas a partir de um exemplo: enquanto as diferentes ocorrências de "reconhecimento facial" ligadas à Parceria Público-Privada (PPP) da iluminação pública no Rio de Janeiro aparecem contabilizadas uma a uma na coluna "Nº de ocorrências relevantes", elas são registradas como uma única ocorrência na coluna "Nº de ocorrências relevantes agrupadas". Por fim, a coluna "Áreas das ocorrências relevantes" oferece ao leitor uma

visão geral de onde o poder público municipal dos grandes municípios utiliza o RF.

Tabela 1. O uso de sistemas de reconhecimento facial em grandes municípios

Município	Nº de documentos com "reconhecimento facial" / "facial"	Nº de ocorrências relevantes	Nº de ocorrências relevantes agrupadas	Áreas das ocorrências relevantes
São Paulo	23	20	9	Segurança Pública; Transporte Público; Saúde Pública
Rio de Janeiro	23	22	6	Assistência Social; Saúde; Transporte Público; Educação; Segurança Pública; Arquivo Geral do Município
Brasília	5	5	3	Segurança Pública; Transporte Público
Fortaleza	10	8	1	Educação
Manaus	25	9	3	Transporte Público; Controladoria-Geral do Município; Secretaria de Gestão
Recife	80	0	0	Não se aplica
Goiânia	135	0	0	Não se aplica
Belém	99	1	1	Transporte Público
Porto Alegre	36	4	4	Segurança Pública; Transporte Público; Departamento de Processamento de Dados
Campinas	640	12	2	Transporte Público
São Luís	442	2	1	Transporte Público
São Gonçalo	35	6	1	Transporte Público
Belo Horizonte	64	0	0	Não se aplica

Fonte: Diários oficiais digitais de municípios diversos – *Elaboração própria*

4. Análise dos dados: um primeiro balanço

Embora o levantamento de dados para a pesquisa ainda esteja em curso, o material coletado permite considerações iniciais sobre o posicionamento do setor público diante dos riscos sociais de vieses algorítmicos, bem como confrontos, também iniciais, com a literatura sobre o tema. Com esses objetivos, debruçamo-nos, especificamente, sobre leis, decretos e resoluções na área de transporte público, guiados por duas preocupações: (i) aprimorar nossa hipótese de trabalho; (ii) verificar, de modo preliminar, a capacidade do setor público municipal para lidar com falhas dos sistemas de RF, dando continuidade, assim, a trabalhos anteriores produzidos por nós, como Brandão et al. (2021).

4.1. Primeiros passos para um estudo de caso sobre a intervenção humana em sistemas biométricos no transporte público

A tabela 1 apresenta nove municípios cujos diários oficiais contêm documentos ligados ao uso de RF no transporte público: São Paulo, Rio de Janeiro, Brasília, Manaus, Belém, Porto Alegre, Campinas, São Luís e São Gonçalo. Em São Paulo e em Brasília, tais documentos não correspondem a leis, decretos ou resoluções. Por isso, as duas cidades

não integram a presente análise. As sete restantes podem ser divididas em dois grupos. Em termos legais, o primeiro deles, composto por Campinas, Manaus e Rio de Janeiro, parece mais bem preparado do que o segundo, formado por São Luís, Porto Alegre, São Gonçalo e Belém, para lidar com eventuais falhas dos sistemas de RF no controle de gratuidades no transporte público.

Entre as leis, os decretos e as resoluções dos membros do segundo grupo, dois elementos se destacam. Primeiramente, esses dispositivos legais não deixam claro se, já dentro do ônibus, o(a) usuário(a) pode, ou não, seguir viagem após ser informado pelo sistema de que (supostamente) está cometendo fraude, ou seja, de que não é o titular do cartão de gratuidade que está utilizando. Em segundo lugar, nos quatro municípios ora em tela, os textos legais não trazem quaisquer indícios de como funciona o processo de bloqueio dos cartões de gratuidade, abrindo margem à seguinte indagação: após o sistema de RF apontar uma possível fraude, como as secretarias de transporte realizam o processo de checagem dessa informação?

A fragilidade dos textos legais desses municípios para prevenir e/ou corrigir *a posteriori* falhas do RF encontra seu exemplo máximo em São Gonçalo. Se, no Art. 1º da Lei nº 819/2018, o município determina que "Fica vedada a utilização da identificação biométrica de idosos e pessoas com deficiência com mobilidade reduzida no transporte público municipal", ele estipula, no Art. 4º da mesma lei, que idosos e pessoas com deficiências que os impossibilitem de se submeter à identificação biométrica poderão ter a identidade checada "por meios tecnologicamente adequados, sendo preferencialmente a biometria facial" – como se o RF não fosse um exemplo de identificação biométrica.

Campinas, Manaus e Rio de Janeiro, por sua vez, deixam claro que, em casos de suspeita de fraude, o(a) usuário(a) poderá seguir viagem. Além disso, apresentam detalhes sobre prazos para que o(a) suposto(a) fraudador(a) apresente recurso administrativo contestando a indicação do sistema de RF. Por fim, Rio de Janeiro e Manaus, de um modo bastante literal, e Campinas, de modo não muito explícito, abordam como o trabalho humano integra o processo de checagem de fraudes.

O Decreto nº 46.852/2019, do Rio de Janeiro, apresenta orientações mínimas sobre como deve se dar a interação entre o sistema de RF e os responsáveis por operá-lo. Já a legislação manauara (Lei nº 2.472/2019 e Portaria Nº 043/2019) instituiu uma comissão para avaliar os recursos administrativos interpostos.

Como se pode perceber, o trabalho de campo realizado até o momento revelou dois elementos empíricos que, se explorados em pesquisas futuras, podem contribuir para a realização de acúmulos teóricos sobre vieses algorítmicos no transporte público. O primeiro deles é a existência, ou não, de previsão legal para o(a) usuário(a) seguir viagem diante de (supostas) fraudes apontadas pelos sistemas de RF. O segundo é a interação entre humanos e sistemas tecnológicos na apuração dessas fraudes. Nas próximas etapas da pesquisa, investigaremos esse segundo tópico. Isso porque "[é] comum que indivíduos que revisam resultados não estejam preparados para avaliar com precisão a qualidade ou a justiça dos *outputs* [dos sistemas], e, frequentemente, respondam a predições de modos enviesados e imprecisos" (Kak, 2020, p. 29, tradução nossa).

Ou seja, mesmo no caso de Manaus, em que existe um comitê legalmente constituído e operante para averiguar potenciais falhas dos sistemas de RF, esse colegiado não pode ser tomado como uma espécie de certeza de que os vieses de gênero e/ou de raça dos sistemas de RF são, de fato, corrigidos – mesmo que *a posteriori*, e, portanto,

em uma situação em que o cidadão já se encontra prejudicado.

Cabe investigar, portanto, como operam comitês como o manauara: (i) como os funcionários desses órgãos fazem a comparação entre os dados biométricos (imagens) obtidos nos validadores dos ônibus e os existentes nas bases de dados das quais dispõem?; (ii) existem protocolos específicos que direcionam essa operação?; (iii) como é o treinamento dos funcionários para a realização dessa função?

É possível que a mera existência de protocolos do tipo não seja suficiente para corrigir vieses algorítmicos de raça e/ou de gênero, uma vez que normas legais nem sempre governam a conduta dos operadores de sistemas biométricos (Chaudhuri, 2019). Diante desse risco, faz-se urgente investigar, para o caso brasileiro, a efetividade dos protocolos em questão. Todavia, precisamos saber, antes disso, se tais protocolos existem na área de transporte público. Por isso, assumimos uma hipótese de trabalho mais específica do que a anunciada inicialmente: municípios que, em seus textos legais, apresentam menções explícitas sobre a revisão humana de potenciais falhas da tecnologia (como Rio de Janeiro e Manaus) devem possuir protocolos mais bem estruturados do que municípios cujos textos legais não abordam essa questão (como, por exemplo, São Gonçalo, cujas confusões conceituais sobre tecnologias biométricas e sistemas de RF evidenciam a falta de familiaridade do poder público local com a tecnologia em questão).

Referências bibliográficas

Berryhill, J. et al. (2019). “Hello, World: Artificial intelligence and its use in the public sector”. OECD Working Papers on Public Governance No. 36.

Brandão, R. et al. (2021). “Reconhecimento facial e discriminação algorítmica nos municípios brasileiros”, 7 de maio de 2021. Disponível em: https://www.migalhas.com.br/depeso/345092/reconhecimento-facial-e-discriminacao-algoritmica-nos-municipios

Buolamwini, J., Gebru, T. (2018). “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”. Proceedings of Machine Learning Research 81:1–15. Conference on Fairness, Accountability, and Transparency.

Chaudhuri, B. (2019). “Paradoxes of Intermediation in Aadhaar: Human Making of a Digital Infrastructure,” *Journal of South Asian Studies* 42 (2019):572–587.

Coelho, J., Burg, T. (2020). “Uso de inteligência artificial pelo poder público”. Transparência Brasil.

Eubanks, V. (2018). Automating Inequality. St.Martin's Publishing Group. Edição Kindle.

Kak, A. (2020). “The State of Play and Open Questions for the Future”. *In:* Regulating Biometrics – Global Approaches and Urgent Questions, Editado por Amba Kak, IA Now, Estados Unidos da América.

MIT Technology Review. (2019). “A US government study confirms most face recognition systems are racist”, 20 de dezembro de 2019. Disponível em: https://www.technologyreview.com/2019/12/20/79/ai-face-recognition-racist-us-government-nist-study/ [consultado em 20-02-2021].

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