



NOTA TÉCNICA

RISCOS DA UTILIZAÇÃO DE TECNOLOGIAS DE RECONHECIMENTO FACIAL PARA FINS DE SEGURANÇA PÚBLICA

2022



AIRES



RAIES



REALIZAÇÃO

AI Robotics Ethics Society na PUCRS - **AIRES na PUCRS**

Rede de Inteligência Artificial Ética e Segura - **RAIES**



AUTORIA

Nicholas Kluge Corrêa - Presidente do Capítulo PUCRS da AIRES

James William Santos - Bolsista RAIES

Paulo Marcelo Pinheiro Pasetti - Bolsista RAIES

Aline Santos Barbosa - Bolsista RAIES

Carolina Del Pino Carvalho - Bolsista RAIES



AIRES na PUCRS:

AI Robotics Ethics Society na PUCRS

<https://www.airespucrs.org/>

airespucrs@airespucrs.org

RAIES:

Rede de Inteligência Artificial Ética e Segura

<https://www.raies.org/>

raies@raies.org

RISCOS DA UTILIZAÇÃO DE TECNOLOGIAS DE RECONHECIMENTO FACIAL PARA FINS DE SEGURANÇA PÚBLICA: UMA NOTA TÉCNICA



No ano de 2022, o prefeito da cidade de Porto Alegre, Sebastião Melo, informou publicamente que planeja ampliar o uso de câmeras de reconhecimento facial em espaços públicos¹. Para os fins de explicação desta nota, iremos nos referir a algoritmos de visão Computacional para reconhecimento facial como **FRT**, isto é, **Facial Recognition Technology**.

No âmbito do Estado do Rio Grande Sul, a Lei n. 15.460, de 26 de março de 2020², já criava o Banco de Dados de Reconhecimento Facial e Digital para a Prevenção ao Desaparecimento de Crianças e Adolescentes. Regulamentada pela Portaria SSP n. 065, de 24 de maio de 2021³, que estabelece objetivos, diretrizes e procedimentos para a operacionalização do Banco de Dados de Reconhecimento Facial e Digital, voltado à Prevenção ao Desaparecimento de Crianças e Adolescentes, no âmbito da Secretaria de Segurança Pública (SSP) e das suas instituições Vinculadas, assim como nos casos de compartilhamento por meio de convênio ou outro instrumento similar.

O objetivo inicialmente proposto para essa tecnologia, utilizada desde o ano de 2021 pelo governo do RS, era, como referido, o de auxiliar na busca por desaparecidos. Entretanto, agora é proposta que a aplicação dessa tecnologia seja ampliada. Assim, técnicas de visão computacional e reconhecimento facial também serão utilizadas para detectar foragidos da Justiça. Atualmente,

1 Disponível em: <https://www.matinaljornalismo.com.br/matinal/reportagem-matinal/governo-rs-reconhecimento-facial-vies-racial/>.

2 Disponível em: <https://leisestaduais.com.br/rs/lei-ordinaria-n-15460-2020-rio-grande-do-sul-cria-o-banco-de-dados-de-reconhecimento-facial-e-digital-para-a-prevencao-ao-desaparecimento-de-criancas-e-adolescentes-e-da-outras-providencias>.

3 Disponível no Diário Oficial do Estado do Rio Grande, nº 119, Porto Alegre, Segunda-feira, 14 de Junho de 2021.

prospectasse o aumento do número de câmeras de vigilância pública em POA, de modo a possibilitar a implementação desta nova aplicação.

É importante ressaltar que o uso de aplicações que utilizam Aprendizagem de Máquina (*Machine Learning*) para fora de suas distribuições de treinamento (*Out-of-Distribution Robustness*)⁴ é um problema comum que enfrentamos em aplicações do mundo real (como a detecção de foragidos da Justiça através de FRTs).

Mudança de distribuição (*Distribution Shift*) refere-se a quando os dados que um modelo trabalha mudam ao longo do tempo (*ou o modelo está sendo usado em um cenário diferente do que foi treinado*), o que faz com que as previsões deste modelo se tornem menos precisas. Assim, de um ponto de vista técnico, esta troca de distribuição (*crianças e adolescentes desaparecidos para foragidos da justiça*) pode acarretar em uma baixa na performance do algoritmo⁵.

Outro ponto importante a ser mencionado é que o uso de biometria⁶ para fins de reconhecimento facial depende da utilização obrigatória de um tipo de dado sensível, isto é, imagens de nossas faces.

Inicialmente, nenhuma empresa ou indivíduo pode fazer uso desse tipo de informação sem consentimento prévio, e explícito, daqueles dos quais os dados se originam. Porém, há situações legais que permitem exceções. Por exemplo, para a utilização em situações ligadas a políticas públicas, e para a preservação da vida ou da integridade física de uma pessoa.

Porém, a salvaguarda de que os direitos fundamentais expressos na Constituição Federal nos confirmam dignidade e proteção frente a atuação do Estado, se enfraquece ao utilizarmos FRTs em espaços públicos.

4 Para mais informações sobre o assunto, recomendamos “Improving Out-of-Distribution Robustness via *Selective Augmentation*”. Disponível em: <https://arxiv.org/abs/2201.00299>.

5 Ao mesmo tempo, vale a pena ressaltar que quase nada sabemos da natureza do algoritmo mencionado. Nossas inferências são baseadas naquilo que sabemos sobre o estado-da-arte em visão computacional. Ou seja, utilização de redes neurais profundas, como CNNs (*Convolutional Neural Networks*), para fins de classificação de imagens.

6 Cálculos que utilizam como base características e medidas do corpo humano.

Ao mesmo tempo, não há transparência alguma no que diz respeito sobre:

- ▲ O tipo de sistema utilizado.⁷
- ▲ O banco de dados utilizado para treinamento do sistema.⁸
- ▲ Os possíveis vieses discriminatórios que tal sistema pode, ou não, possuir.⁹
- ▲ Informações sobre os responsáveis pelo desenvolvimento e monitoração de tal sistema.¹⁰

Pode-se argumentar que esta ausência de transparência¹¹, vai diretamente contra as noções de direitos fundamentais como o ***Direito à Liberdade, o Direito à Igualdade, e o Direito à Segurança***. Sendo que já existem simples ferramentas para se garantir transparência e responsabilização (sem prejudicar a propriedade intelectual) em aplicações que utilizam este tipo de tecnologia.¹²

Os direitos e garantias fundamentais são normas protetivas que têm como objetivo amparar o cidadão da ação do Estado, garantindo assim os requisitos mínimos para que o indivíduo tenha uma vida digna perante a sociedade.

7 Como o paradigma atual em tarefas de Visão Computacional é a Aprendizagem de Máquina, iremos neste documento inferir que estamos falando sobre um modelo de Aprendizagem de Máquina, como uma rede neural profunda (e.g., CNN).

8 Já que modelos de aprendizagem de máquina são treinados a partir de conjuntos de dados, inferimos também que tal sistema foi desenvolvido a partir de algum conjunto de imagens de faces.

9 Modelos desenvolvidos por Aprendizagem de Máquina, quando treinados com dados tendenciosos, podem vir a herdar vieses discriminatórios (tendências de classificação que privilegiam ou prejudicam indivíduos com base em atributos sensíveis) presentes na distribuição de amostras que forma seu conjunto de dados de treinamento.

10 Como modelos estatísticos não podem ser legalmente responsáveis por “decisões falhas”, conexões claras de responsabilização devem existir em processos de automação de decisões de alto risco.

11 O princípio ético mais defendido dentro do estudo da Ética da Inteligência Artificial. Disponível em: <https://arxiv.org/abs/2206.11922>.

12 Para mais informações, recomendamos “Model Cards for Model Reporting”: Disponível em: <https://arxiv.org/abs/1810.03993>.

Na **Constituição Federal** brasileira de 1988, o **artigo 5º** aponta que:



“Art. 5º Todos são iguais perante a lei, sem distinção de qualquer natureza, garantindo-se aos brasileiros e aos estrangeiros residentes no País a inviolabilidade do direito à vida, à liberdade, à igualdade, à segurança e à propriedade [...]”.



Diante disso, observa-se que o Brasil carece de normas legais ou regulamentares como instrumentos apropriados para autorizar o uso de FRTs para Segurança Pública (...)



Isso nos permite conceber que esses direitos são a base do bem viver na sociedade brasileira. A inviolabilidade dos mesmos é a garantia de que a relação entre o indivíduo e o Estado se mantém intacta, juntamente com o Estado Democrático de Direito.

Além disso, a Constituição estabelece o princípio da legalidade como garantia fundamental dos administrados em face do Estado, nos termos do inciso II, art. 5º, “*ninguém será obrigado a fazer ou deixar de fazer alguma coisa senão em virtude de lei*” e, ainda, o *caput* do art. 37, que dentre outros princípios a serem observados pela administração pública direta e indireta de qualquer dos Poderes da União, dos Estados, do Distrito Federal e dos Municípios obedecerá ao princípio da legalidade e princípio da publicidade.

E mais, pela interpretação do art. 59 da Constituição Federal¹³, é possível concluir que o sistema de Direito positivo brasileiro não outorga função legislativa destinada à normas legais restritivas ao direito fundamental à proteção dos dados pessoais pelo Poder Executivo.

Diante disso, observa-se que *o Brasil carece de normas legais ou regulamentares como instrumentos apropriados para autorizar o uso de FRTs para Segurança Pública*, o que vai reforçado pelo disposto no artigo 4º, III, *a*, da LGPD,¹⁴ que excetua a aplicação dessa lei para tal finalidade.

13 Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm.

14 Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm.

Acreditamos que não seja controverso dizer que a implantação de FRTs pelo Poder Público interfere na privacidade de qualquer pessoa. Pois, seus rostos (*dados pessoais sensíveis*, inciso II, do art. 5º, da LGPD)¹⁵ serão filmados e escaneados. Tal coleta ocorre sem o consentimento e sem o seu conhecimento, se não houver notificação para tanto. Isso atrairia o disposto no art. 7º, I, da LGPD (i.e., *o consentimento do titular de dados*), ou mesmo, poderia haver sua dispensa em caso de cumprimento de obrigação legal ou regulatória pelo controlador (art. 11, II, a, da LGPD). Ambas situações para qual não existe nenhum amparo legal na utilização de FRTs.

“

Arguimos que, sem dúvida, o uso de FRTs para fins de vigilância e Segurança Pública devem ser categorizados como “alto risco”.

”

Vale também recordar que a Autoridade Nacional de Proteção de Dados (ANPD), ao publicar a Resolução CD/ANPD nº 2 de 2022,¹⁶ flexibiliza a aplicação da LGPD para os “*agentes de tratamento de pequeno porte*” (*algo que certamente não se aplica ao Estado*). Ao mesmo tempo, estabelece os requisitos da identificação do tratamento de dados de alto risco, em especial que *possa afetar significativamente interesses e direitos fundamentais dos titulares*, (art. 4º, I e II).

Arguimos que, sem dúvida, o uso de FRTs para fins de vigilância e Segurança Pública devem ser categorizados como “alto risco”.

Além disso, o tratamento de dados pessoais, ainda mais quando em aplicações de alto risco, de acordo com o art. 5º, XVII, da LGPD, existe a obrigação do controlador, neste caso, o Poder Público, em realizar o *Relatório de Impacto à Proteção de Dados* (RIPD), conhecido também como *Data Protection Impact Assessment* (DPIA). Este documento conterá a descrição dos processos de tratamento de dados pessoais que podem gerar riscos às liberdades civis e aos direitos fundamentais, bem como medidas, salvaguardas e mecanismos de redução de risco.¹⁷

15 Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm.

16 Disponível em: <https://www.in.gov.br/en/web/dou/-/resolucao-cd/anpd-n-2-de-27-de-janeiro-de-2022-376562019>.

17 Disponível em: https://www.gov.br/governodigital/pt-br/seguranca-e-protecao-de-dados/guias/guia_template_ripd.docx.

Aliás, a ANPD poderá determinar ao controlador, nos termos do art. 38 da LGPD:



[...] “elabore relatório de impacto à proteção de dados pessoais, inclusive de dados sensíveis, referente a suas operações de tratamento de dados, nos termos de regulamento, observados os segredos comercial e industrial.”

O parágrafo único do referido artigo descreve o detalhamento que deverá constar no RPID, como a *descrição dos tipos de dados coletados, a metodologia utilizada para a coleta e para a garantia da segurança das informações e a análise do controlador com relação a medidas, salvaguardas e mecanismos de mitigação de risco adotados*.¹⁸

Tal medida serve para avaliar a sua necessidade/proporcionalidade e os respectivos riscos aos direitos e liberdades dos titulares de dados, bem como comprovar a conformidade com eventual legislação e aferir a inocorrência de viés racial e de gênero em sua tecnologia.

Logo, *se faz necessária a elaboração do RIPD como parte do desenvolvimento e implantação de FRTs*, acompanhado de uma avaliação de riscos quanto aos direitos fundamentais. *Tais informações devem ser disponibilizadas de forma clara, adequada e ostensiva em domínio público como requisito de transparência e responsabilidade*, inclusive, para que os administrados possam ter acesso facilitado às informações sobre o tratamento de seus dados pessoais, como previsto nos arts. 9º e 18, ambos da LGPD.

Vale ressaltar que até o momento, tanto o Estado do Rio Grande do Sul, como a Prefeitura de Porto Alegre, não disponibilizaram o RIPD para a utilização de FRT.

¹⁸ Um modelo de RIPD pode ser adquirido, pelo site do Governo, através deste link: https://www.gov.br/governodigital/pt-br/seguranca-e-protecao-de-dados/guias/guia_template_ripd.docx.

Diante dessa instabilidade jurídica, até que haja a regulação com as salvaguardas necessárias, se mostra ***oportuno suspender o uso de tecnologias de reconhecimento facial pelo Poder Público no Estado do Rio Grande do Sul e na Prefeitura de Porto Alegre.***

Para além da avaliação jurídica, é também necessário que algumas considerações técnicas sobre a natureza de FRTs sejam expostas à luz do debate.

Por exemplo, é importante ressaltar que FRTs (*quando baseadas em Aprendizagem de Máquina e Redes Neurais Profundas*) possui uma série de “vulnerabilidades” associadas a sua utilização. Por exemplo:



Redes neurais profundas são algoritmos que por natureza possuem um alto nível de opacidade¹⁹, tornando assim a interpretação de suas classificações um desafio ainda a ser conquistado pela comunidade. Em situações de alto risco, indivíduos têm o direito de uma explicação para, por exemplo, caso venham a ter seu direito de ir e vir limitado.



Modelos de aprendizagem de máquina são suscetíveis a desenvolverem regras que podem conter vieses discriminatórios.²⁰ Se vieses históricos se encontrarem presentes nos dados utilizados para treinar um modelo (e nenhuma medida de prevenção for tomada), tal modelo irá perpetuar tal estereótipo em suas classificações.

19 Suas classificações/decisões são difíceis de serem explicadas/entendidas em uma linguagem nativa/amigável ao ser humano.

20 Por exemplo, um classificador facial que performa de forma insatisfatória para indivíduos de certos tipos de tons de pele.



Por fim, modelos de aprendizagem de máquina *são vulneráveis a diversos tipos de ataque*²¹, tornando sua performance em aplicações do mundo real, no mínimo, sujeita a exploração adversarial. Dada a forma como tais modelos são treinados, as mesmas técnicas (*otimização via gradiente descendente/ascendente*) podem ser utilizadas para criar amostras capazes de enganar praticamente qualquer classificador.²²

Além disso, vale ressaltar que no quesito “**Ética no Desenvolvimento de Software**”:

Desenvolvedores ainda não recebem apoio e incentivo para trabalhar com questões éticas.

Transparência algorítmica e responsabilização jurídica ainda são problemas em aberto do campo.

Avaliação e mitigação de vieses discriminatórios ainda é algo pouco trabalhado em ambientes empresariais.

Ainda assim, isso não impede que tais sistemas sejam amplamente utilizados pelo poder público. No Brasil, 23 ministérios e agências federais já utilizam sistemas dotados de alguma forma de inteligência artificial (por exemplo, Aprendizagem de Máquina) para automatizar certas funções. Por exemplo:²³

21 Ataques de evasão, envenenamento de dados, extração de modelo, entre outros.

22 Para aprender mais sobre ataques adversariais, nosso repositório de código contém vários exemplos de como tais ataques podem ser realizados. Disponível em: https://github.com/Nkluge-correa/teeny-tiny_castle.

23 Transparência Brasil. (2021). Recomendações de governança - Uso de Inteligência Artificial pelas Autoridades Públicas. https://www.transparencia.org.br/downloads/publicacoes/Recomendacoes_Governanca_Uso_IA_PoderPublico.pdf.

- ▲ O Banco do Brasil utiliza modelos de linguagem para atendimento ao cliente, e modelos de visão computacional para reconhecimento facial;
- ▲ O Departamento da Polícia Federal utiliza IA para reconhecimento facial e modelos de linguagem para detecção de fraude;
- ▲ O Superior Tribunal de Justiça utiliza IA para realizar varreduras automáticas para cada recurso apresentado ao tribunal, cruzar recursos com decisões anteriores (busca de precedentes legais), e gerar recomendações para os juízes.

“

(...) não há como garantir que não haverá nenhum tipo de erro por parte de FRTs.

”

É importante ressaltar que muitos desses modelos (*possivelmente*) sofrem das mesmas vulnerabilidades técnicas mencionadas anteriormente. No mais, o ponto principal que queremos ressaltar aqui é que *não há como garantir que não haverá nenhum tipo de erro por parte de FRTs*.

Para validar tal afirmação, citamos o própria *Defensoria Pública do Estado do Rio de Janeiro*²⁴, onde entre 1º de junho de 2019 e 10 de março de 2020, houve pelo menos 58 casos de reconhecimento facial equivocado, resultando em acusações injustas e até mesmo a prisão de indivíduos inocentes. Vale a pena ressaltar que de todos os acusados injustamente, 70% eram pretos.²⁵ Algo que expõe o problema de “*modelos reproduzirem vieses históricos presentes no corpo de dados utilizados para treiná-los*.”

Embora que alguns sistemas de reconhecimento facial tenham uma precisão de classificação notável (mais de 90%), esses resultados não são universais. Um corpus crescente de estudos

24 Disponível em: <http://www.defensoria.rj.def.br/uploads/imagens/d12a8206c9044a3e92716341a-99b2f6f.pdf>.

25 The Intercept Brasil. EXCLUSIVO: LEVANTAMENTO REVELA QUE 90,5% DOS PRESOS POR MONITORAMENTO FACIAL NO BRASIL SÃO NEGROS. Disponível em: <https://theintercept.com/2019/11/21/presos-monitoramento-facial-brasil-negros/>.

revela disparidades nas taxas de erro entre as categorias demográficas, com indivíduos mulheres, pretas e de 18 a 30 anos apresentando consistentemente a menor precisão.

Para aterrar esta proposta, citamos aqui um estudo onde três algoritmos de classificação de gênero, incluindo aqueles produzidos pela IBM e pela Microsoft, foram avaliados na pesquisa seminal “*Gender Shades*”.²⁶ Mulheres de pele mais escura, homens de pele mais escura, mulheres de pele mais clara e homens de pele mais clara foram divididos em quatro grupos. No subgrupo composto por mulheres de pele mais escura, *todos os três algoritmos tiveram o pior desempenho, com taxas de erro até 34% maiores do que em homens de pele mais clara.*

O Instituto Nacional de Padrões e Tecnologia dos Estados Unidos da América (NIST) corroborou independentemente essas descobertas, indicando que os sistemas de reconhecimento facial em 189 algoritmos são os menos confiáveis em mulheres pretas.²⁷

Assim, seria possível que a implementação de FRTs para fins de vigilância possam impactar de forma desproporcional a comunidade preta? Acreditamos que sim. Mas indiferente de nossa crença, *isso não muda o fato de que avaliações para garantir que algoritmos não possuem vieses discriminatórios em situação de alto risco devem ser feitas*, e seus resultados publicados para toda a população.

Ao mesmo tempo, a vigilância facial, de acordo com a *Algorithmic Justice League*,²⁸ “*põe em perigo direitos como privacidade, liberdade de expressão, liberdade de associação e devido processo*”, independente da cor da nossa pele. Tal vigilância está associada a mudanças comportamentais, como autocensura por medo de retaliação. Lembrando que, assim como o FBI nos Estados Unidos da América, a *Inteligência Brasileira tem uma longa história de monitoramento e repressão de ativistas e líderes pretos.*

26 Disponível em: <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>.

27 Disponível em: <https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf>.

28 Disponível em: <https://www.ajl.org/facial-recognition-technology>.

Além disso, a vigilância contínua instila ansiedade e trauma psicológico nas pessoas, tornando-as vulneráveis a abusos direcionados. No início deste ano, o ator americano Michael B. Jordan *apareceu como suspeito em uma das três imagens disponibilizadas no Termo de Reconhecimento Fotográfico da Polícia Civil do Ceará*. Como resultado, *um adolescente de 17 anos foi apreendido como suspeito de um massacre*.²⁹ Algo que, mais uma vez, demonstra como a cega confiança em FRTs pode causar na violação de direitos fundamentais.

A mobilização **#SaiDaMinhaCara**, é uma reação ao uso de FRTs na segurança pública, após falhas principalmente na identificação de pessoas pretas. Um movimento que, além do Rio Grande do Sul, acontece também no Ceará, Bahia, Rio de Janeiro, Minas Gerais, São Paulo, Rio Grande do Norte, Espírito Santo, Santa Catarina, Pernambuco, Sergipe, Pará, Amapá, Alagoas, Mato Grosso e no Distrito Federal, contando com os esforços de 50 parlamentares.³⁰

Finalizando, gostaríamos de deixar claro que não nos posicionamos contra o uso da tecnologia para fins de otimização e melhoria do bem estar-coletivo de nossa população. Contudo, defendemos que tais melhorias devem ser feitas de forma parcimoniosa, atendendo a regulamentação vigente, com o devido cuidado para que os possíveis efeitos colaterais sejam mapeados (com antecedência) e devidamente mitigados, e por fim, de forma transparente e acessível.

Assim, acreditamos que estaremos rumando para um futuro onde a tecnologia será usada como uma ferramenta de proteção e emancipação humana, e não mais um aparato de automação a opressão de populações vulneráveis.

29 Disponível em: <https://g1.globo.com/ce/ceara/noticia/2022/01/07/astro-do-cinema-michael-b-jordan-aparece-em-lista-de-procurados-pela-policia-do-ceara.ghtml>.

30 Disponível em: <https://sul21.com.br/noticias/politica/2022/06/vereadores-apresentam-projeto-de-lei-contra-uso-de-reconhecimento-facial-em-porto-alegre/>.

AIRES na PUCRS

Presidente

Nicholas Kluge Corrêa

Vice-Presidentes

Diogo Fernando Massmann
Lara Agustina Sosa Marquez

Diretora de Marketing

Carolina Del Pino Carvalho

Diretor Financeiro

Guilherme Camargo

Secretária

Camila Trindade Galvão

Secretário

Rodrigo Mambrini S. Barbosa

RAIES

Coordenador do Projeto RAIES/PUCRS

Nythamar Fernandes de Oliveira

Vice Coordenador do Projeto RAIES/PUCRS

Paulo Caliendo

Jair Tauchen

Henrique Carlos de Castro

Rodrigo Leal

Natália Marroni Borges



AIRES



RAIES

