BRAZILIAN STUDIES PROGRAMME

UNIVERSITY OF OXFORD

LATIN AMERICAN CENTRE

Considerations on the Right to Non-Discrimination and the Use of Facial Recognition Technology by Brazilian Police Forces

Alessandra Prezepiorski Lemos

This is a conference paper, results were presented in Rio de Janeiro, 2019 at the Global Security Challenges Workshop

Brazilian Studies Programme Latin American Centre

St Antony's College, Oxford OX2 6JF

Considerations on the Right to Non-Discrimination and the Use of Facial Recognition Technology by Brazilian Police Forces

Alessandra Prezepiorski Lemos[1]

1.   Introduction

In November 2019, a Commission of Jurists was created by the Brazilian Chamber of Deputies to elaborate draft legislation on the use and processing of personal data in the context of public security. One of the topics being discussed by the commission is how to regulate the use of facial recognition technology (FRT), a technology that is already being used by Brazilian police forces of the States of São Paulo, Rio de Janeiro, Bahia, Paraíba, and Santa Catarina.

The usage of algorithmic technology in law enforcement activities without a robust legal framework can lead to human rights violations. Regarding the use of FRT, specialists have demonstrated the risks this technology poses to the right to non-discrimination (e.g. Ferguson 2017; European Union Agency for Fundamental Rights (FRA) 2019). Any legislation on FRT must address these potential threats by providing safeguards and establishing accountability mechanisms for intentional and unintentional discriminatory practices.

With this paper, the author aims to contribute to the discussion by indicating the threats posed to the right to non-discrimination and suggesting ways to avoid them. Since the discussions around the usage of FRT in Brazil are still in their beginning, the paper will also use international studies and examples to support its claims.

In the first part, it will be presented a brief context of the usage of FRT by Brazilian police forces. Afterward, it will be discussed how bias can creep in during the development, assessment, and usage of FRT. Lastly, some suggestions will be given on how the legislation can address the bias problems discussed.

---

[1] Master of Science in Criminology and Criminal Justice by the University of Oxford (2018). Master of Laws by the Universidade Federal do Paraná (2017).

2.   FRT Use by Brazilian Police Forces

During 2019, police forces of the States of Rio de Janeiro, Bahia, Santa Catarina, and Paraíba started using FRT in their law enforcement activities. According to the research group *Rede de Observatório da Segurança*, the use of FRT enabled the arrest of 151 people between March and October of 2019 (Nunes 2019, 69).

Pablo Nunes clarifies that there is not much information available about arrests circumstances. Only in 62 cases, it was possible to obtain information about the gender of the person arrested - 87,9% of the arrests were of males and 12,1% were of females (Nunes 2019, 69). When it comes to race, only in 42 cases this information was available. Of this total, 90,5% of the arrests were of black people and 9.5% were of white people (Nunes 2019, 69). In a population where the percentage of people who declare themselves as black people is 43,1% and as "pardo" (mixed race) is 7.6% (Instituto Brasileiro de Geografia e Estatística 2013), the demographic profile of the arrests enabled by the FRT seems disproportional and raises a red flag.

Although this data cannot indicate that FRT is being used in a discriminatory way, it is enough to justify the need to discuss the impact that the usage of FRT in public security might have on the right to non-discrimination. The next sections will explore how bias can creep in during the stages of development, assessment, and usage of FRT.

3.   Potential Bias in the Development of FRT

When assessing if an algorithmic technology is human rights compliant, it is essential to understand what kind of inputs are being used to determine the outputs. In the case of facial recognition technology, knowledge about the composition of the database of faces used to train the algorithms is of paramount importance because racial bias can seep into the process, as Andrew Ferguson explicates (2017, 93).

Some studies such as the paper "An Other-Race Effect for Face Recognition Algorithms", published by P. Jonathon Phillips et al. in 2011, suggest that diversity of images from different demographic groups in the

training database is key to ensure that the facial recognition technology will be able to perform with high levels of accuracy, avoiding making discriminatory mistakes.

Phillips et al. conducted two experiments to assess if an other-race effect was present in face recognition technologies. The other-race effect is the aptitude a person has to recognize and remember more easily faces from people of her own race. The first experiment was designed to test the hypothesis that, in an identity matching task, "algorithms will show a performance advantage for faces that characterize the majority race from the geographic region of their origin" (2011, 3). Five East Asian and eight Western algorithms that participated in the Face Recognition Vendor Test 2006 (FRVT 2006) were tested in this experiment. They were confronted with face pairs from one of the FRVT 2006 testing face sets. On average, Western algorithms performed better on Caucasian faces, whereas the East Asian algorithms performed better on Asian faces.

In the second experiment, the researchers used a database with 40 pairs of East Asians and 40 pairs of Caucasian faces, all from young adults (18-35 years old), to test the algorithms. Once again, the results collected in the research seemed to indicate that the national origin of the algorithms influenced their accuracy levels. Although Western and East Asian algorithms performed better on Caucasian Faces with almost the same accuracy rate, East Asian algorithms had higher accuracy rates for East Asian faces (0.75/1.0) than Western algorithms (between 0.5 and 0.55/1.0).

Recent research published by Joy Buolamwini and Timnit Gebru in 2018 has also evidenced the dangers of not having a diverse training database. The authors tested the accuracy of three different software that performed facial recognition (IBM, Microsoft, and Face++) and whose aggregated accuracy levels were extremely high. This accuracy, however, did not hold up when the researchers analysed the results dividing the face images into intersectional subgroups (darker males; darker females; lighters males; lighter females). They found out that all companies performed better on lighter subjects, with an 11.8% to 19.2% difference in performance in comparison with darker subjects. The authors also discovered that all companies performed worst when the face analysed belonged to the subgroup darker female.

FRT inaccuracy is a serious problem because it can lead to false negatives and false positives. Given that inaccuracy rates are higher for minority groups, they are more likely to be mistakenly identified, which

may entail the temporary curtailment of human rights and have stigmatizing consequences for the person wrongfully identified.

The problem of lack of diversity, nonetheless, is not limited to the databases used to train the algorithms. It is also present in the databases used to audit the software and validate whether they are ready to be used.

4.   Potential Bias in the Databases Used to Evaluate the FRT

A benchmark database is a standard database used to audit and validate different algorithms. It allows the comparison of the levels of accuracy of different software. If diversity is important in the creation of a database to train the FRT, it is also necessary for the databases used to benchmark it. Nevertheless, the majority of databases used as benchmark fail to meet this standard. A good example of this is the Labeled Faces in the Wield (LFW) database, the former gold standard benchmark for facial recognition. Its database is composed of 77.5% male faces and 83.5% white faces (Buolamwini, and Gebru 2018, 3). This is probably one of the reasons why the level of accuracy of FRT regarding minority groups is considerably lower even with auditing practices in place.

The underrepresentation of a minority group in a benchmark database is problematic because this means that the benchmark will not be able to assess the real accuracy of the software regarding the underrepresented group. In other words, if there is any bias present in the technology, it will not be addressed.

Even though more diverse benchmark databases are being created, Buolamwini and Gebru are unsure if this should be the focus of future research given that higher levels of accuracy can promote intentionally and unintentionally discriminatory practices. The ethical concerns expressed by the researchers are related to the potential discriminatory usage of FRT which will be addressed in the next section.

5.   The Potential Discriminatory Usage of FRT

Bias can also manifest itself in the watchlists used by police forces and in the choice of places that are going to be surveilled by these technologies.

It is crucial to understand how a watchlist is created and how it is fed to avoid bias. As William Crumbler (2020) argues, "even if an algorithm shows no difference in its accuracy between demographics, its use could still result in a disparate impact if certain groups are over-represented in databases". Transparency over the creation of watchlists and their content, therefore, is essential to guarantee that bias will not creep into the database.

The choice of the places where FRT will be used by police forces is another factor that can lead to discrimination. The address where a person lives or frequently goes to may indicate information about a person´s class, race, sexual orientation, income, wealth, and political sentiment (Ferguson 2017, 9). Therefore, the choice of the places to be surveilled by FRT must be carefully made to ensure that it will not be responsible for discriminating even further communities that are already discriminated against.

The potential discriminatory usage of FRT can either be intentional or unintentional. It is intentional when the focus is to surveil a specific group of people. On the other hand, it can happen unintentionally if it is the unpredicted consequence of adopting certain criteria to decide when and where to use FRT.

The Met Police, the Greater London police force, was accused by NGOs of intentional discrimination because of its use of FRT at the 2017´s Notting Hill Carnival (Dodd 2017). The festival celebrates Caribbean culture and is attended mostly by minority groups. Elena Papamichael (2017) was one of the human rights advocates who condemned the Met Police´s actions. According to her, the Notting Hill Festival has a history of being heavily policed in comparison with other major festivals. Therefore, the choice of using facial recognition technology at the festival was, in her opinion, discriminatory and another example of over-policing of black communities.

Unintentional discrimination can be found if the criteria adopted to decide where to use the FRT is, for instance, the number of arrests carried out in a particular area. Minority/poor neighborhoods are more likely to

have higher arrest figures due to a stricter policing than richer neighborhoods. This means that they are more likely to be the main audience of surveillance even if this was not the initial intent.

Once identified the potential bias in the processes of developing, assessing, and using FRT, it is now possible to discuss how to tackle it.

6. Final Remarks

Even though algorithmic technologies are being promoted as objective and more efficient (e.g. São Paulo State Government 2020), their use can produce consequences that fail to meet this promise, especially in the field of public security. Andrew Ferguson was precise when he stated that in data-driven policing, "each data point translates to real human experience, and many times those experiences remain fraught with all-too-human bias, fear, distrust, and racial tension" (2017, 3).

It is clear from the content of this paper that bias is pervasive and can infiltrate intentionally and unintentionally in the police practices through the usage of FRT. Therefore, the members of the Commission of Jurists have an arduous task ahead of them to create safeguards, standards of use, and accountability mechanisms that will ensure the protection of the right to non-discrimination. To help them in this task, this paper will dedicate its last words to present a few suggestions on how the draft legislation can tackle bias.

The first suggestion is to impose a duty on the police forces to conduct an assessment of the demographic diversity of the databases used to train and audit the facial recognition technology before they decide if they are going to use it in their law enforcement activities. The legislation should provide a framework of principles and directives to guide this analysis. This assessment should be made public to civil society to ensure accountability. If the technology does not perform well regarding all demographic groups, it should not be purchased and used.

A second suggestion is to create a national and independent oversight body that will be responsible for ensuring that the States and Federal police forces are using FRT in a human rights compliant manner. This body could also be responsible for creating a benchmark database that respects the demographic aspects of Brazilian society to assess the accuracy levels of the FRT used by the police forces.

A third suggestion is to prohibit the usage of FRT in certain contexts to avoid bias. The use of FRT should be prohibited in contexts where it is disproportional and have higher risks of being used in a discriminatory manner. For instance, FRT should not be used in protests and manifestations because of its potential chilling effect on minority and dissenting groups.

Preventing FRT bias is not an easy task, but the Brazilian legislators have the chance to create a legal instrument that ensures the protection of human rights and to lead the world debate regarding the regulation of FRT. The adoption of these suggestions can help with achieving this goal.

7. Bibliography

Buolamwini, Joy, and Timmit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparaties in Commencial Gender Classification." *Proceedings of Machine Learning Research* 81: 1-15.

Crumpler, William. 2020. "The Problem of Bias in Facial Recognition." Accessed in August 1, 2020. https://www.csis.org/blogs/technology-policy-blog/problem-bias-facial-recognition.

Dodd, Vikram. 2017. "Met Police to Use Facial Recognition Software at Notting Hill Carnaval." *The Guardian – UK*, August 5, 2017. https://www.theguardian.com/uk-news/2017/aug/05/met-police-facial-recognition-software-notting-hill-carnival.

European Union Agency for Fundamental Rights (FRA). 2019. *Facial Recognition Technology: Fundamental Rights Considerations in the Context of Law Enforcement*. Vienna: European Union Agency for Fundamental Rights (FRA).

Ferguson, Andrew G. 2017. *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York: NYU Press.

Instituto Brasileiro de Geografia e Estatística (IBGE). Censo 2010 – IBGE mapeia a distribuição da população

    perda e parda. *IBGE Official Webpage*, November 6, 2013.

    https://censo2010.ibge.gov.br/noticias-censo.html?busca=1&id=1&idnoticia=2507&t=ibge-mapeia-

    distribuicao-populacao-preta-parda&view=noticia.

Nunes, Pablo. 2019. "Novas Ferramentas, Velhas Práticas: Reconhecimento Facial e Policiamento no Brasil."

    In *Retratos da Violência – Cinco Meses de Monitoramento, Análises e Descobertas (Junho a Outubro –*

    2019), edited by CESEC – Centro de Estudos de Segurança e Cidadania, 67-70. Rio de Janeiro: Rede de

    Observatório da Segurança. http://observatorioseguranca.com.br/wp-

    content/uploads/2019/11/1relatoriorede.pdf.

Papamichael, Elena. 2017. "Face Recognition Technology at Notting Hill Carnival – Another Example of

    Over-Policing of Black Communities that Will Not Make Society Safer." Accessed August 30, 2020.

    https://www.hja.net/face-recognition-technology-notting-hill-carnival-another-example-policing-black-

    communities-will-not-make-society-safer/.

Phillips, P. Jonathan *et al*. 2011. "An Other-Race Effect for Face Recognition Algorithms." *ACM Transactions*

    on Applied Perception 8 (2): 14. https://doi.org/10.1145/1870076.1870082.

São Paulo State Government. Governo inaugural laboratório de reconhecimento facial e digital da Polícia Civil.

    São Paulo State Government Official Webpage, January 28, 2020.

    https://www.saopaulo.sp.gov.br/spnoticias/governo-inaugura-laboratorio-de-reconhecimento-facial-e-

    digital-da-policia-civil/.