



RAFAEL LUCAS BORBA

**ABORDAGENS REGULATÓRIAS EXISTENTES PARA A MINIMIZAÇÃO
DOS VIESES DISCRIMINATÓRIOS NA TECNOLOGIA DE
RECONHECIMENTO FACIAL USADA POR AUTORIDADES POLICIAIS**

Dissertação com vista à obtenção do grau
de Mestre em Direito, na especialidade de
Direito Empresarial e Tecnologia.

Orientador:
Doutor Eduardo José Guedes Magrani

Maio 2022



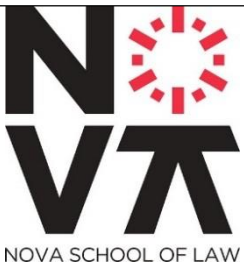
RAFAEL LUCAS BORBA

**ABORDAGENS REGULATÓRIAS EXISTENTES PARA A MINIMIZAÇÃO
DOS VIESES DISCRIMINATÓRIOS NA TECNOLOGIA DE
RECONHECIMENTO FACIAL USADA POR AUTORIDADES POLICIAIS**

Dissertação com vista à obtenção do grau
de Mestre em Direito, na especialidade de
Direito Empresarial e Tecnologia.

Orientador:
Doutor Eduardo José Guedes Magrani

Maio 2022



**ABORDAGENS REGULATÓRIAS EXISTENTES PARA A MINIMIZAÇÃO DOS
VIESES DISCRIMINATÓRIOS NA TECNOLOGIA DE RECONHECIMENTO
FACIAL USADA POR AUTORIDADES POLICIAIS**

DECLARAÇÃO ANTIPLÁGIO

Declaro por minha honra que o trabalho que apresento é original e que todas as minhas citações estão corretamente identificadas. Tenho consciência de que a utilização de elementos alheios não identificados constitui uma grave falta ética e disciplinar.

Lisboa, 26 de maio de 2022.



Rafael Lucas Borba

AGRADECIMENTOS

Ao meu orientador, Dr. Eduardo Magrani, pela generosidade e acolhimento, para além dos conhecimentos compartilhados e incentivos. Agradeço por toda a liberdade, segurança e tranquilidade durante o processo de escrita da dissertação. Mais do que um orientador, ganhei um amigo. Obrigado!

À minha família, em especial aos meus pais, Ruymar Marinho Borba e Maria Célia Lucas Borba, pelo amor e incentivo incondicionais. Mesmo à distância e com muita saudade, vocês sempre me deram muita força. Obrigado!

À minha esposa, Julyana Cerqueira Buery, por todo o amor, carinho e compreensão. Sem você, nada disso teria acontecido. Obrigado por acreditar em mim e me acompanhar nessa aventura. E obrigado por nunca me deixar desistir. Só você sabe quão desafiadora foi essa etapa. Obrigado!

A DEUS, por me guiar em todos os momentos.

DA FORMA DE CITAR E OUTRAS CONVENÇÕES

No corpo do texto e nas notas de rodapé, as citações foram apresentadas sob a forma de referências entre parênteses com os seguintes elementos: apelido do autor ou nome da entidade responsável pela obra; ano de publicação; e páginas. Quando o nome do autor ou da entidade fazem parte do texto, constam entre parênteses apenas o ano e os números das páginas. Serão acrescentados quaisquer outros elementos que se julguem convenientes.

As citações diretas e indiretas de obras estrangeiras incluem traduções realizadas pelo autor para a língua portuguesa, assumindo o autor a responsabilidade por essas.

A utilização de itálico servirá para destacar determinado conceito, terminologia específica, ou estrangeirismos.

A presente dissertação respeita o Novo Acordo Ortográfico da Língua Portuguesa, aprovado pela Resolução da Assembleia da República nº 35/2008, de 29 de julho. Em razão de seu autor ser brasileiro, as variantes utilizadas no Brasil foram as eleitas para esta dissertação. Isso, contudo, não se reflete nas citações diretas de legislação, autores ou obras editadas em Portugal, as quais não se alterou para manter a integralidade.

LISTA DE ABREVIATURAS

ACLU	American Civil Liberties Union
BKA	Bundeskriminalamt
CFTV	Circuito Fechado de Televisão
CNN	Convolutional Neural Network
CNIL	Commission Nationale de l'Informatique et des Libertés
CoE	Conselho da Europa
COVID-19	Coronavirus Disease 2019
DEP	Digital Europe Programme
DNN	Deep Neural Network
DPA 2018	Data Protection Act 2018
DPIA	Data Protection Impact Assessment
EA 2010	Equality Act 2010
ECC	European Cybercrime Center
ECTC	European Counter Terrorism Center
EDPB	European Data Protection Board
EDPS	European Data Protection Supervisor

EDRi	European Digital Rights
EDS	European Data Strategy
EFA	European Free Alliance
EUA	Estados Unidos da América
EUROPOL	Agência da União Europeia para a Cooperação Policial
FRA	European Union Agency for Fundamental Rights
FRIA	Fundamental Rights Impact Assessment
FVRT	Face Recognition Vendor Test
GDPR	General Data Protection Regulation
HLEGAi	High Level Expert Group on Artificial Intelligence
IA	Inteligência Artificial
IBM	International Business Machines Corporation
IDEC	Instituto Brasileiro de Defesa do Consumidor
IFF	Internet Freedom Foundation
INTERPOL	Organização Internacional de Polícia Criminal
LED	Law Enforcement Directive
ONG	Organização Não Governamental

PPB	Pilot Parliaments Benchmark
RBI	Remote Biometric Identification
TELEFI	Towards the European Level Exchange of Facial Images
TJUE	Tribunal de Justiça da União Europeia
TRF	Tecnologia de Reconhecimento Facial
UE	União Europeia

CONTAGEM DE CARACTERES

O corpo desta dissertação, incluindo espaços e notas, ocupa um total de 196.124 caracteres.

Lisboa, 26 de maio de 2022.

A handwritten signature in black ink, reading "Rafael Lucas Borba". The script is cursive and fluid, with the first letters of each name being capitalized and prominent.

Rafael Lucas Borba

RESUMO

O uso de tecnologia de reconhecimento facial (TRF) por autoridades policiais para fins de manutenção da ordem pública vem gerando preocupação devido aos vieses de gênero e raça presentes na tecnologia, que podem levar à identificação incorreta de pessoas de determinados grupos demográficos, violando seus direitos fundamentais. Estudiosos têm destacado que a legislação antidiscriminação da UE é insuficiente para combater especificamente a questão dos vieses discriminatórios na TRF usada por autoridades policiais. Além disso, sugerem duas abordagens principais para lidar com esse problema, quais sejam, a abordagem apresentada pela Proposta de Regulação de IA da Comissão Europeia, e a abordagem apresentada pelo princípio de não discriminação *by design*. Diante disso, o presente trabalho se debruçou sobre essas duas opções regulatórias, para definir qual delas seria a melhor abordagem para concretizar o direito fundamental de não discriminação na TRF usada por autoridades policiais. Após ter sido feita uma análise comparativa entre as duas abordagens, destacando de forma crítica os pontos comuns e divergentes de cada abordagem, chegou-se à conclusão de que as duas se complementam e, por isso mesmo, é oportuno que se desenvolva um sistema regulatório no qual ambas estejam presentes, a fim de mitigar as lacunas existentes na legislação europeia antidiscriminação.

Palavras-chave: Tecnologia de reconhecimento facial. Autoridades Policiais. Vieses discriminatórios. Proposta de Regulamento de IA da Comissão Europeia. Princípio da não discriminação *by design*.

ABSTRACT

The use of facial recognition technology (FRT) by law enforcement authorities for the purposes of maintaining public order has been generating concern due to gender and race biases present in the technology, which may lead to the incorrect identification of people from certain demographic groups, violating their fundamental rights. Scholars have highlighted that EU anti-discrimination legislation is insufficient to specifically tackle the issue of discriminatory biases in FRT used by law enforcement authorities. Moreover, they suggest two main approaches to deal with this problem, namely, the approach put forward by the European Commission's AI Regulation Proposal, and the approach put forward by the principle of non-discrimination by design. In view of this, the present paper looked into these two regulatory options, to define which one would be the best approach to realise the fundamental right of non-discrimination in FRT used by law enforcement authorities. After a comparative analysis between the two approaches, critically highlighting the common and divergent points of each, the conclusion was reached that both approaches complement each other and, therefore, it is beneficial to develop a regulatory system in which both are present, mitigating the existing gaps in the European anti-discrimination legislation.

Keywords: Facial recognition technology. Police Authorities. Discriminatory biases. European Commission Proposal for an AI Regulation. Principle of non-discrimination *by design*.

SUMÁRIO

INTRODUÇÃO	1
1. A TECNOLOGIA DE RECONHECIMENTO FACIAL (TRF)	4
1.1. INTELIGÊNCIA ARTIFICIAL, <i>MACHINE LEARNING</i> E <i>DEEP LEARNING</i> ..	4
1.2. FUNCIONAMENTO DAS TECNOLOGIAS DE RECONHECIMENTO FACIAL	9
1.3. OS USOS DA TRF PELAS AUTORIDADES POLICIAIS	13
1.4. DESAFIOS ATINENTES À TRF	21
1.4.1. FALTA DE PRECISÃO (ACURÁCIA).....	21
1.4.2. ENVIESAMENTO.....	22
1.4.2.1. Causas do enviesamento	26
1.4.2.2. Casos de discriminação de gênero na TRF.....	29
1.4.2.3. Casos de discriminação de raça na TRF	30
2. A ATUAL LEGISLAÇÃO EUROPEIA DE NÃO DISCRIMINAÇÃO	32
2.1. AS PRINCIPAIS POSIÇÕES EXISTENTES ATUALMENTE SOBRE O USO DE TRF.....	32
2.2. A ATUAL LEGISLAÇÃO ANTIDISCRIMINAÇÃO DA UE E SUA APLICAÇÃO À TRF	36
2.3. TENSÕES E LACUNAS NA LEGISLAÇÃO ANTIDISCRIMINAÇÃO DA UE RELACIONADAS AO USO DE TRF	45
3. CAMINHOS REGULATÓRIOS PARA A MINIMIZAÇÃO DAS TENSÕES E LACUNAS EXISTENTES NA LEGISLAÇÃO ANTIDISCRIMINAÇÃO DA UE: <i>AI ACT</i> E PRINCÍPIO DA NÃO DISCRIMINAÇÃO <i>BY DESIGN</i>	48
3.1. A PROPOSTA DE REGULAMENTO DE IA DA COMISSÃO EUROPEIA E A SUA ABORDAGEM EM RELAÇÃO AO USO DE TRF POR AUTORIDADES POLICIAIS	49
3.2. PRINCÍPIO DA NÃO DISCRIMINAÇÃO <i>BY DESIGN</i>	55
3.3. ANÁLISE COMPARATIVA ENTRE AS ABORDAGENS REGULATÓRIAS DA PROPOSTA DE REGULAMENTO DE IA DA COMISSÃO EUROPEIA E O PRINCÍPIO DE NÃO DISCRIMINAÇÃO <i>BY DESIGN</i>	64
CONCLUSÃO.....	70
REFERÊNCIAS	75

INTRODUÇÃO

O debate a respeito das tecnologias de reconhecimento facial (TRF) entrou em ebulição nos últimos anos. O desenvolvimento exponencial (e muitas vezes irresponsável) da tecnologia tem ampliado a sua utilização a campos antes inimagináveis e, ao mesmo tempo, tem gerado o alarme de vozes críticas que se preocupam (e com razão) com os efeitos prejudiciais da proliferação de tal tecnologia sobre indivíduos e grupos demográficos específicos.

Um dos usos da TRF que mais vem ganhando destaque, e gerando preocupação, é o realizado pelas autoridades policiais para fins de manutenção da ordem pública, seja em tempo real ou em tempo diferido (para investigações criminais). Por um lado, estudiosos e entidades, críticas a esse uso, alegam que vieses de gênero e de raça presentes na tecnologia podem levar as autoridades policiais a identificarem incorretamente pessoas de determinados grupos demográficos e, com isso, violarem os direitos fundamentais (como o de não discriminação) de tais pessoas, tornando a tecnologia num instrumento de perpetuação de preconceitos e opressões históricas. Em outras palavras, ao invés de se configurar num meio de manutenção da ordem pública, a TRF nada mais seria do que um meio de manutenção do *status quo*, que aprofunda e reproduz em larga escala perseguições a grupos demográficos já historicamente marginalizados.

Estudos como o de Buolamwini e Gebru (2018, p.77) têm demonstrado que a TRF apresenta taxas de erro mais altas em relação a certos grupos demográficos como, por exemplo, pessoas negras e mulheres (em especial, mulheres negras), o que levaria a um impacto discriminatório sobre essas pessoas. Imagine-se, por exemplo, que em decorrência de um falso positivo da TRF, uma pessoa inocente que não tenha cometido qualquer crime, seja injustamente investigada, detida e até mesmo sentenciada à prisão efetiva. Tais injustiças podem impactar profundamente a vida de pessoas que já se encontram marginalizadas. Primeiro, porque podem privar a liberdade de tais pessoas. Segundo, porque podem causar efeitos secundários, tais como, por exemplo, aumento de preconceitos, negação de oportunidades de emprego, negação de crédito, vigilância indevida.

Por sua vez, os defensores do uso da TRF por autoridades policiais enfatizam que tal tecnologia pode trazer benefícios como, por exemplo, o maior controle de fronteiras, maior eficiência na busca de pessoas desaparecidas (em especial, crianças) e na busca de suspeitos de crimes, contribuindo, assim, para a manutenção da ordem pública e a

disseminação da sensação de segurança no seio da população. Qualquer que seja o lado escolhido, parece que a TRF veio para ficar. Quer seja para fins comerciais, quer seja para fins de manutenção da ordem pública, a presença cada vez maior de sistemas de reconhecimento facial é inegável. Para o bem ou para o mal, tais tecnologias têm moldado cada vez mais as práticas e normas da vida cotidiana e se tornado uma parte cada vez mais integrada do mundo social conectado (Leslie, 2020, p.4-8).

Nota-se que, apesar dos diferentes receios e posicionamentos existentes nas instituições e na sociedade civil, é possível perceber que o uso da TRF na UE avança a passos largos. Por exemplo, em outubro de 2021, o grupo parlamentar “*Greens/EFA (European Free Alliance)*”, publicou um estudo encomendado, no qual buscou mapear os usos atuais da TRF especificamente na UE. Intitulado “*Biometric & Behavioural Mass Surveillance in EU Member States*”, o estudo revelou que autoridades policiais de 11 países membros da UE estavam usando TRF “em tempo diferido” frente a bancos de dados biométricos para fins de investigação criminal, enquanto outros 7 países deveriam seguir esse caminho em breve.

Isso tudo demonstra que a UE precisará realizar rapidamente uma abordagem regulatória harmonizada em relação à TRF usada por autoridades policiais, se quiser se desenvolver tecnologicamente sem comprometer os seus valores primordiais e direitos fundamentais (como a não discriminação). Caso não o faça, a UE arrisca permitir a proliferação de legislações nacionais diferentes e contraditórias, que por sua vez poderão impedir a livre circulação de bens e serviços relacionados à TRF; bem como impedir o estabelecimento de um elevado nível de proteção dos direitos fundamentais comum a todos os países membros da UE, no que diz respeito ao uso da TRF por autoridades policiais.

Diante disso, foi feita uma pesquisa preliminar que constatou dois pontos principais na literatura existente sobre o tema: primeiro, que a legislação antidiscriminação da UE é insuficiente para combater especificamente a questão dos vieses discriminatórios na TRF usada por autoridades policiais; segundo, a sugestão de duas abordagens principais para lidar com esse problema, quais sejam, a abordagem apresentada pela Proposta de Regulação de IA da Comissão Europeia, e a abordagem apresentada por um princípio de não discriminação *by design*.

De posse de tais informações, o presente trabalho vai se debruçar sobre essas duas opções regulatórias (que parecem mais interessantes e eficazes), a fim de definir qual

delas seria a melhor abordagem para concretizar o direito fundamental de não discriminação na TRF usada por autoridades policiais.

Com o objetivo de responder a essa pergunta, foram feitas pesquisas na atual doutrina, legislação e jurisprudência da UE sobre o tema, bem como revisados os estudos publicados por órgãos oficiais. Dentre as principais normas analisadas, destacam-se: a legislação secundária antidiscriminação da UE; o Regulamento Geral de Proteção de Dados; a Diretiva UE 2016/80 (*Law Enforcement Directive*); a Convenção Europeia dos Direitos do Homem; a Carta de Direitos Fundamentais da UE; Convenção 108+; e a Proposta de Regulação de IA da UE (que ainda não está em vigor).

Sendo assim, com base na bibliografia e documentos levantados, será feita, num primeiro momento, uma análise para entender como funciona a TRF, quais são os seus principais usos pelas autoridades policiais e quais são os seus principais desafios, dando especial ênfase à análise dos vieses de gênero e de raça.

Mapeadas tais questões, será feita uma análise da atual legislação antidiscriminação da UE a fim de conferir os seguintes pontos: quais as principais posições existentes dentro da UE atualmente sobre o uso de TRF; quais os tipos de discriminação existentes; se tal legislação realmente é insuficiente para lidar com sistemas de IA e suas formas sutis e não intuitivas de discriminação algorítmica; quais as tensões e lacunas na legislação antidiscriminação da UE relacionadas ao uso de TRF por autoridades policiais.

Por fim, ciente das lacunas e tensões existentes, o último capítulo adentrará na análise das duas abordagens regulatórias específicas aqui especuladas, a fim de descobrir qual delas é a melhor opção para minimizar as tensões e lacunas existentes na legislação antidiscriminação da UE, e resolver de forma eficaz o problema do enviesamento no uso de TRF por autoridades policiais. Primeiramente, será feita uma análise individualizada, tanto da Proposta da Comissão, quanto do princípio de não discriminação *by design*, abordando as vantagens e desvantagens de cada um. Em seguida, finalizando o trabalho, será feita uma análise comparativa entre as duas abordagens, destacando de forma crítica os pontos comuns e divergentes de cada abordagem: a da Proposta (mais ampla e baseada no risco), e a do princípio (mais específica e direcionada).

1. A TECNOLOGIA DE RECONHECIMENTO FACIAL (TRF)

1.1. INTELIGÊNCIA ARTIFICIAL, *MACHINE LEARNING* E *DEEP LEARNING*

A IA, em si, não significa muita coisa palpável, pois, para fazer sentido, ela depende de uma aplicação específica. Por esse motivo é que vários planos estratégicos de IA começam com toda a problemática do conceito. De uma forma geral, é possível dizer que a IA, em princípio, com a ideia de emular o pensamento humano, é um conceito “guarda-chuva”, que abrange muitos elementos. É importante entendermos do que estamos tratando, pois, a depender dos conceitos adotados, podemos gerar análises técnicas e jurídicas completamente diferentes.

Sendo a IA uma área de conhecimento abrangente, seu nascimento teve paralelo com a própria ciência da computação. Por isso, a IA é entendida como um subcampo da informática. Somente a partir da década de 1950 é que o seu conceito começou a tomar uma forma mais próxima do que conhecemos hoje, sendo possível destacar dois marcos fundamentais que orientaram os estudos: O Teste de Turing e o conceito proposto por John McCarthy (Magrani *et al.*, 2021a, p. 11).

Em 1950, Alan Turing publicou um artigo seminal, chamado *Computing Machinery and Intelligence*, que foi o primeiro texto sólido quando se fala na criação de uma área nova baseada na emulação computável do pensamento humano. Em seu texto, Turing propõe um teste/jogo, conhecido como Teste de Turing (ou jogo da imitação), que se baseia na seguinte ideia: a uma pessoa isolada (C), são transmitidas respostas às suas perguntas. Essa pessoa precisa ser capaz de discernir se a resposta que ela está recebendo vem de uma pessoa (B) ou de um dispositivo (A). Se a pessoa (C) não conseguir fazer a diferenciação sobre se as respostas que está recebendo são respostas de um ser humano ou respostas de um computador, significa que o computador (A) passa no Teste de Turing (Turing, 1950, p. 433-460). Ainda hoje, este é um teste que inspira e serve de diretriz de desenvolvimento para alguns modelos de automação encontrados no nosso dia a dia.

No verão de 1956, ocorreu um estudo em *Dartmouth College*, em *Hanover, New Hampshire*, que durou cerca de dois meses e reuniu aqueles que seriam os maiores cientistas ligados à pesquisa de IA na época. Durante o período, foram discutidas questões ligadas aos limites e capacidades das máquinas e o potencial que estas teriam para realizar tarefas tão bem quanto os humanos. O intuito deste grupo de cientistas era de que em apenas 2 meses poderiam fazer um avanço significativo em um ou em mais dos problemas

debatidos (McCarthy *et al.*, 2006, p. 12-14). Aqui, é preciso fazer um parêntese, pois é interessante notar que este primeiro grupo de cientistas que se propôs a estudar IA era composto apenas por homens brancos, com o mesmo tipo de formação acadêmica, que falavam uma mesma língua, e com a mesma visão de mundo. Tais características, por si só, demonstram o enviesamento que permeou os estudos e desenvolvimento da IA desde os seus primórdios.

Apesar disso, foi justamente durante este estudo que John McCarthy cunhou o termo Inteligência Artificial. Segundo o autor, uma máquina poderia ser considerada IA se fosse capaz de realizar uma tarefa que normalmente ligamos à inteligência humana. Importa apenas ser capaz de realizar a tarefa, não a forma como a realiza. (McCarthy, 2000, p. 2-5). Por sua vez, outros estudiosos entendem que o que importa é apenas que a tarefa seja realizada, não havendo necessidade de se balizar no pensamento humano. Outros, por fim, entendem por um conceito meio-termo, que se inspira no raciocínio lógico-racional humano, mas não se limita à ideia de imitar o ser humano (Magrani, 2019, p. 51).

Como é possível perceber, não existe consenso sobre o conceito de IA. Apesar disso, com base nos pontos comuns presentes nos diferentes conceitos, é possível afirmar que a IA é um sistema capaz de aprender e executar tarefas que até então somente o ser humano conseguia realizar. E se antes havia a necessidade de que uma pessoa construísse e treinasse o algoritmo, hoje já existem sistemas de IA capazes de aprender sozinhos à medida que vão desempenhando suas atividades (Magrani *et al.*, 2021a, p. 11-12).

A evolução dos sistemas de IA nas últimas décadas deve-se sobretudo a quatro fatores: aprimoramento de modelos teóricos (com métodos estatísticos e probabilísticos mais sofisticados); melhoria da capacidade computacional (que está mais acessível e relativamente mais barata); maior transformação dos ambientes (devido às novas tecnologias de informação); e disponibilidade cada vez maior da quantidade de dados. Tais fatores se retroalimentam e são todos muito importantes (Cath *et al.*, 2018, p. 1-24). No entanto, assume especial relevo o enorme volume de dados que vêm sendo disponibilizados nos últimos anos. São tais dados, conhecidos como *Big Data*, que fornecem os ingredientes necessários para que a IA possa tomar decisões cada vez mais precisas e fundamentadas. Também é devido a esse grande volume de dados que os sistemas de IA mostram-se aptos a desempenhar diferentes funções, tais como, por exemplo, gerar previsões, tomar decisões, classificar dados, e identificar e classificar imagens (Magrani *et al.*, 2021a, p. 12).

Essas aplicações específicas de IA, cada vez mais usadas no nosso dia a dia, são conhecidas como “IA Fraca ou Estreita” (*Narrow AI*). Nada mais são do que os sistemas de IA criados para tarefas específicas e predeterminadas, incapazes de realizar tarefas diferentes daquelas para as quais foram treinados. Por outro lado, existe também o conceito de “IA Forte” (*Strong AI*), ou “IA Geral” (*Artificial General Intelligence*), que seria o sistema de IA realmente capaz de simular o pensamento lógico-racional humano a fim de se adaptar a diferentes contextos e situações que surjam. Devido à sua maior complexidade, ainda não se encontra desenvolvida e demandará maiores cuidados por causa dos riscos que lhe são inerentes (Magrani, 2019, p. 51).

Apesar da importância do conceito de IA apresentado pelo Grupo de Peritos de Alto Nível em Inteligência Artificial da Comissão Europeia¹, a mais importante definição oficial apresentada em âmbito europeu é a presente na Proposta de Regulamento da Inteligência Artificial na União Europeia, que em seu art. 3(1) diz:

Sistema de inteligência artificial (sistema de IA) significa um programa informático desenvolvido com uma ou várias das técnicas e abordagens enumeradas no anexo I, capaz de, tendo em vista um determinado conjunto de objetivos definidos por seres humanos, criar resultados, tais como conteúdos, previsões, recomendações ou decisões, que influenciam os ambientes com os quais interage.

Outro conceito importante é o de aprendizado de máquina (*machine learning*). Uma vez que a IA é uma área de estudo, um termo guarda-chuva, *machine learning* nada mais é do que uma das suas sub-áreas. Nas palavras de Otterlo (2013, p.46):

O aprendizado de máquinas é um ramo da IA que procura desenvolver sistemas de computador que melhoram seu desempenho automaticamente com a experiência, inspirando-se na ciência da computação, estatística, otimização e

¹ "Inteligência artificial (IA) refere-se a sistemas projetados por humanos que, dado um objetivo complexo, agem no mundo físico ou digital percebendo seu ambiente, interpretando os dados estruturados ou não estruturados coletados, raciocinando sobre o conhecimento derivado destes dados e decidindo a (s) melhor (s) ação (ões) a tomar (de acordo com parâmetros pré-definidos) para atingir o objetivo dado. Os sistemas de IA também podem ser projetados para aprender a adaptar seu comportamento, analisando como o ambiente é afetado por suas ações anteriores. Como disciplina científica, a IA inclui várias abordagens e técnicas, tais como aprendizagem de máquinas (das quais aprendizagem profunda e aprendizagem de reforço são exemplos específicos), raciocínio de máquinas (que inclui planejamento, programação, representação e raciocínio do conhecimento, busca e otimização), e robótica (que inclui controle, percepção, sensores e atuadores, bem como a integração de todas as outras técnicas em sistemas ciberfísicos)" (HLEGAI, 2019a, p.7).

psicologia. [...] Usamos o termo geral aprendizado de máquina e o tomamos para significar qualquer metodologia e conjunto de técnicas que encontram novos padrões e conhecimentos em dados, e gera modelos (por exemplo, perfis) que podem ser usados para previsões eficazes sobre os dados.

Por sua vez, Géron (2019, p. 4) define *Machine Learning* como sendo “a ciência (e a arte) da programação de computadores para que eles possam aprender com os dados”. Segundo o autor, é ciência porque envolve a parte computacional e de matemática, mas também é arte porque muitas das coisas feitas dentro de aprendizado de máquina são aleatórias. Os programadores fazem, mas não necessariamente entendem como as coisas funcionam. Eles sabem que as coisas acontecem, têm como mensurar que aquilo aconteceu, mas não têm como explicar exatamente quais foram as regras e o que aconteceu dentro dessa caixa preta. Isso com certeza é um grande problema quando falamos de explicabilidade de algoritmos.

É preciso destacar que o aprendizado de máquina também possui subdivisões como, por exemplo, “*K-means*”, “Árvores de Decisão” e o “Aprendizado Profundo” (*Deep Learning*). Dentre todas as técnicas, o Aprendizado Profundo é a que hoje assume especial relevo, tendo sido impulsionado principalmente a partir de 2006 com o artigo publicado por Geoffrey Hinton e Ruslan Salakhutdinov². Em tal estudo, os autores demonstraram a maneira de treinar uma rede neural profunda de forma a torna-la capaz de reconhecer dígitos manuscritos com uma altíssima precisão (>98%). A partir daí, ficou comprovado que, com a ajuda do alto poder computacional e grandes quantidades de dados, o Aprendizado Profundo era capaz de alcançar feitos impressionantes, impossíveis de serem alcançados por qualquer outra técnica de Aprendizado de Máquina.

Mas o que é exatamente o Aprendizado Profundo? É um subconjunto de Aprendizagem de Máquinas, mais sofisticado, no qual as tarefas são divididas e distribuídas em algoritmos de aprendizado de máquinas que são organizados em camadas consecutivas. A partir do resultado obtido na camada anterior, constrói-se a camada subsequente. Reunidas, as camadas formam uma espécie de rede neural artificial, que se inspira na abordagem distribuída de resolução de problemas realizada pelos neurônios num cérebro humano. Nas palavras de LeCun, Bengio e Hinton (2015, p. 436):

² Ver: <http://www.cs.toronto.edu/~hinton/>.

O aprendizado profundo permite modelos computacionais que são compostos de múltiplas camadas de processamento para aprender representações de dados com múltiplos níveis de abstração. Estes métodos melhoraram drasticamente o estado da arte em reconhecimento de fala, reconhecimento visual de objetos, detecção de objetos e muitos outros domínios, tais como descoberta de medicamentos e genômica. O aprendizado profundo descobre uma estrutura intrincada em grandes conjuntos de dados usando o algoritmo de retropropagação para indicar como uma máquina deve alterar seus parâmetros internos que são usados para calcular a representação em cada camada a partir da representação na camada anterior. As redes profundas convolucionais trouxeram avanços no processamento de imagens, vídeo, fala e áudio, ao passo que as redes recorrentes brilharam em dados seqüenciais, como texto e fala.

Constata-se, portanto, que *Deep Learning* é só mais um algoritmo que faz parte do *Machine Learning* (que por sua vez, faz parte do campo de estudo da IA). E sendo um algoritmo, é só mais um meio de resolver uma tarefa. Não necessariamente é o melhor meio para qualquer tipo de tarefa, tendo em vista que a maioria das tarefas pode ser resolvida com técnicas mais simples como, por exemplo, os métodos “*Esemble*” e “Florestas Aleatórias” (Géron, 2019, p. XVI). Mas com certeza é um meio robusto que resolve tarefas que envolvem os mais diferentes tipos de dados.

Dentre as aplicações de IA que fazem grande uso de *Deep Learning*, podemos destacar a visão computacional. Como o próprio nome indica, é a ideia de se dar “visão” a uma máquina, a fim de que ela seja capaz de enxergar coisas. É usada para processar imagens de fotografias e vídeos, e deles extrair dados e padrões. No entanto, apesar do uso do termo “visão”, é preciso deixar claro que a forma como a máquina detecta imagens é completamente diferente da forma como o olho humano detecta. E é importante termos isso em consideração, uma vez que não é possível imaginar que a solução regulatória para um possa ser igual à do outro. A dinâmica do contexto é completamente diferente. O que o algoritmo da máquina faz é detectar certos pontos de contato na imagem e geometricamente estabelecer que a forma encontrada corresponde à forma de uma coisa “x”, pré-determinada em um banco de dados. No entanto, na falta de um termo computacional próprio para descrever esse processo, preferiu-se fazer uso do antropomorfismo e utilizar o termo “visão computacional”. Dessa forma, conseguia-se expressar a ideia pretendida de forma mais simples e catalisar todo o debate em torno do tema. Tendo isso em mente, é possível destacar o reconhecimento facial como um dos maiores exemplos de aplicação de visão computacional.

1.2. FUNCIONAMENTO DAS TECNOLOGIAS DE RECONHECIMENTO FACIAL

Graças à biometria, é possível usar informação biológica dos indivíduos para identifica-los e verificá-los (Bennett, 2001, p. 151). A biometria fundamenta-se em *“qualquer característica física ou traço pessoal mensurável, robusto e distintivo de um indivíduo, que possa ser usado para identificar ou verificar a identidade reivindicada por ele”* (Woodward *et al.*, 2001, p. 67). Assim como a digitalização da íris e as impressões digitais, o reconhecimento facial é só mais uma técnica biométrica (Bennett, 2001, p. 151).

A importância do reconhecimento facial para as Administrações Públicas pode ser percebida desde meados do século XIX, quando as polícias começaram a armazenar fotos de infratores para a sua posterior identificação. No início da década de 1990, Matthew Turk e Alex Pentland desenvolveram o método das *eigenfaces*, o que foi considerado por muitos o surgimento do reconhecimento facial automatizado propriamente dito, ou seja, operado por um programa informático. Desde então, a tecnologia tem se desenvolvido e se tornado cada vez mais sofisticada (Adler e Schuckers, 2007, p. 1248).

Em linhas gerais, a Tecnologia de Reconhecimento Facial (TRF) pode ser entendida como um sistema biométrico capaz de processar automaticamente imagens digitais que contém rostos de pessoas (ou seja, dados biométricos)³, a fim de identificar, autenticar/verificar ou categorizar essas pessoas (Article 29 Data Protection Working Party, 2012, p.2). Uma vez que a imagem é capturada, seja através de câmeras de vigilância, seja por meio de uma foto publicada em rede social, os dados biométricos faciais são extraídos, processados e podem acabar sendo usados para inúmeros fins. Além disso, o sistema de IA está constantemente aperfeiçoando a sua performance e precisão, pois aprende com os dados que utiliza. Não bastasse, é capaz de padronizar comportamentos considerados de interesse, bem como de *“interpretar e prever ações humanas e a classifica-las como ‘normal’, ‘anormal’ ou ‘danosa/nociva’”* (Fontes e Lütge, 2022, p. 93).

³ Comissão Europeia, Proposta de Regulamento de Inteligência Artificial, 2021/0106 (COD), art. 33 (3): Dados biométricos significam dados pessoais resultantes de um tratamento técnico específico das características físicas, fisiológicas ou comportamentais de uma pessoa singular, os quais permitem obter ou confirmar a identificação única dessa pessoa singular, nomeadamente imagens faciais ou dados dactiloscópicos.

A função de autenticação/verificação procura responder à pergunta "você é quem você diz ser?". Aqui, a biometria é utilizada em uma comparação de um-para-um. Ela procura certificar a identidade de uma pessoa comparando o modelo biométrico que esta pessoa apresenta com um determinado modelo biométrico pré-gravado/armazenado na galeria⁴ (Fieux-Castagnet e Santucci, 2020, p. 5). Este procedimento é utilizado, por exemplo, para controle de fronteiras em aeroportos, em que a imagem do passaporte é comparada com uma imagem ao vivo obtida no local (FRA, 2019b, p. 7).

Já a função de identificação busca responder à pergunta "quem é você?". Aqui, o modelo biométrico facial de uma pessoa é comparado com modelos biométricos faciais de muitas outras pessoas armazenados em um banco de dados/galeria, a fim de descobrir se a imagem dessa pessoa está armazenada ali. Ou seja, é uma comparação de um-para-muitos (FRA, 2019b, p. 7-8). Este procedimento fornece uma "pontuação de similaridade" para cada correspondência feita, a fim de indicar a probabilidade de se ter encontrado o indivíduo. Em alguns países, este procedimento é usado para fornecer funções de segurança como encontrar criminosos, terroristas e pessoas desaparecidas (Fieux-Castagnet e Santucci, 2020, p. 4). Normalmente, os rostos das imagens de vídeo são comparados com os rostos armazenados no banco de dados, a fim de descobrir se a pessoa está em uma lista de vigilância. Tais sistemas geralmente são utilizados em ambientes abertos, sujeitos a interferência externa. Por isso, é maior a probabilidade de que resultem em falsas combinações quando comparados a sistemas de ambiente controlado, como verificação/autenticação. Em geral, o uso da TRF para este tipo de função torna a tecnologia numa poderosa ferramenta de vigilância.

Por fim, a função de categorização busca analisar o modelo biométrico da pessoa a fim de classificá-la de acordo com características como sexo, idade e etnia (Article 29 Data Protection Working Party, 2012/opinião3, p. 5). A análise da expressão facial permite detectar doenças (como a depressão), inferir se alguém está mentindo e até reconhecer emoções (Fieux-Castagnet e Santucci, 2020, p. 5). Portanto, percebe-se que, neste caso, a tecnologia não serve para comparar um indivíduo com ele mesmo ou com vários outros indivíduos. Ela serve apenas para identificar as características deste indivíduo de forma anônima. Entretanto, se essas características estiverem ligadas a outros dados (por exemplo, dados de localização), torna-se possível identificar tal

⁴ "A galeria é o conjunto de modelos biométricos contra os quais é feita qualquer tarefa de verificação ou identificação. Para criar a galeria, as imagens do rosto de cada indivíduo precisam ser registradas pela SRF [sistemas de reconhecimento facial]" (Introna e Nissenbaum, 2010, p. 18).

indivíduo (FRA, 2019b, p. 8). Embora esta função da TRF seja promissora para o setor comercial, seu uso levanta sérias questões éticas, pois implica na categorização dos indivíduos com base em imagens faciais.

Os estudos do presente trabalho focarão principalmente na função de identificação, uma vez que esta é a função associada aos sistemas de vigilância e ao poder de polícia do Estado. Na formação das chamadas “listas de vigilância”, a TRF fornece ao Estado os melhores resultados, ou os resultados que atingiram um nível de similaridade acima de um valor pré-definido. Dessa forma, as pessoas que compõem a lista de vigilância são marcadas para investigações mais aprofundadas (Garvie *et al.*, 2016, p. 9).

A Aprendizagem de Máquina (através da Aprendizagem Profunda, geralmente supervisionada) é a técnica principal que embasa o funcionamento da TRF. Por meio desta técnica, os algoritmos são treinados com grandes volumes de dados a fim de melhorarem a sua performance e serem capazes de identificar padrões adicionais nos dados. Dessa forma, a sua tomada de decisão é aperfeiçoada e os algoritmos tornam-se capazes de tirar novas conclusões a partir da análise de novos dados (Surden, 2019, p. 1311-1312).

Este aperfeiçoamento se dá através dos seguintes procedimentos: 1) estipulação clara do problema; 2) coleta de dados; 3) anotação de dados (na aprendizagem supervisionada); 4) limpeza e representação de dados; 5) treinamento utilizando conjunto específico de dados de treinamento; 6) teste e ajuste utilizando um conjunto de dados de desenvolvimento; 7) avaliação e validação utilizando um conjunto de dados de teste. No aprendizado de máquina, tais passos não ocorrem necessariamente numa ordem linear. É um processo que vai e volta através destas etapas, sempre com o objetivo de construir e refinar os modelos (Lehr e Ohm, 2017, p. 655).

Tendo em vista a importância dos dados para o treinamento desses algoritmos, o fator decisivo para que uma tecnologia de Aprendizado de Máquinas tenha precisão e sucesso é justamente a utilização de grandes quantidades de dados de elevada qualidade (Surden, 2019, p. 1315). Com efeito, para que a TRF tenha um desempenho aceitável, é primordial que os dados utilizados possuam alta qualidade e diversidade, tanto em relação aos atributos das pessoas (gênero, etnia, idade), quanto em relação às peculiaridades das imagens (iluminação, ângulo, tamanho) (FRA, 2019a, p.1-18). Não bastasse, o desempenho também é influenciado pela qualidade das imagens presentes na galeria e das imagens de comprovação (Introna e Nissenbaum, 2010, p. 17).

Devido à sua maior precisão, as Redes Neurais Convolucionais (*Convolutional Neural Networks – CNN*) – uma das espécies de Redes Neurais Profundas (*Deep Neural Networks – DNN*) –, têm se tornado, nos últimos anos, a principal técnica que compõe as TRFs. Especificamente a partir 2014, com a utilização das *CNNs* para analisar imagens, ficou clara a evolução exponencial da precisão das TRFs (Shepley, 2019, p. 2).

As *DNNs* inspiram-se no funcionamento do cérebro humano, na medida em que – à semelhança dos neurônios – funcionam com base em múltiplas camadas de pequenas unidades de processamento interligadas entre si. Em linhas gerais, o funcionamento das *DNNs* pode ser dividido da seguinte forma: (I) camada de entrada (onde as informações da imagem de entrada são extraídas); (II) diversas camadas ocultas, que processam as informações (a potência da *DNN* está diretamente ligada ao maior número dessas camadas); (III) camada de saída (que na TRF geralmente subdivide-se entre a produção de uma pontuação de semelhança – identificação e verificação –, e a produção de uma classificação – categorização) (Guo e Zhang, 2019, p. 2).

Cada software funciona de um jeito. Não existe um padrão universal de reconhecimento facial, e cada banco de dados é criado e treinado para uma finalidade específica. No entanto, apesar da diversidade de técnicas usadas, as TRFs normalmente operam a partir de etapas comuns (Wechsler, 2007, p. 144). A primeira etapa consiste na aquisição da imagem digital, na qual consta a imagem do rosto de uma pessoa. Essa captura se dá a partir de um vídeo (câmeras CFTV) ou de uma foto (câmeras digitais). Em seguida, opera-se a detecção da face na imagem e o rosto é submetido a um procedimento de ajuste e adequação a fim de ficar com os mesmos padrões de cor, tamanho e iluminação das imagens da galeria. A terceira etapa consiste na extração das características. Neste momento a *CNN* funciona analisando a geometria do rosto, localizando fatores faciais únicos da pessoa, tais como, por exemplo, a distância entre os olhos, cantos da boca, distância entre a testa e o queixo, globos oculares. Com base nisso, codifica-se as informações biométricas e cria-se um descritor de características – uma espécie de “assinatura facial” (uma fórmula matemática) – . Por fim, compara-se a assinatura facial a uma galeria (banco de dados) que contém rostos (pré-coletados e armazenados) de uma pessoa (quando se dá a verificação) ou de várias pessoas (quando se dá a identificação). Se o cálculo das pontuações de similaridade atingirem um valor acima de um determinado patamar, considera-se que as imagens faciais comparadas são da mesma pessoa. No entanto, é preciso destacar que não existe uma TRF 100% precisa. (Article 29 Data Protection Working Party, 2012Opinião 2, p. 2; Gordon, 2021, p. 8-9).

Apesar dos ganhos que a *CNN* traz à TRF, ela aumenta a opacidade da tecnologia. E isso é uma grande preocupação, na medida em que, por estarem ocultas as camadas de processamento, torna-se praticamente impossível saber quais características e padrões faciais estão servindo de base para o reconhecimento. Nem sequer os programadores conseguem saber ou entender. Tal fato contribui para a acentuação da característica de “caixa-preta”⁵ que existe na tecnologia e acaba por dificultar a detecção e a correção dos enviesamentos que o sistema possa vir a ter.

1.3. OS USOS DA TRF PELAS AUTORIDADES POLICIAIS

As TRFs têm sido usadas tanto pelo setor privado quanto pelo setor público. Suas finalidades variam e muitas vezes se confundem: vão desde usos comerciais até usos para a segurança pública e vigilância. Apenas a título exemplificativo, é possível destacar os seguintes usos: (I) captação e classificação das reações dos consumidores quanto a produtos/serviços⁶ e publicidade⁷; (II) na educação, para checar a presença⁸ e a concentração dos alunos⁹; (III) nas igrejas, para detecção de presença e comportamento dos fiéis¹⁰; (IV) controle de aeroportos¹¹ e fronteiras¹²; (V) localização de desaparecidos¹³; (VI) acesso a dispositivos móveis¹⁴ e marcação de fotos em redes

⁵ Nesse sentido: “Caixa-preta de IA e explicabilidade. Algumas técnicas de aprendizagem de máquinas, embora muito bem-sucedidas do ponto de vista da precisão, são muito opacas em termos de compreensão de como elas tomam decisões. A noção de caixa-preta de IA se refere a tais cenários, onde não é possível rastrear a razão de certas decisões. A explicabilidade é uma propriedade dos sistemas de IA que, em vez disso, podem fornecer uma forma de explicação por suas ações.” (HLEGAI, 2019a, p. 5)

⁶ Caso Hering, no Brasil: <https://idec.org.br/noticia/apos-denuncia-do-idec-hering-e-condenada-por-uso-de-reconhecimento-facial>.

⁷ Caso da Linha 4 (Amarela) do Metrô de São Paulo, Brasil: <https://idec.org.br/idec-na-imprensa/viaquatro-e-condenada-por-reconhecimento-facial-sem-autorizacao-no-metro-de-sp>.

⁸ No Brasil: <https://olhardigital.com.br/2019/10/24/noticias/professores-brasileiros-realizam-chamada-por-reconhecimento-facial/>; uso em escola na cidade de Lockport, em Nova York: <https://www.govtech.com/education/k-12/judge-ends-nyclu-suit-against-school-facial-recognition>.

⁹ Na China: <https://www.scmp.com/news/china/society/article/2146387/pay-attention-back-chinese-school-installs-facial-recognition>.

¹⁰ Ver: <https://www.cartacapital.com.br/sociedade/empresas-lancam-servico-de-reconhecimento-facial-para-igrejas-no-brasil/>.

¹¹ No Brasil: <http://intra.serpro.gov.br/tema/noticias-tema/reconhecimento-facial-intensifica-seguranca-nos-aeroportos>; nos EUA: <https://www.nytimes.com/2021/12/07/travel/biometrics-airports-security.html>.

¹² Projeto Fronteira Tech, no Brasil: <https://www.aen.pr.gov.br/Noticia/Ponte-da-Amizade-ganha-monitoramento-inteligente>.

¹³ No Brasil: <https://www.uol.com.br/tilt/noticias/redacao/2020/12/20/app-usa-reconhecimento-facial-na-busca-por-pessoas-desaparecidas.htm>.

¹⁴ Ver: <https://www.tecmundo.com.br/software/232810-iphone-ganha-desbloqueio-via-face-id-mascara-ios-15-4.htm>.

sociais¹⁵; (VII) monitoramento de cidades inteligentes¹⁶; (VIII) *Social Credit Scores*¹⁷; (IX) compras¹⁸; (X) acesso a serviços públicos¹⁹; (XI) segurança pública e vigilância²⁰; (XII) combate a fraudes²¹; (XII) combate à pandemia de COVID-19²²; (XIV) guerras – como na guerra entre Rússia e Ucrânia²³.

A TRF acaba assumindo especial relevância para as autoridades policiais, na medida em que permite que estas exerçam vigilância sem serem notadas. Por meio da tecnologia, torna-se possível localizar e identificar pessoas a longas distâncias e em tempo real, sem que estas saibam ou consentam. Não bastasse, por também ser possível identificar suspeitos “em tempo diferido”, a tecnologia também acaba assumindo importante valor para as investigações criminais.

A Proposta de Regulamento de IA da União Europeia tem o intuito de ser aplicável a todos os sistemas de identificação biométrica remota (*remote biometrics identification systems – RBI*), incluindo as TRFs. Tais sistemas possuem as seguintes características: são operados à distância sem se saber antecipadamente se a pessoa visada estará presente na área; coletam dados biométricos; comparam tais dados a um banco de dados; e são usados especificamente para identificar uma pessoa²⁴.

¹⁵ Ver: <https://www.techtudo.com.br/noticias/2017/12/facebook-muda-reconhecimento-facial-e-avisa-quando-voce-aparece-em-fotos.ghml>.

¹⁶ Projeto “Cidade Inteligente”, de São José dos Campos, Brasil: <https://revistasegurancaeletronica.com.br/sao-jose-dos-campos-adquire-mil-novas-cameras-com-reconhecimento-facial-para-reforçar-segurança/>; Projeto “Vila A”, em Foz do Iguaçu, Brasil: <https://agenciabrasil.ebc.com.br/geral/noticia/2020-07/foz-do-iguacu-tera-o-primeiro-bairro-inteligente-do-pais>

¹⁷ Na China: <https://www.forbes.com/sites/bernardmarr/2019/01/21/chinese-social-credit-score-utopian-big-data-bliss-or-black-mirror-on-steroids/>.

¹⁸ Na China: <https://www.theguardian.com/world/2019/sep/04/smile-to-pay-chinese-shoppers-turn-to-facial-payment-technology>; Ver também: <https://www.bbc.com/news/business-55748964>.

¹⁹ Em Portugal: <https://tek.sapo.pt/noticias/internet/artigos/reconhecimento-facial-nos-servicos-publicos-vai-avancar-associado-a-chave-movel-digital>.

²⁰ Na União Europeia: <https://www.wired.com/story/europe-police-facial-recognition-prum/>; programa RIO+SEGURO, no Rio de Janeiro, Brasil: <https://prefeitura.rio/cidade/programa-rioseguro-chega-a-taquara-e-vai-se-expandir-ainda-mais-pela-zona-oeste/>; Na Inglaterra: <https://www.wired.co.uk/article/met-police-facial-recognition-new>; nos Estados Unidos da América: <https://www.washingtonpost.com/technology/2019/07/07/fbi-ice-find-state-drivers-license-photos-are-gold-mine-facial-recognition-searches/>.

²¹ Na Irlanda: <https://findbiometrics.com/biometrics-news-irish-department-social-protection-confirms-use-facial-recognition-021302/>.

²² Na Coreia do Sul: <https://www.institutodeengenharia.org.br/site/2021/12/15/coreia-do-sul-testara-reconhecimento-facial-para-rastrear-casos-de-covid-19/>; no Canadá: <https://www.poder360.com.br/internacional/hospital-usa-reconhecimento-facial-para-monitorar-pacientes-com-covid-19/>.

²³ Ver: <https://www.opendemocracy.net/en/technology-and-democracy/facial-recognition-ukraine-clearview-military-ai/>.

²⁴ Ver considerando 8 e art. 3 (33) da Proposta de Regulamento de IA da Comissão Europeia, 2021/0106 (COD).

Ainda de acordo com a proposta da Comissão Europeia, tendo em vista as diferentes características e formas como são utilizados, bem como os diferentes riscos inerentes, deve ser feita uma diferenciação entre os sistemas de identificação biométrica remota “em tempo real” (*real-time biometric identification systems*) e os sistemas de identificação biométrica remota “em tempo diferido” (*post biometric identification systems*). Os primeiros são definidos como sistemas que capturam dados biométricos e executam os processos de comparação e identificação de forma instantânea (ou sem significativo atraso), com base em imagens obtidas “em tempo real” ou “quase em tempo real”, geradas por câmeras ou dispositivos parecidos. Por sua vez, os sistemas que operam “em tempo diferido” são aqueles que coletam os dados biométricos e processam a comparação e identificação após um atraso significativo. Funcionam com base em imagens geradas por CFTV ou dispositivos privados.²⁵

De maneira geral, as autoridades podem usar a TRF para realizar quatro tarefas principais de identificação: (I) “prender e identificar” (sujeita-se a fotografia da pessoa presa ao reconhecimento facial); (II) “parar e identificar” (utiliza-se a tecnologia – instalada em um smartphone ou carro de polícia – a fim de descobrir a identidade de uma pessoa detida na rua); (III) “investigar e identificar” (sujeita-se a imagem de um suspeito ao reconhecimento facial); (IV) “videovigilância em tempo real” (usa-se a TRF para comparar de forma continuada imagens coletadas em câmeras de segurança ao vivo com imagens de pessoas relacionadas em listas de vigilância) (Garvie *et al.*, 2016, p. 10).

A grande preocupação é que o uso indiscriminado da TRF por autoridades²⁶ potencializa uma série de riscos, tais como: vigilância em massa²⁷; racismo²⁸;

²⁵ Ver: Considerando 8 da Proposta de Regulamento de IA da Comissão Europeia, 2021/0106 (COD).

²⁶ “Autoridade policial significa (a) uma autoridade pública competente para efeitos de prevenção, investigação, detecção ou repressão de infrações penais ou execução de sanções penais, incluindo a proteção contra ameaças à segurança pública e a prevenção das mesmas; ou (b) qualquer outro organismo ou entidade designados pelo direito de um Estado-Membro para exercer autoridade pública e poderes públicos para efeitos de prevenção, investigação, detecção ou repressão de infrações penais ou execução de sanções penais, incluindo a proteção contra ameaças à segurança pública e a prevenção das mesmas.” (art. 3(40), Proposta de Regulamento de IA da UE).

²⁷ A possibilidade de interação da TRF com praticamente todos os tipos de dispositivos de vídeo (CFTV; smartphones; telas de publicidade) e captação de som, pode levar a um nível de vigilância sem precedentes. Seria um aprofundamento ainda maior da vigilância em massa, que praticamente acaba com a privacidade e o anonimato, e afeta o exercício de outros direitos fundamentais (proteção de dados pessoais, liberdade de expressão e reunião, direito de protesto, liberdade de consciência e liberdade de religião). A generalização do reconhecimento facial no espaço público pode corroer o anonimato e criar uma sensação de vigilância ininterrupta, afetando o comportamento das pessoas no espaço público (CNIL, 2019, p.7).

²⁸ As diferenças significativas em relação à acurácia das TRFs na avaliação de pessoas não brancas demonstram que a tecnologia não é neutra e reflete o racismo estrutural existente na sociedade (LAPIN, 2021, p.8); ver também: Leslie D., Understanding bias in facial recognition technologies, The Alan Turing Institute, 2020. Disponível em: <https://www.turing.ac.uk/research/publications/understanding-bias-facial-recognition-technologies>.

transfobia²⁹; violação de direitos de crianças e adolescentes³⁰; e, mais especificamente, violação de direitos fundamentais. Quanto aos últimos (direitos fundamentais), é possível destacar os seguintes direitos ameaçados: privacidade; proteção de dados pessoais; liberdade de ir e vir; liberdade de reunião e associação; liberdade de expressão; liberdade religiosa; julgamento justo; e o direito à não discriminação (FRA, 2019b, p. 4).

Diante desse cenário de tantos possíveis danos, muito se tem perguntado se a tecnologia estaria madura o suficiente para ser utilizada na sociedade. Após denúncias de enviesamento, o setor privado, representado pelas *Big Techs*, tem revisto o seu posicionamento e, em alguns casos, até mesmo pausado ou abandonado suas pesquisas. Em junho de 2020, a IBM³¹, por exemplo, anunciou que não investiria mais em TRF, tendo em vista que esta tecnologia estaria sendo usada para controle social e opressão pelas forças policiais. Na mesma época, a Amazon proibiu que sua TRF fosse usada para fins policiais por um ano. Posteriormente, em maio de 2021, decidiu estender tal proibição por tempo indeterminado³². Seguindo essa tendência, a Microsoft³³ também indicou que não venderá sua TRF para a polícia norte-americana.

Por sua vez, o setor público (em alguns países) começa a dar passos em direção ao banimento ou moratória da tecnologia. Por exemplo, cidades como São Francisco³⁴, Nova York³⁵, Cambridge³⁶ e Portland³⁷ decidiram banir a tecnologia. Na cidade do Rio de Janeiro, tramita na Câmara Municipal um projeto de lei³⁸ que visa proibir o uso de

²⁹ Por funcionar com base em critério binários, a TRF pode reforçar a exclusão e o estigma de pessoas transgênerosexuais e não-binárias. Isso porque a tecnologia pode reiteradamente negar visibilidade a identidades divergentes e conflitar com a auto-identificação de gênero. Tal fator tem o potencial de acirrar violências e reiterar o cerceamento de direitos às pessoas transexuais e não-binárias (LAPIN, 2021, p. 9). Ver também: Coding Rights, Relatório Reconhecimento Facial no Setor Público e Identidade Trans, 2021. Disponível em: <https://codingrights.org/docs/rec-facial-id-trans.pdf>.

³⁰ Devido à menor precisão com que a tecnologia detecta rostos jovens que mudam rapidamente, crianças e adolescentes são indivíduos vulneráveis que merecem um padrão de proteção mais elevado. Além disso, por não terem a capacidade de discernimento completamente formada, estão mais vulneráveis ao mau uso de seus dados (EPRS, 2021, p.8; LAPIN, 2021, p.9).

³¹ Ver: <https://www.theverge.com/2020/6/8/21284683/ibm-no-longer-general-purpose-facial-recognition-analysis-software>.

³² Ver: <https://g1.globo.com/economia/tecnologia/noticia/2021/05/19/amazon-estende-proibicao-de-uso-de-reconhecimento-facial-pela-policia-por-tempo-indeterminado.ghtml>.

³³ Ver: <https://canaltech.com.br/apps/microsoft-se-recusa-a-vender-reconhecimento-facial-a-policia-nos-eua-166337/>.

³⁴ Ver: <https://www.bbc.com/news/technology-48276660>.

³⁵ Ver: <https://www.nyclu.org/en/publications/what-you-need-know-about-new-yorks-temporary-ban-facial-recognition-schools>.

³⁶ Ver: <https://tab.uol.com.br/noticias/redacao/2020/01/16/cambridge-bane-reconhecimento-facial-entenda-por-que-temos-de-pensar-nisso.htm>.

³⁷ Ver: <https://www.dwt.com/blogs/privacy--security-law-blog/2021/01/portland-facial-recognition-ban>.

³⁸ Projeto de Lei nº 824/2021, de autoria do vereador Reimont. Disponível em: <http://aplicnt.camara.rj.gov.br/APL/Legislativos/scpro2124.nsf/ab87ae0e15e7d4dd0325863200569395/33b9222f266e43710325872700723005?OpenDocument>

TRF pelo Poder Público Municipal. Também na Assembleia Legislativa do Estado do Rio de Janeiro tramita um projeto de lei³⁹ no mesmo sentido. De fato, os efeitos do uso de uma TRF enviesada são extremamente lesivos para as vítimas de uma possível discriminação perpetrada pelo Estado. Tais efeitos podem constituir-se, por exemplo, em perseguições políticas e religiosas, investigações infundadas, falsas acusações de fraude, perseguições a jornalistas e ativistas, prisões injustas⁴⁰.

Em outubro de 2021, o grupo parlamentar “*Greens/EFA (European Free Alliance)*”, publicou um estudo encomendado, no qual buscou mapear os usos atuais da TRF especificamente na UE. Intitulado “*Biometric & Behavioural Mass Surveillance in EU Member States*”, o estudo revelou que as implantações atuais da TRF na Europa são principalmente experimentais e localizadas, mas com potencial de se tornarem sistemas de vigilância biométrica em massa.

Até à data de publicação do estudo, autoridades policiais de 11 países membros da UE estavam usando TRF “em tempo diferido” frente a bancos de dados biométricos para fins de investigação criminal: Áustria, Finlândia, França, Alemanha, Grécia, Hungria, Itália, Letônia, Lituânia, Holanda e Eslovênia. Além disso, outros 7 países deveriam seguir esse caminho em breve: Croácia, República Tcheca, Romênia, Espanha, Suécia, Chipre, Estônia.

Foi relatado, ainda, que a Interpol possui um sistema de reconhecimento facial baseado em imagens faciais recebidas de mais de 160 países. A Europol, por sua vez, possui duas subunidades que utilizam a ferramenta de busca de reconhecimento facial e o banco de dados conhecido como FACE: o Centro Europeu de Combate ao Terrorismo (*European Counter Terrorism Center – ECTC*) e o Centro Europeu de Combate ao Cibercrime (*European Cybercrime Center – ECC*).

Em relação às bases de dados utilizadas, o estudo apontou que alguns países limitam as buscas a bases de dados criminais (Áustria, Alemanha, França, Itália, Grécia, Eslovênia, Lituânia, Reino Unido), enquanto outros países realizam as buscas também nas bases de dados civis (Finlândia, Holanda, Letônia, Hungria). Infere-se disso que as categorias de pessoas que terão seus dados processados por uma TRF podem variar

³⁹ Projeto de Lei nº 5240/2021, de autoria da deputada Dani Monteiro: http://www3.alerj.rj.gov.br/lotus_notes/default.asp?id=144&url=L3NjcHJvMTkyMy5uc2YvMThjMWRkNjhmOTZiZTNlNzgzMjU2NmVjMDAxOGQ4MzMvYTdkYTU1NGFiZWUzMTBhYTAzMjU4N2E1MDA1YmNmMDE%2FT3BlbkRvY3VtZW50&s=09#.

⁴⁰ Por exemplo, uma mulher foi detida por engano no Rio de Janeiro, Brasil: <https://www.poder360.com.br/brasil/reconhecimento-facial-da-pm-do-rio-falha-e-mulher-e-detida-por-engano/>

substancialmente. Em relação aos bancos de dados criminais, a variação pode ir desde suspeitos e condenados, até requerentes de asilo, estrangeiros, pessoas não identificadas, imigrantes, requerentes de visto. Apenas para se ter ideia dos tamanhos dos bancos de dados criminais, pode-se citar como exemplo: o INPOL, da Alemanha, com 6,2 milhões de indivíduos; o TAJ, na França, com 21 milhões; e o AFIS, na Itália, com 9 milhões (Ragazzi *et al.*, 2021, p.39).

A partir do momento em que também se utilizam bancos de dados civis, amplia-se o leque para também processar os dados de indivíduos que têm a sua identidade conhecida de vários documentos e procedimentos civis. Como exemplo de bancos de dados civis, pode-se citar o Registro de Imagens Faciais da Hungria, que contém 30 milhões de modelos faciais (Ragazzi *et al.*, 2021, p.39).

De acordo com o relatório, a TRF também compõe soluções integradas de cidade segura, como o sistema “*NEC Technology Bio-IDiom*”, implantado em Lisboa⁴¹ e Londres para fins de investigação criminal. Neste caso específico, o reconhecimento pode ocorrer via reconhecimento facial, bem como pela combinação de múltiplas tecnologias biométricas, tais como impressão da palma da mão, impressão digital, reconhecimento da íris, reconhecimento das veias do dedo (Ragazzi *et al.*, 2021, p.39).

O estudo destaca que, devido à estrita estrutura legal à qual as autoridades policiais são submetidas, a aquisição de imagens e as subsequentes buscas individuais são realizadas como parte das investigações judiciais quando um limiar legal de suspeita é atingido. É preciso que as autoridades sigam os limites processuais estabelecidos pelo processo judicial (Ragazzi *et al.*, 2021, p.39). Em relação às buscas, estas são realizadas por especialistas ou por grupos de policiais, e podem enumerar até mil suspeitos para análise humana. É preciso ressaltar que cada país treina os operadores da TRF de forma diferente, e nem todos aplicam padrões de qualidade às imagens (TELEFI, 2021, p.29-31).

O primeiro país da UE a usar TRF para investigações criminais foi a Alemanha, que hoje tem a *BKA*⁴² e mais 16 departamentos de polícia estaduais utilizando a tecnologia como ferramenta (TELEFI, 2021, p.22). A França, por outro lado, é o país que mais usa a TRF, com a sua *Gendarmerie*⁴³ utilizando a tecnologia para realizar buscas e indicar suspeitos para linhas de investigação (TELEFI, 2021, p.31).

⁴¹ Para mais informações: https://pt.nec.com/pt_PT/press/202201/20220104_01.html.

⁴² Bundeskriminalamt (Departamento Federal de Polícia Criminal).

⁴³ É uma força militar encarregada da realização de funções de polícia no âmbito da população civil.

Também é possível identificar na UE alguns casos de uso de TRF para “vigilância inteligente”, e não propriamente para investigação policial. Tais sistemas inteligentes não coletam ou processam informações biométricas per se. Eles acabam sendo usados em “tempo diferido” para auxiliar no processamento de grandes quantidades de informações gravadas, ou para orientar a atenção dos operadores de CFTV no monitoramento de grande número de transmissões de vídeo ao vivo simultaneamente (Ragazzi *et al.*, 2021, p.41). Podem ser citados os seguintes exemplos: (I) detecção de comportamentos anormais (Toulouse); (II) busca de características (Marbella); (III) resumo de vídeos longos combinado com reconhecimento facial (Vannes, Roubaix e Moirans); (IV) reconhecimento de movimento (Marbella); (V) reconhecimento de áudio, como complemento a imagens (Saint-Etienne, Rouen, Estrasburgo); (VI) reconhecimento de emoções, para vigiar ameaças (projetos-piloto nunca implantados em Nice e Eindhoven); (VII) sistemas integrados para cidades inteligentes seguras (Veneza, Valenciennes, Brienon-sur-Armançon, Nice, Marselha, Roubaix, Madri, Lisboa e Londres); (VIII) gerenciamento de multidões (Valenciennes, Marselha, Mannheim, Veneza, Amsterdã, Eindhoven) (Ragazzi *et al.*, 2021, p.41-44).

Os exemplos acima destacados correspondem a variações de aplicações de segurança baseadas no processamento algorítmico de imagens e sons. O número de implantações que correspondem de fato à TRF em tempo real é relativamente pequeno. Geralmente eles são apresentados como “pilotos”, limitados no tempo e muitas vezes rapidamente interrompidos por razões legais (Ragazzi *et al.*, 2021, p.44). Podem ser listados os seguintes exemplos de TRF ao vivo na UE: (I) Projeto Dragonfly, na Hungria, de 2019 até o presente (35.000 câmeras CFTV ligadas a bancos de dados biométricos permitem que a polícia identifique qualquer cidadão pelo rosto, já tendo resultado em 6.000 correspondências, 250 buscas e 4 prisões); (II) projeto piloto de TRF ao vivo durante o Carnaval de Nice, França, em 2019; (III) projeto piloto de TRF ao Vivo no Aeroporto Internacional de Bruxelas, Bélgica, de 2017 a 2019 (foram usadas 4 câmeras ligadas a uma TRF, a fim de capturar e isolar rostos e depois verificá-los em uma lista de suspeitos. Ocorreram vários falsos positivos e a tecnologia se mostrou ineficaz. O sistema foi suspenso, pois operava sem base legal); (IV) projeto-piloto de TRF ao vivo em Südkreuz, Berlin, Alemanha, de 2017 a 2018 (a Polícia Federal Alemã usou TRF na Estação Ferroviária de Südkreuz para combinar rostos capturados em CFTV com fotos de alta qualidade de pessoas. Descobriu-se que a tecnologia criava um número significativo de falsos positivos e o projeto foi encerrado em 2018); (V) Hamburgo, Alemanha, de 2017

a 2020 (após os protestos no G20, a polícia de Hamburgo passou a usar TRF para traçar o perfil dos manifestantes. Houve uma batalha legal de três anos com a Autoridade de Proteção de Dados de Hamburgo) (Ragazzi *et al.*, 2021, p.65-106).

Recentemente, foi revelado que a UE está construindo um enorme sistema internacional de reconhecimento facial, que permitirá às forças policiais dos países membros vincularem seus bancos de dados de fotos com milhões de imagens de rostos⁴⁴. Essa expansão da TRF pela UE deriva de um plano mais amplo⁴⁵ de modernizar o policiamento em todo o continente. Mais especificamente, a expansão da TRF encaixa-se na proposta de compartilhamento de dados do Prüm II⁴⁶, que visa ampliar a quantidade de informações que podem ser compartilhadas entre as autoridades policiais, incluindo imagens faciais e informações de carteiras de motorista. No entanto, de acordo com a proposta, o Prüm não permitirá o uso de TRF “em tempo real”, mas apenas o uso de TRF “em tempo diferido”. Isso quer dizer que as autoridades policiais somente poderão usar imagens gravadas de CFTV, fotos de mídias sociais e telefones (Burgess, 2022).

Com o novo sistema, um país poderá comparar uma foto com os bancos de dados de outros países e descobrir se há correspondência. Como os bancos de dados variam de um país para o outro, as imagens poderão incluir suspeitos, condenados por crimes, requerentes de asilo e “cadáveres não identificados” (Burgess, 2022). De acordo com o porta-voz da Comissão Europeia, não será permitida a correspondência de imagens com a população em geral. Somente poderão ser trocadas imagens faciais de suspeitos ou criminosos condenados. E mais: não haverá um banco de dados central gigante, mas sim um roteador central que conecta as autoridades policiais, agindo como um intermediário de mensagens entre os Estados (Comissão Europeia, 2021). Não será

⁴⁴ Ver: Europe Is Building a Huge International Facial Recognition System. Wired. 2022. Disponível em: <https://www.wired.com/story/europe-police-facial-recognition-prum/>.

⁴⁵ Código de Cooperação Policial da UE, que visa reforçar a cooperação policial entre os Estados-Membros e fornecer aos agentes policiais da UE ferramentas mais modernas para o intercâmbio de informações. Este Código visa melhorar as operações transfronteiriças, fornecer canais e prazos claros para o intercâmbio de informações e dar à Europol um papel mais forte. O pacote de cooperação policial inclui três propostas legislativas: (I) Recomendação do Conselho sobre a cooperação policial operacional; (II) Diretiva relativa ao intercâmbio de informações entre as autoridades policiais dos Estados-Membros; (III) Regulamento sobre Intercâmbio Automatizado de Dados para a Cooperação Policial (Prüm II), que permite a troca mais rápida de dados de especial importância para a aplicação da lei e expande os dados disponíveis de DNA, impressões digitais, dados de registro de veículos agora para imagens faciais e registros policiais. Comissão Europeia, 2021, disponível em: https://ec.europa.eu/home-affairs/news/boosting-police-cooperation-across-borders-enhanced-security-2021-12-08_en.

⁴⁶ Ver: Proposta de Regulamento do Parlamento Europeu e do Conselho relativo ao intercâmbio automatizado de dados para efeitos de cooperação policial (Prüm II), que altera as Decisões 2008/615/JAI e 2008/616/JAI do Conselho e os Regulamentos (UE) 2018/1726, (UE) 2019/817 e (UE) 2019/818 do Parlamento Europeu e do Conselho. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:52021PC0784&from=EN>.

armazenado nenhum dado. Por fim, as correspondências serão revisadas por um humano, que decidirá se estão corretas, antes que outra ação seja tomada (Burgess, 2022).

1.4. DESAFIOS ATINENTES À TRF

1.4.1. FALTA DE PRECISÃO (ACURÁCIA)

Antes de aprofundarmos a discussão sobre vieses nas TRF é preciso discutir um problema técnico que permeia essas tecnologias: a falta de precisão (acurácia). No início de 2020, um homem negro, chamado Robert Williams⁴⁷, foi preso em Detroit (EUA) depois da TRF utilizada pela polícia estadual errar ao apontar similaridades entre a foto de Williams e as imagens das câmeras de segurança de uma relojoaria assaltada. Williams, que ficou preso por mais de trinta horas, só foi liberado após o dono da loja confirmar que não foi ele quem cometeu o delito. Este foi o primeiro caso documentado de uma pessoa presa indevidamente devido a erro de um *software* (Garvie, 2020).

No Reino Unido, as TRFs usadas em aeroportos têm confundido viajantes com criminosos foragidos e servido como instrumento de detenção de pessoas sem respeito ao devido processo legal⁴⁸. Estudo feito pela Universidade de Essex constatou que a TRF usada pela polícia de Londres para fazer o reconhecimento facial em tempo real possui taxa de acerto de 19,05%: ou seja, da identificação de 42 indivíduos como suspeitos, só 8 estavam corretas (Fussey e Murray, 2019, p.10).

Esse problema na acurácia sujeita as TRF a fraudes. Estudo feito pela *Consumentenbond*, na Holanda, mostrou que, de 110 smartphones testados (de diferentes marcas), 42 conseguiam ser desbloqueados por meio da foto do dono (Zizi, 2019). Até mesmo o *TrueDepth*⁴⁹, sistema da Apple, pode ser fraudado através de uma impressão 3D ou uma Deepfake. Além disso, algumas TRF⁵⁰ têm dificuldade de fazer a diferenciação correta entre gêmeos idênticos. Pesquisa da Universidade de Virgínia

⁴⁷ Ver: <https://www.aclu.org/press-releases/man-wrongfully-arrested-because-face-recognition-cant-tell-black-people-apart>.

⁴⁸ Ver: <https://www.theguardian.com/politics/2019/sep/07/border-control-systems-face-fire-from-travellers-wrongly-delayed>.

⁴⁹ Tecnologia que, além de usar Deep Learning, projeta 50 mil pontos de infravermelho na face, usando cada um desses pontos como informação adicional. Essa tecnologia pega muito mais informação além da imagem em si. A taxa de erro é menor, porque existe mais informação. Nesse sentido: <https://support.apple.com/en-us/HT208108>.

⁵⁰ Por exemplo, o FaceId, da Apple. Ver: <https://www.youtube.com/watch?v=e8-yupM-6Oc>.

Ocidental (EUA) demonstrou que, apesar de ser possível fazer a diferenciação entre gêmeos idênticos, a capacidade de identificação do algoritmo se deteriora com o aumento do número de imagens no banco de dados.

Sendo assim, diante da constatação de que os diferentes tipos de TRF existentes atualmente ainda sofrem de uma preocupante falta de precisão, torna-se necessário analisarmos a tecnologia com cautela. Não é aconselhável que depositemos nossa confiança de forma cega na tecnociência e a consideremos uma autoridade inquestionável. No fim das contas, o conhecimento orientado por dados é limitado pela discricionariedade humana, a qual é enviesada (Hong, 2020, p.187-192).

1.4.2. ENVIESAMENTO

A confiança na objetividade da ciência não pode ser motivo para o negligenciamento de questões como o enviesamento. Sistemas de segurança não estão isentos de conter e reproduzir preconceitos enraizados na sociedade (Gilliom e Monahan, 2013, p.125). De uma forma geral, é possível constatar que há nas pessoas uma ideia de que as tecnologias baseadas em IA (dentre elas, a TRF), por utilizarem algoritmos complexos, são capazes de atuarem sem o preconceito e a discricionariedade humanos. Há a sensação de que a substituição da atividade humana pela tecnologia resultará em menos vieses inconscientes (Polli, 2017).

Nos últimos anos, no entanto, tem aumentado entre estudiosos⁵¹ e instituições⁵² o debate a respeito da precisão e do enviesamento de sistemas de IA. Um número crescente de estudos⁵³ vem tentando entender como preconceitos humanos (dentre eles, o de gênero

⁵¹ Por exemplo: David Danks and Alex John London, ‘Algorithmic Bias in Autonomous Systems’ (2017) 17 IJCAI 4691.

⁵² Por exemplo: Proposta de Regulamento de IA da Comissão Europeia, 2021/0106 (COD); Council of the European Union, ‘Presidency conclusions - The Charter of Fundamental Rights in the context of Artificial Intelligence and Digital Change’ (2020) 11481/20; Comissão Europeia, ‘Livro Branco sobre a inteligência artificial – Uma abordagem europeia virada para a excelência e a confiança, COM(2020) 65 final.

⁵³ Por exemplo: Leslie D., Understanding bias in facial recognition technologies, The Alan Turing Institute, 2020. Disponível em: <https://www.turing.ac.uk/research/publications/understanding-bias-facial-recognition-technologies>; LAPIN – Laboratório de Políticas Públicas e Internet. Vigilância automatizada: uso de reconhecimento facial pela Administração Pública, julho de 2021. Disponível em: <https://lapin.org.br/2021/07/07/vigilancia-automatizada-uso-de-reconhecimento-facial-pela-administracao-publica-no-brasil/>; Buolamwini, J. and Gebru, T. ‘Gender shades: Intersectional accuracy disparities in commercial gender classification’ (2018) Conference on fairness, accountability and transparency p.77-91; Phillips, J.P. *et al.*, ‘An other-race effect for face recognition algorithms’ (2009) 8(2) ACM Transactions on Applied Perception (TAP), p.1-11; Grother, P.J., Ngan, M. and Hanaoka, K. ‘Face Recognition Vendor Test (FVRT): Part 3, Demographic Effects’ (2019) National Institute of Standards and Technology; Klare, B.F. *et al.*, ‘Face Recognition Performance: Role of Demographic Information’ (2012) 7(6) IEEE Transactions on Information Forensics and Security, p.1789-1801.

e de raça) se incorporam às novas tecnologias (como a TRF)⁵⁴. É importante debater sobre os vieses embutidos nas TRFs, a fim de desmistificar a ideia de que a tecnologia é uma alternativa neutra e imparcial. Com efeito, quanto mais enraizado for um preconceito em uma determinada sociedade, maior é a probabilidade de que esse preconceito seja replicado por um algoritmo treinado com os dados fornecidos por essa mesma sociedade (Gilliom e Monahan, 2013, p.126).

O viés está relacionado à visão de mundo limitada que uma pessoa ou grupo tem. Quando adstrito a um indivíduo, ele opera na razão de um-para-um. Mas quando relacionado a uma IA, ele opera na razão de um-para-muitos, amplificando o seu impacto sobre a sociedade de forma muito perigosa. Tem-se com isso, portanto, que os algoritmos podem se transformar em verdadeiros vetores de perpetuação de preconceitos, discriminações e injustiças (Frazão, 2019, p.39).

Em relação aos estudos que vêm discutindo o enviesamento das novas tecnologias, é possível destacar o “The Maximum Surveillance Society: The Rise of CCTV”, realizado por Norris e Armstrong, em 1999. Por meio deste estudo, os pesquisadores constataram que a vigilância realizada por CFTV na Inglaterra é direcionada e está condicionada a variáveis como idade, classe, gênero e raça. Dentre outras conclusões, os autores constataram que a chance de a população negra ser alvo de vigilância era de 150% a 250% maior do que a população branca.

Em 2003 foi publicado o estudo “2002 NIST Face Recognition Vendor Test (FRVT)”, que demonstrou a existência de diferenças na precisão da TRF quando aplicada a grupos demográficos distintos. Com base na avaliação de dez algoritmos em um conjunto de dados de 120.000 imagens, constatou-se maior sucesso no reconhecimento de homens do que no de mulheres (Phillips *et al.*, 2003).

Já em 2009, foi publicado o estudo “An other-race effect for face recognition algorithms”. Tal estudo objetivou analisar o desempenho de dois algoritmos distintos (um ocidental⁵⁵ e outro do leste asiático⁵⁶) ao executarem tarefas de verificação que envolveram indivíduos caucasianos e do leste asiático. A conclusão foi que, assim como os humanos, os algoritmos também estão sujeitos a um efeito “other-race”. Isso significa que temos a capacidade limitada de identificar rostos de raças diferentes das nossas

⁵⁴ Nesse sentido, ver: University of Bath. Biased bots: Human prejudices sneak into AI systems. Bath, 20 dez. 2018. Disponível em: <https://www.bath.ac.uk/announcements/biased-bots-human-prejudices-sneak-into-ai-systems/>.

⁵⁵ Algoritmo criado através da fusão de algoritmos da França, da Alemanha e dos EUA.

⁵⁶ Algoritmos criado por meio da fusão de algoritmos da China, Japão e Coreia.

(Phillips *et al.*, 2009, p.1-2). Sendo assim, ao dar prioridade à precisão (focando somente em baixas taxas de aceitação falsas⁵⁷), o algoritmo ocidental teve melhor desempenho em relação às pessoas caucasianas, e o algoritmo do leste asiático teve em relação aos asiáticos do leste.

Em 2010 foi publicado o estudo “*Report on the Evaluation of 2D Still-Image Face Recognition Algorithms*”, que avaliou dez algoritmos em três conjuntos de dados relacionados a 4 milhões de pessoas. Assim como no “*2002 NIST Face Recognition Vendor Test (FRVT)*”, ficou constatada a maior dificuldade de reconhecimento facial de mulheres, do que de homens. No entanto, ao contrário do que viria a se tornar predominante em pesquisas posteriores, as pessoas negras foram reconhecidas com maior sucesso do que as brancas (Grother *et al.*, 2010, p. 56).

Já em 2012, foi publicado o estudo “*Face Recognition Performance: Role of Demographic Information*”, que procurou avaliar a precisão de seis algoritmos em relação a diferentes etnias e gêneros. Por meio de tal estudo, constatou-se ser mais difícil reconhecer mulheres do que homens, e mais difícil reconhecer negros do que hispânicos e brancos.

Buolamwini e Gebre, por sua vez, lançaram em 2018 o estudo “*Gender shades: Intersectional accuracy disparities in comercial gender classification*”, no qual analisaram a precisão de três algoritmos distintos (utilizados para classificação comercial⁵⁸) em relação aos seguintes grupos: homens negros, mulheres negras, homens brancos, mulheres brancas. As próprias autoras criaram o conjunto de dados de avaliação usado na pesquisa, chamado de *Pilot Parliaments Benchmark (PPB)*, que buscou representar de forma equilibrada diferentes tipos de pele e gêneros. Nas análises, a TRF teve o pior desempenho em relação às mulheres mais escuras, e melhor desempenho em relação a grupos mais claros do que em relação aos mais escuros.

A fim de avaliar o impacto produzido por seu estudo, Raji e Buolamwini (2019) utilizaram a mesma metodologia para realizar um segundo estudo, chamado “*Actionable auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products*”. O resultado foi que, em sete meses, todos os algoritmos testados anteriormente tiveram o seu desempenho melhorado em relação a mulheres e a

⁵⁷ “Taxa de aceitação falsa” mede a probabilidade de um sistema de segurança por reconhecimento facial aceitar incorretamente a identificação de um usuário de forma incorreta. Quanto menor for, melhor será a precisão do desempenho do algoritmo.

⁵⁸ Apesar dos sistemas avaliados serem comerciais, alguns deles são usados por órgãos de segurança pública dos EUA.

pessoas mais escuras. Apesar disso, dois algoritmos, dentre eles o *Rekognition* da Amazon, continuavam com taxas altas de imprecisão em relação a mulheres mais escuras.

Os dois estudos conduzidos por Buolamwini foram alvos de críticas, especialmente em relação aos testes aplicados ao algoritmo *Rekognition* da Amazon. Por exemplo, criticou-se o fato de tais estudos não terem levado em consideração que a Amazon recomenda o estabelecimento de um limite de confiança de 99% para que as autoridades policiais possam usar a sua TRF. Esse limite significa que o algoritmo só pode gerar resultado quando a pontuação de similaridade atinge um patamar de 99% ao realizar uma tarefa de identificação (AWS, 2020, p.159). No entanto, as pesquisadoras explicaram que na vida real nem sempre os utilizadores da TRF observam as recomendações feitas pelo fornecedor (no caso, a Amazon). Por isso, tentaram simular essa circunstância na pesquisa. E de fato, a única autoridade policial que a Amazon reconheceu como cliente (Departamento do Xerife do Condado de Washington, no Oregon) admitiu que não usa o *Rekognition* da maneira como a Amazon alega que recomenda. No entanto, ressalta-se que a autoridade não se baseia apenas na TRF para agir. A tecnologia, nesse caso, é apenas mais uma de suas ferramentas de investigação (Menegus, 2019).

Outro ponto alegado pela Amazon foi o de que a empresa não conseguiu reproduzir os resultados encontrados por Buolamwini. Segundo a gigante da tecnologia, não foi encontrada diferença significativa na precisão do *Rekognition* em relação a grupos demográficos distintos. Esse argumento, no entanto, está incorreto, pois uma avaliação algorítmica depende de forma direta do tipo de dados de avaliação utilizados (Introna e Nissenbaum, 2010, p.18). Explica-se: suponha um algoritmo que reconhece melhor homens brancos do que mulheres negras. Se tal algoritmo for avaliado por um conjunto de dados de avaliação no qual predominam imagens de homens brancos, por óbvio tal algoritmo apresentará elevado nível de precisão. O cenário seria outro, no entanto, se o conjunto de dados fosse mais diversificado. Como a Amazon não é transparente quanto aos seus conjuntos de dados de avaliação, os resultados alcançados com o conjunto de dados de avaliação criado por Buolamwini (o *Pilot Parliaments Benchmark – PPB*) são mais confiáveis.

Não bastasse, o estudo “*Face Recognition Vendor Test (FVRT): Part 3, Demographic Effects*”, publicado em 2019, comprovou que tanto algoritmos de identificação facial quanto algoritmos de verificação facial apresentavam taxas de falso-

positivo maiores para mulheres e pessoas asiáticas, africanas e nativo-americanas, do que para homens e pessoas do leste europeu.

Por fim, também é relevante destacar o estudo “*Amazon’s Face Recognition Falsely Matched 28 Members of Congress With Mugshots*”, conduzido pela União Americana pelas Liberdades Cívicas (*American Civil Liberties Union – ACLU*), no qual o sistema *Rekognition* da Amazon identificou incorretamente 28 congressistas norte-americanos como pessoas que já haviam sido presas. Dos 28 erros, 11 correspondiam a pessoas negras. Ou seja, 39,2% dos erros relacionavam-se a pessoas negras, sendo que estas representavam apenas 20% do Congresso (Snow, 2018).

Diante de todo o exposto, percebe-se que os estudos têm confirmado o enviesamento de raça e de gênero na TRF, devido à falta de precisão principalmente no reconhecimento de pessoas negras e mulheres. Como decorrência, tem-se que as pessoas desses grupos têm mais chances de não serem reconhecidas ou de serem identificadas incorretamente (e serem injustamente investigadas pelas autoridades). Além disso, os estudos também expuseram como o preconceito está relacionado ao país de origem do algoritmo e à baixa representatividade dos diferentes grupos demográficos no conjunto de dados de treinamento.

1.4.2.1. Causas do enviesamento

Se a execução de um algoritmo afetar de forma distinta diferentes grupos demográficos, significativamente privilegiando uns em detrimento de outros (por exemplo, homens brancos em detrimento de mulheres e pessoas negras), pode-se dizer que tal algoritmo é enviesado (Drozdowski *et al.*, 2020, p. 89).

Na tentativa de entender as origens dos vieses algorítmicos, Friedman e Nissenbaum (1996, p.330) propuseram a categorização dos vieses em: vieses preexistentes, vieses técnicos, e vieses emergentes. Na esteira desse pensamento, David Leslie (2020, p.4-13) explica que o viés nos algoritmos de reconhecimento facial é o resultado da combinação de dois fatores: (I) racismo histórico e privilégio masculino branco nas tecnologias de reprodução visual; (II) e o desenvolvimento de “*novas fontes de viés e discriminação decorrentes dos novos contextos sócio-técnicos de inovação*”

*digital e design*⁵⁹ e *produção algorítmica*”. Segundo o autor, no desenvolvimento da fotografia, ao longo dos anos, a “*composição química do filme foi projetada para ser a melhor para capturar a pele clara. O filme colorido era insensível à grande variedade de tipos de pele não branca e muitas vezes não mostrava os detalhes de rostos de pele mais escura*”. Essa padronização da pele clara como o tom de pele ideal resistiu com a chegada das câmeras digitais e foi transferida para os *softwares* de reconhecimento facial. Além disso, durante o desenvolvimento da TRF, novas condições de discriminação foram surgindo. Elas vão desde a “*forma como os preconceitos humanos se introduziram nas escolhas de design de modelos*” até à “*forma como os conjuntos de dados massivos que alimentam a aprendizagem de máquinas orientadas por dados tenderam a super-representar os grupos dominantes e a marginalizar as pessoas de cor*” (Leslie, 2020, p.13).

Danks e London (2017, p.4691), por sua vez, propuseram a categorização dos vieses em: (I) viés de treinamento de dados; (II) viés de foco algorítmico; (III) viés de processamento algorítmico; (IV) viés de contexto de transferência; e (V) viés de interpretação.

O viés de treinamento de dados está diretamente ligado à qualidade dos dados usados. Isso porque, a depender do conjunto de dados utilizado, o algoritmo é capaz de reconhecer um grupo demográfico, enquanto outros não. Como se sabe, no aprendizado de máquinas utiliza-se um conjunto de dados de treinamento a fim de ensinar o algoritmo a detectar padrões dos grupos demográficos. Uma vez treinado, o algoritmo mostra-se capaz de tirar novas conclusões a partir de novos dados inseridos (Surden, 2019, p.1312-1314). Dessa forma, um treinamento que ensina o algoritmo a detectar padrões de apenas um tipo de grupo demográfico, compromete o desempenho do algoritmo, na medida em que este não consegue reconhecer tão bem os padrões de outros grupos demográficos (Klare *et al.*, 2012, p.1791). A representatividade equilibrada de diversos grupos demográficos é, assim, de extrema importância para que um conjunto de dados seja considerado de qualidade (FRA, 2019b, p.27).

Além disso, também é importante destacar que a qualidade dos dados também diz respeito às imagens presentes na galeria⁶⁰. Explica-se: se houver na galeria mais fotos de

⁵⁹ Para maiores estudos sobre o racismo no design de software, ver: Sandvig, Christian; Hamilton, Kevin; Karahalios, Karrie; Langbort, Cedric. “When the Algorithm itself is a Racist: Diagnosing Ethical Harm in the Basic Components of Software,” *International Journal of Communication* 10(2016), p. 4872-4990.

⁶⁰ Definição de galeria: ver a nota de rodapé nº 4.

peessoas negras do que de pessoas brancas, a galeria torna-se uma fonte de reprodução de preconceitos históricos e pode levar a TRF a perpetuar tais preconceitos na sociedade, oprimindo cada vez mais grupos demográficos já marginalizados. Isso ocorre porque, se há mais fotos de pessoas negras na galeria, o algoritmo trabalha mais para fazer correspondências em relação especificamente a esse grupo demográfico, mesmo que isso resulte em falsos positivos.

Como se sabe, a sociedade é permeada de preconceitos direcionados a determinados grupos demográficos. O fato de pessoas negras terem sido presas mais do que pessoas brancas não significa que aquelas tenham uma maior tendência para o cometimento de crimes. Significa, como numerosos estudos já demonstraram, que as pessoas negras são em grande parte vítimas de uma discriminação institucionalizada na sociedade, que leva autoridades policiais e judiciárias a prenderem mais negros do que brancos (Xenidis e Senden, 2020, p.5-6). Com o agravante de que, se antes esse racismo institucionalizado era reproduzido de forma individualizada pelas autoridades, agora ele pode ser reproduzido de forma massificada pela TRF, que opera em vários lugares ao mesmo tempo.

Já o viés de foco algorítmico diz respeito à utilização de informações indesejáveis pelo algoritmo (Danks e London, 2017, p.4693). Um dos motivos para que isso ocorra é a escolha incorreta das características consideradas relevantes ao serem anotados os dados de treinamento do algoritmo de aprendizagem da TRF (Hao, 2019a). Por exemplo: caso um algoritmo tenha sido treinado para reconhecer padrões específicos de apenas um grupo demográfico, ele não será capaz de reconhecer os padrões de outros grupos.

Por sua vez, o viés de processamento algorítmico significa que um algoritmo concebido de forma defeituosa ou deficiente pode agir de forma discriminatória ao processar dados. Já o viés de contexto de transferência diz respeito à utilização do algoritmo dentro do propósito e contexto corretos (Danks e London, 2017, p.4694). Por exemplo: se um algoritmo for criado no país “A”, ele aprenderá a reconhecer os padrões dos rostos da raça predominante naquele país. Se tal algoritmo for utilizado em outro país “B”, no qual predominam outras raças, ele não saberá reconhecer com precisão os padrões dos rostos dessas raças (Phillips *et al.*, 2009, p.1-2).

Por último, destaca-se o viés de interpretação, que está relacionado à interpretação errada ou enviesada que um usuário dá ao resultado gerado pela TRF (Danks e London, 2017, p.4694). Se a pessoa responsável pela supervisão humana da TRF, por exemplo, não compreender como funciona a tecnologia, ou se for impregnada de preconceitos

enraizados, ela pode vir a interpretar incorretamente um resultado gerado pela TRF (McGregor *et al.*, 2019, p.338).

Diante do exposto, é necessário que os dados utilizados na TRF reflitam representações equilibradas dos diferentes grupos demográficos que serão impactados. Do contrário, tais dados simplesmente reproduzirão as hierarquias sociais, as relações de poder e as opressões que já compõem o contexto social do qual foram retirados. No entanto, não basta o controle correto dos dados. É necessário que também haja um controle sobre os designers e programadores, a fim de que eles hajam de forma ética e em favor dos direitos humanos desde a concepção da TRF (Leslie, 2020, p.15-17). É preciso que criadores de conjuntos de dados e designers de algoritmos priorizem *“a obtenção de capacidades técnicas para identificar, compreender e corrigir desequilíbrios potencialmente discriminatórios na representação das características demográficas e fenotípicas de seus dados”* (Leslie, 2020, p.15-17). Dessa forma, impede-se que os preconceitos sistêmicos se infiltrem na TRF.

1.4.2.2. Casos de discriminação de gênero na TRF

Autores como Gilliom e Monaham (2013, p.137) alertam que as tecnologias de videovigilância podem se tornar ferramentas de voyeurismo. Segundo os autores, estudos já demonstraram que operadores de salas de controle usavam o CFTV para acompanhar mulheres, focar em suas partes íntimas e imprimir as imagens. Nada garante que tais práticas não continuarão com as novas tecnologias, muito mais invasivas.

E o que é pior, um vazamento dos dados biométricos faciais pode fomentar *softwares de deepfake*⁶¹ que visam oprimir mulheres. Cita-se, por exemplo, o aplicativo *DeepNude* que visa inserir rostos femininos em corpos nus, criados digitalmente, de forma a parecer que a mulher está nua. Importante destacar que tal aplicativo não funciona com rostos de homens. Tal prática, por certo, causa incalculáveis danos emocionais, psicológicos e à imagem da mulher (Hao, 2019b). Pode ser usada para diversos fins como, por exemplo perseguir ativistas que lutam pelos direitos das mulheres⁶² e realizar pornô de vingança.

⁶¹ Refere-se a falsificações digitais hiper-realistas de vídeos, imagens e sons humanos (Chesney e Citron, 2019, p.1757). É um tipo de mídia sintética, gerada ou manipulada por IA (Schick, 2020, p.8).

⁶² Ver, por exemplo, o caso da jornalista Rana Ayubb, que teve o seu rosto inserido em um vídeo pornô e divulgado nas redes sociais (por grupos de extrema direita), como represália aos protestos que a jornalista fez em razão do estupro e morte de uma garota de oito anos de idade na Índia. Disponível em:

A Discriminação de gênero algorítmica também está presente nas redes sociais. Por exemplo, apesar do Facebook ter anunciado em 2014 que a sua TRF tinha 97% de acurácia, a precisão da tecnologia não era tão precisa quando relacionada à análise de indivíduos negros e mulheres. O motivo é que a base de dados usada para treinar o algoritmo era composta de 77% de homens e 80% de pessoas com a pele clara (Buolamwini, 2019).

Outro caso de imprecisão relacionado a rostos femininos é o da empresa Idemia, que fornece a TRF usada para controle de migrações nos EUA, Austrália e França. Estudos demonstraram que quando se tratava de reconhecer rostos de mulheres negras, a taxa de erro aumentava em dez vezes em relação às mulheres brancas. Isso explicitou que as opressões geradas pela TRF são interseccionais (Simonite, 2019). Por interseccionalidade entende-se a forma pela qual sistemas discriminatórios (como a opressão de classe, a discriminação de gênero e de raça) atuam conjuntamente para oprimir e subjugar certos grupos (Crenshaw, 2002, p.177).

1.4.2.3. Casos de discriminação de raça na TRF

Como dito anteriormente, a TRF aplicada para fins de policiamento e vigilância, se enviesada, pode reproduzir preconceitos e opressões pré-existentes na sociedade. É um “círculo vicioso”: se o algoritmo se basear em suspeitas infundadas e pré-concebidas, a tecnologia intensificará a vigilância sobre os grupos demográficos já historicamente marginalizados, estigmatizando-os ainda mais.

Destaca-se o caso do *software Faception*⁶³, no qual a TRF desenvolvida por uma empresa israelense foi capaz de classificar como terroristas 9 das 11 pessoas que realmente praticaram atos terroristas em 2015, em Paris. Preocupante foi o fato de que o algoritmo considerou a informação “ascendência árabe” como suficiente para identificar tais pessoas como terroristas (Adee, 2016).

Outros casos emblemáticos foram: (I) o do algoritmo de reconhecimento facial da *Google Photos*, que em 2015 identificou duas pessoas negras como gorilas⁶⁴. Chama

<https://www.indiatoday.in/trending-news/story/journalist-rana-ayyub-deepfake-porn-1393423-2018-11-21>.

⁶³ Nesse sentido, ver: <https://www.dailymail.co.uk/news/article-3606811/Can-spot-terrorist-just-looking-face-Israeli-company-claims-predict-paedophiles-geniuses-ace-poker-players-analysing-features.html>.

⁶⁴ Ver: “A major flaw in Google’s algorithm allegedly tagged two black people’s faces with the word ‘gorillas’”. Disponível em: <https://www.businessinsider.com/google-tags-black-people-as-gorillas-2015-7>.

atenção a solução apresentada pela empresa, que demonstra claramente como o racismo está enraizado não só nos dados, mas nas próprias escolhas dos programadores, designers, engenheiros e dirigentes de empresas: a Google simplesmente optou, após 3 anos do ocorrido, por remover as imagens de primatas de sua base de dados, bem como excluir de seu *software* os termos relacionados aos animais (Hern, 2019); (II) o do algoritmo das câmeras fotográficas da Nikon, que incorretamente indicava que pessoas asiáticas haviam piscado nas fotos quando, na verdade, estavam de olhos abertos (Rose, 2010).

Há ainda o risco de que a TRF, ao ser usada para policiamento e vigilância, influencie o comportamento do policial e crie um viés, fazendo com que o agente de segurança aborde o suspeito de forma violenta. Como dito, falsos positivos acontecem desproporcionalmente em relação às pessoas negras, grupo já socialmente marginalizado em diversos países. Abordagens policiais enviesadas aumentarão ainda mais a opressão sobre este grupo demográfico.

Segundo estudo feito pela Rede de Observatórios da Segurança (Ramos, 2019, p. 68-69), de março a outubro de 2019, 151 pessoas foram presas no Brasil com o uso de reconhecimento facial. Dessas 151, 90% eram negras, presas por crimes de baixo potencial ofensivo. O grande problema do policiamento preditivo, com sua natureza antecipatória, é que os indivíduos podem ser presos, investigados e discriminados não por causa do que fizeram, mas por causa do que conclusões baseadas em dados dizem que eles poderiam fazer. Se o uso da TRF pelas autoridades policiais em espaços públicos não for tratado com o devido cuidado, existe o risco de que grupos demográficos inteiros sejam vigiados de forma intensiva e extensiva com o simples intuito de se punir e controlar tais grupos (Norris, 2003, p.9).

2. A ATUAL LEGISLAÇÃO EUROPEIA DE NÃO DISCRIMINAÇÃO

2.1. AS PRINCIPAIS POSIÇÕES EXISTENTES ATUALMENTE SOBRE O USO DE TRF

Antes de analisarmos propriamente o quadro regulatório de não discriminação existente na UE, é necessário entender como andam as discussões sobre usos de TRF e quais os principais posicionamentos adotados pelas instituições. Em resumo, é possível destacar atualmente 4 posições principais em relação ao uso de sistemas de identificação biométrica remota: promoção ativa; moratória; proibição total; e promoção com salvaguardas. De uma forma geral, a Comissão Europeia e o Conselho Europeu apoiam o desenvolvimento desses sistemas. Por exemplo, documentos da Comissão como o “*White Paper on AI*” e a recente Proposta de Regulamentação de IA (*AI Act*) procuram transformar a UE em um centro global para IA confiável, no qual promove-se a inovação ao mesmo tempo em que se garante a proteção dos direitos humanos. Nesse contexto, a TRF assume importância na promoção de segurança nas cidades inteligentes (Ragazzi *et al.*, 2021, p.56).

Por outro lado, *GDPR (General Data Protection Regulation)* e *LED (Law Enforcement Directive)* criaram a base para contestações sobre o uso de TRF em espaços públicos. Campanhas online⁶⁵ foram lançadas para alertar a população sobre os riscos da TRF e para pressionar a Comissão Europeia, ao mesmo tempo em que relatórios⁶⁶ importantes foram publicados. Cada um desses posicionamentos expõe diferentes questões de privacidade e direitos humanos, bem como propõe diferentes estratégias de mitigação, como a moratória, a proibição total e a promoção com salvaguardas (Ragazzi *et al.*, 2021, p.56).

Dentre os que propõem uma promoção ativa da TRF na UE, destacam-se: diferentes governos locais e nacionais, que buscam na tecnologia soluções para desenvolver serviços públicos e cidades inteligentes e seguras; bem como autoridades policiais, que enxergam a tecnologia como um instrumento eficiente de investigação criminal. Apesar de representarem uma voz importante, os defensores da promoção ativa

⁶⁵ Por exemplo: #ReclaimYourFace.

⁶⁶ Por exemplo: FRA 2018; EDRi 2020; FRA, 2019b; Fernandez *et. al.*, 2020; Gonzalez Fuster, 2020; CNIL 2019.

da TRF não encontram eco nas principais discussões da UE (Ragazzi *et al.*, 2021, p.56-57).

Já em relação à moratória, é possível destacar que a Autoridade Europeia para a Proteção de Dados (*European Data Protection Supervisor – EDPS*) defendia tal posicionamento num primeiro momento, pois entendia ser necessário um debate público e democrático antes que tecnologias de identificação biométrica remota fossem implantadas em locais públicos (EDPS, 2020). No entanto, após verem que os seus apelos por moratória não foram abordados pela Comissão Europeia no Projeto de Lei de IA, a EDPS mudou de posição e juntou-se ao Comité Europeu para a Proteção de Dados (*European Data Protection Board – EDPB*) para defenderem a proibição geral do uso de sistemas biométricos remotos em espaços públicos. Tais órgãos entendem que o uso desses sistemas pode violar profundamente a privacidade e os direitos fundamentais das pessoas (EDPB, 2021).

Por sua vez, o Parlamento Europeu aprovou uma resolução sobre IA em janeiro de 2021, na qual propõe à Comissão Europeia uma avaliação sobre os usos de TRF e, a depender dos resultados, uma consequente moratória até que a tecnologia amadureça e consiga estar em conformidade com os direitos fundamentais. Visa-se, com isso, evitar resultados enviesados e discriminatórios e, ao mesmo tempo, assegurar que a tecnologia seja usada de forma proporcional (Parlamento Europeu, 2021).

Dentre os que pedem uma proibição total do uso de sistemas de identificação biométrica remota, pode-se destacar organizações não governamentais (ONGs) nacionais e da UE, bem como partidos políticos da UE. Tais entidades argumentam que o processamento em massa de dados biométricos em espaços públicos cria um sério risco de vigilância em massa que infringe os direitos fundamentais. A relação de ONGs inclui, por exemplo: *EDRi*⁶⁷; *La Quadrature du Net*⁶⁸; *Algorithm Watch*⁶⁹. Já entre os partidos, destaca-se os Verdes, que junto com outros Eurodeputados (em sua maioria da esquerda), publicaram uma carta direcionada à Comissão Europeia em abril de 2021, na qual expressaram suas preocupações relacionadas à Proposta de Regulamento de IA, em

⁶⁷ Ver: EDRi. 2020. Ban Biometric Mass Surveillance!. Disponível: <https://edri.org/our-work/blog-ban-biometric-mass-surveillance/>.

⁶⁸ Ver: La Quadrature du Net. 2020. Our Legal Action against the Use of Facial Recognition by the French Police. Disponível: <https://www.laquadrature.net/en/2020/09/21/our-legal-action-against-the-use-of-facial-recognition-by-the-french-police/>.

⁶⁹ Ver: Algorithm Watch. 2020. Automating Society Report 2020. Disponível: <https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/12/Automating-Society-Report-2020.pdf>.

especial em relação às possíveis isenções presentes no art. 4 (Ragazzi *et al.*, 2021, p.60-61).

Ao mesmo tempo, foram lançadas campanhas⁷⁰ que buscam conscientizar a população e pressionar os governos a proibir o uso de TRF. Não bastasse, um grupo formado por *Access Now*, *Human Rights Watch*, *European Digital Rights (EDRI)*, Anistia Internacional, *Internet Freedom Foundation (IFF)* e Instituto Brasileiro de Defesa do Consumidor (IDEC), lançou uma carta aberta (assinada por quase 200 organizações), na qual pediram a proibição total do uso de TRF que permite vigilância em massa e discriminatória (Ragazzi *et al.*, 2021, p.60-61).

Por fim, aqueles que são a favor da promoção da TRF com salvaguardas argumentam que a implantação da tecnologia deve ser rigorosamente monitorada devido aos potenciais riscos que elas representam, tal como o perigo de criminalizar e estigmatizar grupos demográficos que já sofrem discriminação. Nesta categoria, podem ser destacados: Conselho da UE; Comissão da UE; a Agência dos Direitos Fundamentais (FRA); Autoridades de Proteção de Dados nacionais como a CNIL; o Conselho da Europa (CoE); alguns partidos políticos da UE; e alguns tribunais (Ragazzi *et al.*, 2021, p.57).

O Conselho da UE lançou o “O Guia Prático sobre o Uso de Dados Pessoais no Setor de Polícia”, que complementa a Convenção 108+ e enfatiza que, devido ao alto risco que sistemas biométricos automatizados representam para os direitos individuais, salvaguardas específicas devem ser implementadas em tais sistemas, com o apoio e orientação de Autoridades de Proteção de Dados (Conselho da Europa, 2018). Além disso, no documento “Diretrizes sobre Reconhecimento Facial”, o Conselho requer que sejam estabelecidas condições para o uso de TRF por autoridades policiais bem como critérios claros para a criação de bancos de dados, à luz de propósitos específicos, legítimos e explícitos (Comitê Consultivo da Convenção para a proteção dos indivíduos com relação ao processamento automático de dados pessoais, 2021, p.5-6).

Em relação à FRA, destaca-se o seu relatório de 2019 sobre TRF, que exige uma abordagem moderada em relação ao uso da tecnologia. De acordo com a FRA, o uso de TRF leva a implicações sobre os direitos fundamentais que variam consideravelmente dependendo do propósito, escopo e contexto. Por esse motivo, defende que deve haver uma compreensão ampla sobre como funciona exatamente a TRF e qual o impacto

⁷⁰ Por exemplo: *#ReclaimYourFace*; *Ban the Scan*, da Anistia Internacional; *Ban Automated Recognition of Gender and Sexual Orientation*, da *Access Now*; *Project Panopticon*, do rastreador *Panoptic*, baseado na Índia.

causado sobre os direitos humanos. A entidade ainda destaca uma série de questões baseadas na estrutura de direitos fundamentais da UE, bem como na legislação de proteção de dados da UE. Ao final, defende a necessidade de uma estrutura legal clara e detalhada sobre o uso da TRF, bem como o monitoramento próximo e uma avaliação completa e contínua do impacto de cada implantação da tecnologia (FRA, 2019b, p.18-34).

A Autoridade de Proteção de Dados francesa (*Commission Nationale de l'Informatique et des Libertés – CNIL*), assume uma posição parecida em seu relatório “*Facial Recognition. For a debate living up to the challenges*”. A entidade alerta para a necessidade de segurança dos dados biométricos (devido à sua natureza sensível) e o potencial de vigilância da TRF sobre a sociedade (CNIL, 2019, p.6). Em suas recomendações, a CNIL reconhece a legitimidade e a proporcionalidade de alguns usos de TRF, ao mesmo tempo em que exige uma vigilância especial sobre a tecnologia, inclusive no que diz respeito aos “projetos-piloto” (CNIL, 2019, p.10).

Por sua vez, a Comissão Europeia, no comunicado “*Shaping Europe’s Digital Futures*” (que acompanha o *White Paper on AI*), apresentou suas diretrizes e estratégias para adequar a Europa à era digital. Dentre tais estratégias, destaca-se a Estratégia Europeia de Dados (*European Data Strategy – EDS*), que visa tornar a Europa em líder global na economia orientada por dados e, ao mesmo tempo, assegurar a soberania tecnológica da Europa, desbloqueando o potencial de novas tecnologias como a IA. O financiamento dessas iniciativas será feito pelos programas “*Connecting Europe Facility 2*”, “*Horizon Europe*”, e “*Digital Europe Programme (DEP)*” (Newsroom, 2020).

É possível perceber, portanto, que o uso da TRF na UE avança a passos largos, apesar dos diferentes receios e posicionamentos existentes nas instituições e na sociedade civil. Isso indica que a UE precisará realizar rapidamente uma abordagem regulatória harmonizada se quiser promover o seu desenvolvimento tecnológico sem comprometer os seus valores primordiais e direitos fundamentais (dentre eles, o da não discriminação, previsto no art. 21 da Carta de Direitos Fundamentais da UE). Caso não o faça, a UE arrisca permitir a proliferação de legislações nacionais diferentes e contraditórias, que por sua vez poderão impedir a livre circulação de bens e serviços relacionados à TRF; bem como impedir o estabelecimento de um elevado nível de proteção dos direitos fundamentais comum a todos os países membros da UE, no que diz respeito ao uso da TRF por autoridades policiais.

Algumas instituições, percebendo os riscos reais aos direitos fundamentais, começaram a fazer alertas. O Parlamento Europeu (2017), por exemplo, trouxe ao debate questões de direitos fundamentais relacionadas a Big Data e decisões automatizadas. Por sua vez, o Conselho da UE (2020) alertou para questões de direitos fundamentais relacionadas ao uso de dados enviesados no aprendizado de máquinas. Já a Comissão Europeia publicou a sua Proposta de Regulamento de IA (*AI Act*), que tem um de seus focos na minimização dos riscos relacionados aos direitos fundamentais. Com efeito, nos Considerandos 1 e 2 da Proposta, a Comissão enfatiza ser necessário assegurar aos direitos fundamentais um nível de proteção elevado e coerente em toda a UE a fim de evitar que divergências fragmentem o mercado interno e reduzam a segurança jurídica para os operadores que desenvolvem ou utilizam sistemas de IA.

Diante de todo o exposto, é possível perceber que diferentes órgãos e países da UE entendem que a TRF pode contribuir para tornar as investigações criminais mais eficientes. No entanto, devido ao elevado risco que a tecnologia tem de se tornar enviesada e perpetuar opressões sobre grupos demográficos já oprimidos e marginalizados, é crucial que ocorra na UE uma abordagem regulatória focada na não discriminação. A definição de tal abordagem, no entanto, passa, antes, pela análise do atual ordenamento jurídico europeu no que diz respeito à não discriminação, a fim de entendermos se tal ordenamento é suficiente.

2.2. A ATUAL LEGISLAÇÃO ANTIDISCRIMINAÇÃO DA UE E SUA APLICAÇÃO À TRF

Em primeiro lugar, é preciso destacar que a discriminação ocorre quando um indivíduo ou um grupo é tratado menos favoravelmente do que outro é, foi ou seria tratado em uma situação comparável, devido a uma característica pessoal percebida ou real (como, por exemplo, gênero e raça) (FRA, 2019b, p.27).

A não discriminação é um direito humano fundamental e está consagrada em diversas normas internacionais de direitos humanos como, por exemplo, no art. 7º da Declaração Universal dos Direitos Humanos e no art. 26 do Pacto Internacional dos Direitos Civis e Políticos, que proíbem a discriminação em geral; na Convenção sobre a Eliminação de Todas as Formas de Discriminação contra as Mulheres; e na Convenção Internacional sobre a Eliminação de Todas as Formas de Discriminação Racial.

No ordenamento jurídico europeu especificamente, as regras de não discriminação estão espalhadas em uma estrutura multinível de proteção, estando consagradas tanto na legislação primária quanto na secundária. Tais regras fazem parte do conjunto de normas (juntamente com as normas de privacidade, proteção de dados e a Proposta de Regulação da IA) que estabelecem parâmetros críticos para o desenvolvimento e uso de TRF, e se aplicam amplamente aos operadores públicos e privados de sistemas TRF.

No nível primário de proteção, a não-discriminação está presente no art. 2º do Tratado da UE, e no art. 10º do Tratado sobre o Funcionamento da UE, como um dos valores fundantes do bloco europeu. Também é possível destacar a Carta dos Direitos Fundamentais da UE, que em seu art. 20 estabelece que todos são iguais perante a lei, e no art. 21 proíbe qualquer discriminação por qualquer motivo. Inclusive, devido a sua formulação aberta e amplo alcance, o art. 21 parece conceitualmente adequado para tratar de casos de discriminação algorítmica. Também é importante destacar que, apesar do art. 51(1) determinar que as disposições da Carta têm por destinatários as instituições, órgãos e organismos da União, bem como os Estados-Membros (quando apliquem o direito da União), as relações entre as partes privadas também podem ser afetadas (efeito horizontal) (Frantziou, 2020, p.225-231).

Além disso, a proibição expressa no art. 21 da Carta assume especial importância por três motivos. Primeiro, porque influencia proibições correspondentes em outros dispositivos jurídicos, tais como o art. 14 e o Protocolo nº 12, ambos da Convenção Europeia dos Direitos do Homem; segundo, porque a proibição presente na Carta é ainda mais ampla do que nestes dispositivos e na legislação de nível secundário, uma vez que estabelece uma lista não exaustiva e aberta de novos fundamentos que podem ser protegidos; terceiro, porque a proibição de discriminação prevista no art. 21 da Carta é um direito aplicável independentemente de qualquer outro direito. Por exemplo, ao contrário do art. 14 da Convenção Europeia dos Direitos do Homem, que estabelece que a discriminação é proibida somente quando relacionada ao exercício de outros direitos previstos na Convenção, o art. 21 da Carta estabelece que a não discriminação independe da violação de qualquer outro direito.

Outro ponto relevante é que o direito à não discriminação, assim como demais direitos fundamentais, não é absoluto. Sendo assim, pode estar sujeito a limitações tanto no âmbito da Carta de Direitos Fundamentais da UE, quanto no da Convenção Europeia dos Direitos do Homem. O art. 52(1) da Carta e o arts. 8(2), 9(2), 10(2) e 11(2) da Convenção, por exemplo, estabelecem requisitos necessários para que um direito

fundamental possa ser flexibilizado. Nesse sentido, para que o direito à não discriminação possa ser restringido, é preciso que haja previsão na lei; que o objetivo perseguido seja legítimo; e que a restrição seja necessária, proporcional e respeite a essência do direito.

Por sua vez, no nível secundário, temos a legislação que busca efetivar as normas primárias de direitos fundamentais, governando com mais detalhes a criação e a implementação de tecnologias emergentes. Como exemplo, é possível citar a Diretiva de Igualdade de Gênero 2006/54/EC, Diretiva de Igualdade Racial 2000/43/EC, Diretiva de Bens e Serviços de Gênero 2004/113/EC, e a Diretiva de Igualdade no Emprego 2000/78/EC. A proibição de discriminação nessas diretivas tem um alcance menor do que no art. 21 da Carta e se estende a 3 áreas políticas: acesso a bens e serviços, emprego e sistema de bem-estar (Hacker, 2018, p.1155). Além disso, uma diretiva anti-discriminação horizontal, que visa preencher as lacunas entre as diretivas acima citadas, encontra-se bloqueada há anos no Conselho da UE⁷¹.

Outra questão importante a se ressaltar é que a legislação da UE, como um todo, protege tanto contra a discriminação direta quanto contra a indireta. De acordo com o art. 2º da Diretiva 2000/43/CE, do Conselho, discriminação direta é a prática em que um indivíduo (ou grupo) é explicitamente tratado de forma menos favorável do que outro em situação comparável, devido a uma característica particular sua (que é protegida). Por sua vez, discriminação indireta é quando uma regra, critério, ou prática supostamente neutra, baseia-se na característica particular de um indivíduo (ou grupo) para colocá-lo em situação de significativa desvantagem em comparação com outras pessoas (ou grupos). Pode-se dizer que a discriminação indireta é mais velada e sutil, em comparação com a direta.

É importante dizer que tanto a discriminação direta quanto a indireta podem ser justificadas, caso respeitem os requisitos destacados anteriormente⁷². Neste caso, um tratamento diferenciado justificado não configurará discriminação (FRA, 2018a, p.103). No entanto, a discriminação indireta pode ser mais facilmente justificada do que a direta, na medida em que altera o foco do “tratamento diferente” para “efeitos diferentes”. Não há um tratamento de distinção que incide diretamente sobre o indivíduo, mas sim os efeitos de um tratamento aparentemente neutro. Ou seja, o fato da discriminação indireta resultar de um tratamento aparentemente neutro e não afetar o indivíduo diretamente, faz

⁷¹ Para mais informações: <https://www.europarl.europa.eu/legislative-train/theme-area-of-justice-and-fundamental-rights/file-anti-discrimination-directive?sid=03-2019>.

⁷² Previsão em lei; objetivo legítimo; e restrição necessária, proporcional, que respeita a essência do direito.

com que tal tipo de discriminação possa ser justificada de forma mais flexível (desde que respeitados os requisitos estabelecidos). A discriminação direta, por outro lado, só pode ser justificada sob condições muito estritas e deve perseguir objetivos específicos (FRA, 2018a, p.103-123)⁷³.

No âmbito específico do aprendizado de máquinas, nota-se que a discriminação direta prevalece menos do que a indireta (Wachter *et al.* 2020, p.41-45). Isso porque a discriminação indireta parece ser mais adequada para englobar um conjunto de resultados algorítmicos que supostamente deveriam ser neutros, mas que de fato são discriminatórios (Hacker, 2018, p.1151). Ora, é mais provável que um resultado algorítmico discriminatório decorra de um conjunto de dados (de treinamento, por exemplo) enviesados, do que de uma ação intencional do desenvolvedor de introduzir um viés no sistema de IA (Xenidis e Senden, 2020, p.19). Além disso, a discriminação decorrente de um resultado algorítmico se aproxima da discriminação indireta por também ser sutil, de difícil constatação, e ter o potencial de atingir um maior número de pessoas, como grupos demográficos inteiros (Xenidis e Senden, 2020, p.21).

É importante destacar que o fato da TRF usar dados biométricos faciais (que por sua vez são dados pessoais⁷⁴) atrai também a aplicação das normas relativas à proteção de dados, tais como o Regulamento Geral de Proteção de Dados (*General Data Protection Regulation – GDPR*), a Diretiva (UE) 2016/680 do Parlamento e do Conselho (*Law Enforcement Directive – LED*), e a Convenção 108+. Assim, apesar do foco de tais normas ser a proteção da privacidade (art. 7º da Carta de Direitos Fundamentais da UE) e dos dados (art. 8º da Carta de Direitos Fundamentais da UE), elas contribuem substancialmente para a proteção contra a discriminação no uso de TRF pelas autoridades policiais. Nos termos do art. 2(1), GDPR e art. 2, LED, tanto o GDPR quanto a LED são aplicáveis ao tratamento automatizado de dados pessoais e ao tratamento manual que faz parte de um sistema de arquivamento. No entanto, enquanto a LED se aplica ao processamento nos países membros de Schengen, o GDPR se aplica ao processamento no Espaço Econômico Europeu. Outra diferença é que, nos termos dos Considerandos 11 e 12 da LED e Considerando 19 do GDPR, a LED é um regime mais específico do que o

⁷³ Podem se basear, por exemplo: em exceções particulares à discriminação por idade (art. 6 da Diretiva de Igualdade no Emprego 2000/78/CE); em exceções em relação a instituições religiosas (art. 4(2) da Diretiva de Igualdade no Emprego 2000/78/CE); em requisitos profissionais genuínos (art. 14(2) da Diretiva de Igualdade de Gênero (reformulada) 2006/54/CE, art. 4 da Diretiva de Igualdade Racial 2000/43/CE, art. 4(1) da Diretiva de Igualdade no Emprego 2000/78/CE).

⁷⁴ De acordo com o art. 4(14) do GDPR e art. 3(13) da Diretiva (UE) 2016/680.

GDPR, sendo aplicável quando as autoridades públicas processam dados pessoais para a prevenção, investigação, detecção de processos penais.

Pois bem, analisando a legislação, percebe-se do art. 9(1) do GDPR, art. 10 da LED, e art. 6(1) da Convenção 108+, que os dados biométricos, por serem sensíveis, recebem um nível maior de proteção. Dessa forma, conforme estabelece o art. 10 da LED, o processamento de dados biométricos faciais por uma TRF operada por autoridades policiais só é permitido se for estritamente necessário e se estiver sujeito a salvaguardas adequadas à proteção dos direitos e liberdades do titular dos dados. O art. 6(2) da Convenção 108+, por sua vez, corrobora tais exigências e ainda determina explicitamente que as salvaguardas devem proteger principalmente contra o risco de discriminação. Destaca-se, ainda, o Considerando 51 e o art. 19(1), ambos da LED. Através de uma interpretação sistemática dos dois dispositivos, é possível concluir que a legislação da UE obriga que os controladores adotem todas as medidas técnicas e organizacionais apropriadas a fim de assegurar que o processamento dos dados pessoais (dentre eles, os biométricos faciais) esteja de acordo com a lei e minimize ao máximo o risco de discriminação.

Além disso, o art. 5 do GDPR e o art. 4 da LED estabelecem que qualquer tratamento de dados pessoais (tais como as imagens faciais) deve seguir os seguintes princípios essenciais: ser lícito, justo e transparente; seguir um objetivo específico, explícito e legítimo; cumprir os requisitos de minimização de dados, exatidão de dados, limitação do armazenamento, segurança de dados e responsabilidade. Será apresentado abaixo uma análise sucinta de como esses princípios podem ajudar a impedir ou minimizar os riscos de violação ao direito fundamental de não discriminação.

Destaca-se, de início, que o art. 5(1)(a), GDPR, o art. 4(1)(a), LED, e o art. 5(4)(a), Convenção 108+ protegem contra a discriminação ao estabelecerem que os dados pessoais devem ser processados de forma justa, lícita e transparente. Nessa esteira, o Considerando 40, GDPR e o Considerando 35, LED estabelecem que o processamento deve satisfazer os requisitos de bases legais específicas para ser considerado legal. O art. 6, GDPR e as transposições nacionais do art. 8, LED podem ser usados como bases legais para a videovigilância. No entanto, se forem processadas categorias especiais de dados (como dados biométricos faciais), o processador deve satisfazer também os requisitos mais rígidos do art. 9, GDPR e art. 10, LED. Além disso, os requisitos do art. 22(2) e (4), GDPR ou do art. 11(1) e (2), LED precisam ser satisfeitos para que possam ser tomadas decisões baseadas unicamente no processamento automatizado. No contexto específico

de TRF, tem sido comum o uso do art. 9(2)(g), GDPR como base jurídica. Tal dispositivo permite o tratamento de dados pessoais com base na legislação da União ou dos Estados-Membros, se for necessário por razões de interesse público substancial.

Cabe destacar ainda que a legislação de proteção de dados exige transparência no processamento de dados pessoais, como bem destaca o 5(4)(a) da Convenção 108+. Isso significa que os desenvolvedores e utilizadores de uma TRF devem fornecer informações sobre todas as etapas do processo de tomada de decisão da IA, a fim de que as autoridades de supervisão possam investigar, caso necessário, se tal processamento possui efeitos discriminatórios (Zuiderveen Borgesius, 2018, p.38). De acordo com o Considerando 39, GDPR, o princípio da transparência exige que as informações ou comunicações relacionadas com o tratamento de dados pessoais sejam de fácil acesso e compreensão, e formuladas numa linguagem clara e simples. No entanto, o Considerando 26, LED determina que o princípio da transparência, por si só, não impede que autoridades de aplicação da lei exerçam atividades tais como investigações encobertas ou videovigilância. Com efeito, o art. 13(3), LED, estabelece que os Estados-Membros podem adotar medidas excepcionais que limitem a transparência aos titulares dos dados, a fim de se evitar que as investigações criminais em andamento sejam prejudicadas, ou para proteger a segurança pública, a segurança nacional e os direitos e liberdades de terceiros. Tais exceções, no entanto, precisam ser necessárias, proporcionais, ter em conta os direitos fundamentais dos indivíduos e serem fortemente justificadas.

Por sua vez, o princípio da justiça (ou imparcialidade) exige que o titular não tenha os seus dados pessoais processados de forma injustificadamente prejudicial, discriminatória, inesperada ou enganosa (EDPB, 2020a, p.17). Tal princípio visa compensar a vulnerabilidade do titular de dados frente ao controlador, e deve ser sempre levado em conta pelos desenvolvedores ao projetar a TRF e pelos operadores ao conceberem planos de implementação (Madiega e Mildebrath, 2021, p.14).

Já o princípio da limitação da finalidade, previsto no art. 5(1)(b), GDPR e art. 4(1)(b), LED, determina que o processamento dos dados pessoais deve ser voltado para uma finalidade precisamente definida, explícita e legítima, e que só sob condições estritas tais dados podem ser usados para uma finalidade incompatível com a finalidade inicialmente designada. Como a TRF possui grande potencial de ter a sua função desvirtuada, é necessário que sejam estabelecidas salvaguardas no processamento realizado por ela, a fim de que a sua utilização seja impedida para fins não autorizados

como, por exemplo, perseguir e discriminar grupos demográficos específicos (FRA, 2017, p.21-23).

O art. 5(1)(c), GDPR e art. 4(1)(c), LED estabelecem o princípio da minimização dos dados, que determina que os dados pessoais sejam adequados, pertinentes e limitados ao mínimo necessário relativamente às finalidades para as quais são tratados. Tal princípio é essencial quando se trata de TRF, pois como a tecnologia nunca foi totalmente precisa, ela necessita ser alimentada com um grande volume de dados sensíveis a fim de aperfeiçoar o algoritmo. O perigo é que, sob a desculpa de tornar o algoritmo preciso e evitar viés, sejam coletados indefinidamente dados biométricos faciais de certos grupos demográficos. Mais do que a quantidade, é preciso que haja qualidade nos dados (EDPS, 2019).

O art. 5(1)(e), GDPR e art. 4(1)(e), LED, por sua vez, preveem o princípio da limitação de armazenamento, segundo o qual os dados pessoais devem ser conservados de forma a permitir a identificação dos titulares dos dados apenas durante o período necessário para as finalidades para as quais são tratados. Quanto maior for o tempo de armazenamento dos dados, maior deve ser a justificação sobre a legitimidade da finalidade e a necessidade do armazenamento (EDPB, 2020b, p.28). Além disso, é preciso que os controladores de dados assegurem que os dados extraídos de uma imagem digital para construir um modelo (*template*) não sejam excessivos e contenham apenas as informações necessárias para a finalidade especificada. Dessa forma, evita-se um possível processamento posterior, que pode vir a ser discriminatório (EDPB, 2020b, p.21).

Já o art. 5 (1)(d), GDPR, o art. 4(1)(d), LED, e o art. 5(4)(d) da Convenção 108+ exigem que os dados pessoais processados sejam precisos e atualizados (princípio da precisão dos dados), o que impõe aos desenvolvedores e usuários de TRF o dever de zelar para que a tecnologia não apresente diferenças de precisão devido às características dos diferentes grupos demográficos (por exemplo, cor da pele, idade e gênero). E caso ocorra essa diferenciação, que sejam tomadas todas as medidas necessárias, para eliminá-la o quanto antes, evitando assim a discriminação involuntária. Nesse sentido, tendo em vista que a baixa qualidade das imagens aumenta o número de erros, é preciso que os controladores ou processadores de dados verifiquem a qualidade das imagens e modelos biométricos presentes nas “listas de vigilância” a fim de evitar falsas correspondências. Também é preciso que os algoritmos sejam treinados com conjuntos de dados que possuam maior representatividade dos diferentes grupos demográficos e que não possuam vieses ocultos (Madiaga e Mildebrath, 2021, p.16).

O princípio da segurança de dados, previsto no art. 5(1)(f), GDPR e art. 4(1)(f), LED, estabelece que deve ser garantida a segurança adequada dos dados pessoais durante o seu tratamento, incluindo a proteção contra o seu tratamento não autorizado ou ilícito e contra a sua perda, destruição ou danificação accidental, adotando as medidas técnicas ou organizacionais apropriadas. Por sua vez, o art. 32, GDPR e art. 29, LED estabelecem a necessidade de tanto o controlador quanto o processador implementarem medidas técnicas e organizacionais que impeçam os dados pessoais de serem divulgados ou acessados por terceiros não autorizados. Nesse sentido, a fim de se proteger adequadamente todas as etapas do processamento de dados, devem ser tomadas medidas como, por exemplo, a criptografia de dados biométricos; a compartimentação dos dados durante a transmissão e armazenamento; o armazenamento de modelos biométricos e dados de identidade em bancos de dados diferentes; a associação de um código de integridade aos dados, como a assinatura *hash*; proibição de acesso externo aos dados biométricos; integração de medidas adequadas para a detecção de fraude; etc (EDPB, 2020b, p.21).

Também importante é o princípio da responsabilidade, previsto no art. 5(2), GDPR e art. 4(4), LED, que estabelece que o controlador de dados é responsável pelo cumprimento (e pela demonstração do cumprimento) dos princípios de processamento de dados pessoais. Na execução dessa tarefa, é essencial que os controladores usem as ferramentas previstas nos seguintes dispositivos: art. 24, GDPR e art. 19, LED (implementação de medidas técnicas e organizacionais apropriadas); art. 33(5), GDPR e art. 30(5), LED (documentação das violações de dados); art. 30, GDPR e art. 24, LED (registro das atividades de processamento).

Ainda no contexto do princípio da responsabilidade, é importante destacar os arts. 35 e 36, GDPR, os arts. 20(1), 27 e 28, LED, e o art. 10(2) (3) da Convenção 108+, que estabelecem o princípio da privacidade *by design* e a obrigação de realizar Avaliação de Impacto da Proteção de Dados (*Data Protection Impact Assessment – DPIA*). Nesse sentido, tanto o princípio quanto a avaliação requerem que os responsáveis pelo tratamento (e os processadores, quando aplicável) examinem os riscos e possíveis impactos do tratamento de dados pretendido e estabeleçam medidas técnicas e organizacionais antes do início de tal tratamento, de forma a evitar ou minimizar o risco de interferência sobre os direitos fundamentais das pessoas, tais como o direito à não discriminação. Logo, dado o elevado risco que a utilização da TRF por autoridades policiais representa para o direito fundamental de não discriminação, torna-se

fundamental levar em conta em uma DPIA o risco da violação de tal direito (Comitê Consultivo da Convenção para a proteção dos indivíduos com relação ao processamento automático de dados pessoais, 2021, p.14).

Trazendo a discussão para o âmbito de Portugal, é importante destacar a Lei nº 58/2019, que assegura a execução do GDPR na ordem jurídica nacional, e a Lei nº 59/2019, que transpõe a LED para o ordenamento jurídico português. Dessas normas, os seguintes dispositivos mostram-se relevantes quando se trata de TRF: (I) art. 19 da Lei nº 58/2019, que impõe limites ao uso de sistemas de videovigilância para fins de segurança privada. Apesar desse dispositivo não dizer respeito exatamente ao uso de TRF por autoridades policiais, ele serve como um indicativo de que a tecnologia deve ser usada de forma restrita e proporcional; (II) art. 6º da Lei nº 59/2019, que determina que o tratamento dos dados biométricos por autoridades policiais só pode ser feito se for estritamente necessário, e se estiver sujeito a garantias adequadas de proteção dos direitos e liberdades do titular dos dados, dentre outros requisitos. Além disso, proíbe as definições de perfis que conduzam à discriminação de pessoas com base em dados biométricos.

Por fim, fechando a análise do ordenamento jurídico europeu no que diz respeito ao direito de não discriminação e o uso de TRF por autoridades policiais, é preciso destacar um caso do Reino Unido⁷⁵, em que o Tribunal de Apelação anulou decisão de primeira instância, entre outros motivos, porque a *Polícia de South Wales* usou TRF em diversos locais públicos, em clara violação ao art. 8(2), Convenção Europeia dos Direitos do Homem, que estipula que não pode haver ingerência da autoridade pública no direito à vida privada das pessoas, senão quando esta ingerência estiver prevista na lei. Além disso, a justificativa legal usada pela polícia não era precisa e dava aos policiais individuais excessiva discricionariedade sobre onde o TRF poderia ser implantado e quem poderia ser colocado em uma lista de vigilância. Isso significava que a avaliação de impacto da proteção de dados preparada pela Polícia não atendeu aos requisitos da Seção 64 do “*Data Protection Act (DPA) 2018*”. O Tribunal também entendeu que a Polícia não havia feito tudo o que podia razoavelmente para cumprir o Dever de Igualdade do Setor Público, previsto na Seção 149 do “*Equality Act (EA) 2010*”, uma vez que não tomou as medidas necessárias para investigar e evitar a possibilidade de vieses algorítmicos. Por fim, a Corte considerou que como a TRF é uma tecnologia nova e controversa, é preciso que todas as

⁷⁵ Ver: *R (Bridges) v CC South Wales & ors* (2020) EWCA Civ 1058, (2020).

autoridades públicas que pretendam usá-la no futuro tomem todas as medidas necessárias para que o *software* não reproduza preconceito de gênero e de raça.

2.3. TENSÕES E LACUNAS NA LEGISLAÇÃO ANTIDISCRIMINAÇÃO DA UE RELACIONADAS AO USO DE TRF

Muitos estudiosos entendem que a atual legislação antidiscriminação da UE não é suficiente para lidar com sistemas de IA e proteger contra as suas formas sutis e não intuitivas de discriminação. Xenidis e Senden (2020, p.25), por exemplo, afirmam que a abordagem da justiça individual não é forte o suficiente para combater a discriminação algorítmica, pois uma série de problemas impede as pessoas de iniciar uma ação legal para proteger seu direito à igualdade. Tais problemas podem variar desde incerteza sobre o resultado e medo de vitimização, até problemas financeiros (por exemplo, custo dos processos, falta de assistência jurídica) e problemas institucionais (por exemplo, duração dos processos, falta de especialização e assistência, falta de confiança no judiciário, falta de compensação suficiente).

Além disso, a legislação secundária não é tão clara, o que gera insegurança para os demandantes. Por exemplo, não há uma definição clara do quão significativo deve ser um dano sofrido para que ele configure discriminação indireta. Da mesma forma, também não há uma definição clara de quais são os grupos desfavorecidos e quais são os grupos de comparação. Tais definições ocorrem de forma flexível, de acordo com o caso concreto, e dependem da argumentação das partes. Não há, por parte dos tribunais, uma abordagem consistente e coerente para avaliar corretamente a discriminação. Caso houvesse, desenvolvedores de sistemas, controladores, reguladores e usuários teriam ao seu dispor orientações e parâmetros legais suficientes para criar sistemas e mecanismos de governança capazes de detectar, remediar e prevenir a discriminação automatizada (Wachter *et al.*, 2020, p. 44). Essa incoerência na jurisprudência tem até mesmo o efeito de reduzir o impacto do art. 21 da Carta de Direitos Fundamentais da UE em detrimento das pessoas que deveriam ser protegidas (Ward, 2018, p.56).

Outro ponto destacado pelos estudiosos é o de que a lei de não discriminação da UE permite uma justificação fácil de algumas formas de discriminação algorítmica. Tal facilidade ocorre principalmente em relação à discriminação indireta, que requer apenas que um tratamento supostamente neutro seja adequado e proporcional para atingir um objetivo legítimo previsto em lei. Nesse sentido, questões como segurança pública,

segurança nacional e prevenção de crimes poderiam ser alegados para justificar o uso de TRF por autoridades públicas, mesmo que direitos fundamentais sejam violados (Hacker, 2018, p.1161). No entanto, existem autores que entendem que os tribunais não aceitarão com tanta facilidade o argumento de que a TRF é um meio adequado e proporcional. Segundo eles, a necessidade de se interpretar de forma restrita qualquer exceção ao direito de não discriminação, fará com que os tribunais sempre considerem que houve uma violação ao direito, a menos que o infrator comprove que utilizou o algoritmo menos enviesado (Xenidis e Senden, 2020, p.22).

Alguns estudiosos destacam também que a legislação secundária antidiscriminação da UE apresenta âmbito de aplicação limitado e alguns requisitos com limiares desproporcionalmente altos para que a proteção seja conferida (Madiega e Mildebrath, 2021, p.19). Com efeito, não há uma legislação secundária que possa ser aplicada de forma transversal, cobrindo todos os âmbitos da vida em sociedade. A legislação secundária é aplicada de forma segmentada, com leis específicas que cobrem apenas alguns setores, tais como os mercados de emprego e consumo. Logo, não há uma lei secundária da UE que estabeleça requisitos precisos que possam ser cumpridos a fim de se evitar a discriminação no uso de TRF por autoridades policiais (Pires, 2021, p.31), bem como que estabeleça limiares menos elevados para que os indivíduos possam buscar proteção. Além disso, o art. 51 da Carta de Direitos Fundamentais da UE estabelece que o direito fundamental à não discriminação tem por destinatários as instituições, órgãos e organismos da UE, bem como os Estados-Membros, quando estes apliquem o direito da União. Isso faz com que a aplicabilidade de tal direito seja incerta quando se trata de entidades privadas (Ward, 2018, p.53-56). Não bastasse, a jurisprudência do Tribunal de Justiça da União Europeia (TJUE) tem aplicado o direito de não discriminação de forma menos flexível e transversal do que se esperava (Gerards e Xenidis, 2021, p.65).

Por fim, há o entendimento de que mesmo que a legislação de não discriminação da UE tenha o seu campo de aplicação ampliado, ela não será capaz de combater a discriminação algorítmica de forma eficaz (Madiega e Mildebrath, 2021, p.20). Isso se deve, dentre outros motivos, porque: (I) a discriminação perpetrada por sistemas de IA é difícil de ser detectada pelas vítimas. Estas, muitas vezes, nem tomam consciência da discriminação, devido à velocidade, escala, e complexidade das decisões algorítmicas (Wachter *et al.* 2020, p.6-10); (II) mesmo que a vítima suspeite da discriminação, ela terá dificuldade para comprovar a sua existência, pois as provas das decisões algorítmicas estão na posse do operador ou fornecedor (Xenidis e Senden, 2020, p.20-24); (III) direitos

de propriedade intelectual, segredos comerciais e sigilo de investigação podem ser invocados pelos controladores como motivos de recusa ao acesso (Wachter *et al.*, 2020, p.10); (IV) mesmo que houvesse acesso ao algoritmo, este não seria inteligível por leigos no assunto (vítimas, advogados e juízes) (Gerards e Xenidis, 2021, p.75); (V) alguns sistemas de IA não são explicáveis nem mesmo pelos próprios programadores que o criaram, o que impede o entendimento do seu funcionamento e a identificação de decisões discriminatórias (Wachter *et al.*, 2020, p.12).

3. CAMINHOS REGULATÓRIOS PARA A MINIMIZAÇÃO DAS TENSÕES E LACUNAS EXISTENTES NA LEGISLAÇÃO ANTIDISCRIMINAÇÃO DA UE: *AI ACT* E PRINCÍPIO DA NÃO DISCRIMINAÇÃO *BY DESIGN*

A fim de superar as limitações identificadas no tópico anterior, diferentes medidas são sugeridas pelos estudiosos. Por exemplo, alguns entendem que é suficiente que disposições gerais de não discriminação sejam introduzidas nas legislações nacionais, a fim de atuarem como uma rede de segurança nos casos não abrangidos por leis gerais ou específicas (Gerards e Xenidis, 2021, p.141). Já outros sugerem que as lacunas sejam preenchidas por instrumentos legislativos tais como fiscalização pública, direito de acesso e informação, e reparação coletiva. (Hacker, 2018, p.1170).

Apesar de tais sugestões serem interessantes e terem o seu mérito, o presente trabalho focará na análise de duas abordagens regulatórias específicas, a fim de descobrir qual delas é a melhor opção para minimizar as tensões e lacunas existentes na legislação antidiscriminação da UE, e resolver de forma eficaz o problema do enviesamento no uso de TRF por autoridades policiais.

Uma das abordagens é a apresentada pela Comissão da UE, que em abril de 2021 publicou a sua Proposta de Regulamento de IA. Tal proposta adota uma regulação geral baseada no risco, aplicável à maioria dos sistemas de IA, dentre eles a TRF. Conforme afirma a Comissão na Exposição de Motivos, a Proposta complementa o direito da União em vigor em matéria de não discriminação com requisitos específicos que visam minimizar o risco de discriminação algorítmica.

Já a outra abordagem que será analisada é a que foca na promoção e desenvolvimento de um “princípio de não discriminação *by design*” (ou igualdade *by design*) a fim de superar as lacunas e tensões presentes na atual legislação antidiscriminação da UE (Xenidis e Senden, 2020, p.26-29). Tal princípio estabelece que medidas técnicas e organizacionais adequadas devem ser implementadas por fornecedores⁷⁶ e utilizadores⁷⁷ em todas as etapas de desenvolvimento de uma TRF (ou

⁷⁶ Conceito de fornecedor presente no art. 3(2)(3) da Proposta de Regulamento de IA da Comissão: "Fornecedor" é uma pessoa singular ou coletiva, autoridade pública, agência ou outro organismo que desenvolva um sistema de IA ou que tenha um sistema de IA desenvolvido com vista à sua colocação no mercado ou colocação em serviço sob o seu próprio nome ou marca, a título oneroso ou gratuito. Já "Fornecedor de pequena dimensão" é um fornecedor que seja uma micro ou pequena empresa na aceção da Recomendação 2003/361/CE da Comissão.

⁷⁷ Conceito de utilizador presente no art. 3(4) da Proposta de Regulamento de IA da Comissão: "Utilizador" é uma pessoa singular ou coletiva, autoridade pública, agência ou outro organismo que utilize, sob a sua

seja, desde a fase de projeto até a fase de uso), a fim de que sejam evitados ou minimizados vieses de gênero e de raça (Pires, 2021, p.33).

Por óbvio, é preciso destacar que cada uma das abordagens apresenta um nível diferente de regulamentação. Enquanto a Proposta representa um tipo de regulamentação geral, aplicável a numerosos sistemas de IA e voltada a diferentes questões de direitos fundamentais, o princípio da não discriminação *by design* representa uma regulamentação específica, voltada exclusivamente para a questão da discriminação perpetrada por TRF usada por autoridades policiais. No entanto, apesar de tais diferenças, ambas as abordagens apresentam aspectos importantes quando se trata de combater vieses em sistemas de reconhecimento facial usados por autoridades policiais.

3.1. A PROPOSTA DE REGULAMENTO DE IA DA COMISSÃO EUROPEIA E A SUA ABORDAGEM EM RELAÇÃO AO USO DE TRF POR AUTORIDADES POLICIAIS

A Proposta de Regulamento da Comissão Europeia surge como o resultado de debates e orientações publicadas anteriormente pela entidade, tais como as “Orientações Éticas para a IA Confiável” (publicada em 2019 pelo Grupo Independente de Peritos de Alto Nível sobre a IA, da Comissão), e o *White Paper On AI – A European approach to excellence and trust* (publicado em 2020 pela Comissão).

Ao escolher um regulamento geral aplicável a diferentes tipos de IA, a Comissão optou por uma abordagem regulatória baseada em regras detalhadas. Por um lado, essa abordagem apresenta pontos negativos, tais como: a possibilidade das regras não alcançarem o verdadeiro objetivo visado pela regulamentação, uma vez que os entes regulados podem ser levados a cumprirem a regra literalmente, mas não a sua finalidade; menor flexibilidade e durabilidade em comparação aos princípios; maior restrição à discricionariedade dos entes regulados; menor campo de aplicação (Ford, 2008, p.7). Por outro lado, a abordagem baseada em regras possui pontos positivos, tais como: maior certeza e previsibilidade do que os princípios; maior clareza e facilidade de serem aplicadas uniformemente; impossibilidade de conformidade mínima por regulações mal-intencionadas; redução da arbitrariedade de reguladores e regulados; menor abertura à

autoridade, um sistema de IA, salvo se o sistema de IA for utilizado no âmbito de uma atividade pessoal de caráter não profissional.

interpretação, o que confere maior poder aos supervisores e aos sistemas de conformidade (Black *et al.*, 2007, p.191-195).

Tendo isso em mente, é preciso destacar que a Proposta de Regulamento da Comissão busca estabelecer uma classificação para sistemas de IA, com diferentes requisitos e obrigações, adaptados a uma abordagem baseada no risco. Esse tipo de abordagem diferencia entre as utilizações de IA que criam: (I) um risco inaceitável; (II) um risco elevado; (III) um risco baixo ou mínimo.

Os usos de risco inaceitável são aqueles particularmente nocivos e que violam os valores da UE (tais como os direitos fundamentais). Representam uma clara ameaça para a segurança, subsistência e direitos das pessoas e, por isso mesmo, são proibidos. O art. 5º da Proposta prevê tais usos como os concebidos para manipular o comportamento humano através de técnicas subliminares e pontuação social por parte dos governos. Também é proibido o uso de sistemas de identificação biométrica à distância em tempo real em espaços acessíveis ao público para efeitos de manutenção da ordem pública, a não ser que se apliquem determinadas exceções limitadas.

Já os usos de risco elevado criam um impacto adverso na saúde e segurança das pessoas ou nos seus direitos fundamentais. De acordo com os arts. 6º e 7º da Proposta, a classificação de um sistema de IA como de risco elevado não depende só da função deste sistema, mas também da finalidade específica e das modalidades para as quais ele é utilizado. O Capítulo 1 do Título III da Proposta fixa as regras de classificação e reconhece 2 conjuntos principais de sistemas de IA de elevado risco: (I) primeiro, sistemas de IA criados para serem usados como elementos de segurança de produtos que são suscetíveis a uma avaliação da conformidade *ex ante* realizada por terceiros; (II) segundo, sistemas de IA autônomos especificamente elencados no Anexo III da Proposta, devido às consequências que trazem para os direitos fundamentais.

Essa lista do Anexo III inclui um número limitado de sistemas de IA cujos riscos já se materializaram ou são suscetíveis de se materializar num futuro próximo, e pode ser atualizada pela Comissão (art. 7º da Proposta). Tal lista inclui sistemas de IA utilizados para: identificação biométrica e categorização de pessoas singulares; formação educacional ou profissional; infraestruturas críticas (por exemplo, transporte); emprego; componentes de segurança do produto (por exemplo, aplicações de IA em cirurgia assistida por robôs); aplicação da lei; migração, asilo e gestão do controle de fronteiras (por exemplo, verificação da autenticidade dos documentos de viagem); administração da

justiça e processos democráticos; serviços públicos e privados essenciais (por exemplo, pontuação de crédito que nega aos cidadãos oportunidades de obter um empréstimo).

Tais sistemas de risco elevado são autorizados no mercado europeu, mas antes de serem colocados à disposição (para comercialização) terão de cumprir uma série de requisitos de segurança (previstos no capítulo 2 do Título III da Proposta, relativos, por exemplo, à gestão dos riscos, à supervisão humana e à gestão dos dados) (art. 16, “a”; e Considerandos 27 e 43, todos da Proposta), bem como terão de ser submetidos a uma avaliação de conformidade *ex ante* (art. 16, “e”; e Considerandos 54 e 62, todos da Proposta). Além disso, o art. 61 da Proposta (tal como o Considerando 54) estabelecem que os fornecedores devem criar e documentar um sistema de acompanhamento e supervisão *ex post* (ou seja, pós-comercialização), a fim de assegurar o cumprimento das obrigações e requisitos para todos os sistemas de IA de alto risco já colocados no mercado.

Por fim, existem os usos de IA que criam um risco baixo ou mínimo. Os que criam um risco baixo estariam sujeitos a um conjunto limitado de obrigações, tais como transparência. Por sua vez, os que representam um risco mínimo poderiam ser desenvolvidos e utilizados na UE sem obrigações legais adicionais, para além da legislação existente (Madiega e Mildebrath, 2021, p.24).

Levando em conta o rápido desenvolvimento da tecnologia e dos mercados relacionados à IA, a proposta de Regulamento da Comissão, (I) por um lado, procura adotar uma posição tecnologicamente neutra e preparada para o futuro, (II) e por outro, pretende representar uma abordagem horizontal, aplicável a sistemas de IA desenvolvidos, implantados e usados em todas as áreas, incluindo as áreas relacionadas às autoridades policiais e judiciárias. Nessa esteira, a Proposta tem o intuito de também ser aplicável a todos os sistemas de identificação biométrica remota (*remote biometrics identification systems – RBI*), incluindo as TRFs usadas por autoridades policiais. De acordo com o Considerando 8 e art. 3(33) da Proposta, tais sistemas possuem as seguintes características: são operados à distância sem se saber antecipadamente se a pessoa visada estará presente na área; coletam dados biométricos; comparam tais dados a um banco de dados; e são usados especificamente para identificar uma pessoa.

Ainda de acordo com a Proposta da Comissão, tendo em vista as diferentes características e formas como são utilizados, bem como os diferentes riscos inerentes, deve ser feita uma diferenciação entre os sistemas de identificação biométrica remota “em tempo real” (*real-time biometric identification systems*) e os sistemas de identificação

biométrica remota “em tempo diferido” (*post biometric identification systems*). Os primeiros são definidos como sistemas que capturam dados biométricos e executam os processos de comparação e identificação de forma instantânea (ou sem significativo atraso), com base em imagens obtidas “em tempo real” ou “quase em tempo real”, geradas por câmeras ou dispositivos parecidos. Por sua vez, os sistemas que operam “em tempo diferido” são aqueles que coletam os dados biométricos e processam a comparação e identificação após um atraso significativo. Funcionam com base em imagens geradas por CFTV ou dispositivos privados.

Diante do contexto da Proposta, diferentes cenários podem ser identificados para a regulação do reconhecimento facial, a depender do tipo de TRF e finalidade para a qual é usada. Nesse sentido, o primeiro cenário está previsto no art. 5(1)(d), e Considerandos 18 e 19 da Proposta, que propõem a proibição da IA ser usada para realizar identificação biométrica remota em tempo real de pessoas em espaços acessíveis ao público para efeitos de manutenção da ordem pública. Tal proibição geral decorre do fato desse tipo de TRF ameaçar de forma inaceitável os direitos e valores fundamentais da UE. Com efeito, o Considerando 18 da Proposta esclarece que tais sistemas são considerados particularmente intrusivos para os direitos e liberdades das pessoas envolvidas, afetam a vida privada de grande parte da população, podem originar uma sensação de vigilância ininterrupta e indiretamente dissuadir o exercício da liberdade de reunião e de outros direitos fundamentais. Não bastasse, os sistemas que funcionam em tempo real aumentam os riscos de violação dos direitos e liberdades das pessoas visadas pelas autoridades policiais, uma vez que tal tipo de tecnologia produz efeitos imediatos e não há mecanismos hábeis para controlá-la e corrigi-la a tempo em caso de erro. No fim das contas, a proibição prevista na Proposta visa impedir que as autoridades policiais cometam excessos como, por exemplo, implantar TRF em locais públicos para identificar indivíduos que apenas praticaram delitos menores, ou para identificar manifestantes que participam de atos públicos. Procura-se impedir a ocorrência de uma vigilância em massa (Christakis e Becuywe, 2021).

O segundo cenário para regulação da TRF também está previsto no art. 5(1)(d) e Considerando 19 da Proposta, e diz respeito a 3 exceções à regra geral de proibição, que se justificariam por haver um interesse público substancial que supera os riscos para os direitos fundamentais. Nesse sentido, o uso de identificação biométrica remota em tempo real em espaços públicos para efeitos de manutenção da ordem pública só é permitido para as seguintes exceções, caso estritamente necessárias: (I) investigação seletiva de

potenciais vítimas de crimes, especialmente crianças desaparecidas; (II) prevenção de uma ameaça específica, substancial e iminente à vida ou à segurança física de pessoas ou de um ataque terrorista; (III) detecção, localização, identificação ou instauração de ação penal relativamente a infratores ou suspeitos de infrações penais graves, referidas no artigo 2º, nº 2, da Decisão-Quadro 2002/584/JAI do Conselho e punível no Estado-Membro em causa com pena ou medida de segurança privativas de liberdade de duração máxima não inferior a três anos. Esse limiar da pena serve como baliza para assegurar que somente infrações suficientemente graves podem justificar o uso de sistemas de identificação biométrica à distância em tempo real. Além disso, essa exceção agilizaria o procedimento de extradição entre os países da UE de pessoas que cometeram crimes graves.

De acordo com o art. 5(2) e (3) e Considerandos 20, 21 e 23 da Proposta, o uso excepcional deste tipo de sistema (em cada uma das 3 situações elencadas acima) está sujeito ao atendimento de requisitos adicionais, tais como: realização de avaliações de necessidade e proporcionalidade; proporcionalidade na lista de vigilância (a base de dados de pessoas utilizada como referência deve ser adequada a cada utilização em cada uma das situações); observância de salvaguardas e condições necessárias e proporcionadas em relação a tais utilizações (como, por exemplo, limitações temporais, geográficas e das pessoas visadas); respeito aos princípios consagrados no GDPR; sujeição à autorização expressa e específica de autoridades judiciais ou administrativas independentes do Estado-Membro no qual o sistema é implementado (a princípio, a autorização deve ser obtida antes da sua utilização, salvo em situações de urgência devidamente justificadas e limitadas ao mínimo absolutamente necessário). O art. 5(4) e o Considerando 22 da Proposta ainda estabelecem que cabe aos Estados-Membros a decisão de implementar (total ou parcialmente), ou mesmo de não implementar, as 3 exceções em suas leis nacionais. Com efeito, a Proposta reconhece que questões de segurança nacional dizem respeito à competência dos Estados-Membros e, com base nisso, busca conciliar segurança nacional e ordem pública, por um lado, e proteção de dados e demais direitos fundamentais (como não discriminação) por outro lado.

Já o terceiro cenário previsto na Proposta para a regulação da TRF está previsto no art. 6(2), no Considerando 33, e no parágrafo 1(a) do Anexo III, todos da Proposta. De acordo com tais dispositivos, os demais sistemas de identificação biométrica remota, seja em tempo real, seja em tempo diferido, devem ser classificados como de risco elevado, independentemente de sua finalidade específica (manutenção da ordem ou não) e do local

de uso (espaços acessíveis ao público ou não). O motivo se deve ao fato de que as imprecisões técnicas de tais sistemas podem conduzir a resultados enviesados e ter efeitos discriminatórios, principalmente no que diz respeito à idade, etnia, gênero ou às deficiências das pessoas.

Este terceiro cenário abrange um amplo conjunto de usos de TRF, tais como: o uso por entes privados de TRF em tempo real em espaços acessíveis ao público (para, por exemplo, controlar a entrada em transportes, escolas e eventos esportivos; realizar pagamentos em lojas; para fins de saúde pública); o uso de TRF em tempo real por autoridades públicas para outros fins que não a manutenção da ordem pública (para, por exemplo, controlar o acesso a um prédio público); o uso de TRF em tempo real, seja por autoridades policiais ou não, em espaços não acessíveis ao público; o uso de TRF em tempo diferido, por autoridades policiais ou não (para, por exemplo, identificar a pessoa que cometeu um crime) (Christakis e Becuywe, 2021; Kind, 2021).

Os sistemas de reconhecimento facial elencados no terceiro cenário não entram na proibição geral da Proposta. No entanto, são classificados como de elevado risco e, por isso mesmo, estão sujeitos ao cumprimento de exigências específicas e várias obrigações de conformidade, a fim de que possam ser colocados no mercado da UE ou colocados em serviço. Procura-se, segundo o Considerando 27 da Proposta, garantir que o uso de tais sistemas não represente riscos inaceitáveis aos interesses públicos da UE, conforme reconhecidos e protegidos pela legislação da UE.

Nesse sentido, a fim de reduzir os riscos de violação dos direitos fundamentais, o uso da TRF de elevado risco (seja em tempo real ou em tempo diferido) está condicionado ao cumprimento das seguintes medidas de controle *ex ante*: criação, implementação, documentação e manutenção de um sistema de gestão de riscos através de medidas adequadas de avaliação e mitigação, que devem ser executadas e atualizadas continuamente (art. 9, Proposta); utilização de conjuntos de dados (de treino, validação e teste) que cumpram os critérios de qualidade e práticas apropriadas de governança de dados (art. 10, Proposta); elaboração e atualização constante da documentação técnica detalhada (art. 11, Proposta); desenvolvimento de capacidades que permitam o registro automático de eventos enquanto a IA estiver funcionando, e que forneçam, no mínimo, informações sobre cada período de uso da TRF, a galeria, os dados de entrada e a identidade das pessoas físicas que verificam os resultados (art. 12, Proposta); garantia de transparência e fornecimento de informações adequadas aos utilizadores, a fim de que estes possam interpretar o resultado do sistema de IA e utilizá-lo corretamente (art. 13,

Proposta); implementação de medidas adequadas de supervisão humana, de modo a assegurar que nenhum utilizador possa tomar uma ação ou decisão com base na identificação resultante do sistema de identificação biométrica remota, exceto se o resultado for verificado e confirmado por no mínimo 2 pessoas físicas (art. 14, Proposta); garantia de que a TRF seja criada e desenvolvida de modo a alcançar níveis adequados de exatidão, robustez e cibersegurança, bem como tenha um desempenho coerente em relação a tais aspetos durante o ciclo de vida (art. 15, Proposta).

Além disso, a Proposta, no seu Título III (arts. 16 a 29), sujeita os fornecedores (incluindo importadores e distribuidores) e utilizadores de TRF de risco elevado ao cumprimento de rigorosas obrigações *ex ante*. Tais obrigações envolvem, por exemplo: a criação de um sistema de gestão de qualidade (art. 17, Proposta); sujeição a uma avaliação de conformidade (art. 19, Proposta); adoção de medidas corretivas necessárias (art. 21, Proposta); dever de informação e cooperação com as autoridades nacionais competentes (arts. 22 e 23, Proposta). De acordo com o art. 43 e o Considerando 64 da Proposta, uma entidade independente deve realizar a avaliação de conformidade na TRF de risco elevado (ao contrário da autoavaliação aplicável a outros tipos de sistemas de IA de risco elevado), a menos que existam normas harmonizadas ou especificações comuns.

De acordo com a Proposta, cumpridos todos os requisitos e obtida a certificação, a TRF de risco elevado poderá ser comercializada e usada por agentes públicos e privados, sempre respeitando a legislação da UE. Com efeito, o Considerando 24 da Proposta esclarece que o tratamento de dados biométricos e de outros dados pessoais envolvidos no uso de TRF de risco elevado deve continuar a cumprir todos os requisitos decorrentes do art. 9(1), GDPR, art. 10, LED, e art. 10(1), Regulamento (UE) 2018/1725.

Todas essas medidas de controle *ex ante* são complementadas com um sistema *ex post* robusto de monitoramento do mercado e supervisão das TRFs de alto risco. De fato, o art. 61 da Proposta determina que os fornecedores devem criar e documentar um sistema de acompanhamento pós-comercialização que seja proporcionado à natureza das tecnologias de inteligência artificial e aos riscos do sistema de IA de risco elevado. Além disso, o sistema de monitorização pós-comercialização deve basear-se num amplo plano de acompanhamento pós-comercialização.

3.2. PRINCÍPIO DA NÃO DISCRIMINAÇÃO *BY DESIGN*

Outra abordagem regulatória que se mostra como opção para minimizar as tensões e lacunas existentes na legislação antidiscriminação da UE, e resolver de forma eficaz o problema do enviesamento no uso de TRF por autoridades policiais, é a abordagem que foca no desenvolvimento de um princípio de não discriminação *by design* (ou igualdade *by design*) (Xenidis e Senden, 2020, p.26-29). Por meio deste princípio, medidas técnicas e organizacionais adequadas devem ser implementadas por desenvolvedores, fornecedores e usuários em todas as etapas de desenvolvimento de uma TRF usada por autoridades policiais (ou seja, nas fases de concepção/projeto, desenvolvimento e implementação/utilização), a fim de que sejam evitados ou minimizados vieses de gênero e de raça e que sejam garantidas salvaguardas às pessoas afetadas (Pires, 2021, p.50).

Com efeito, estudiosos reconhecem que uma abordagem de não discriminação *by design* poderia minimizar as deficiências da lei antidiscriminação da UE atualmente em vigor (Wachter *et al.* 2020, p.47). Em primeiro lugar, tal abordagem minimizaria as incertezas e ambiguidades da atual legislação antidiscriminação da UE em relação à TRF, uma vez que estabeleceria medidas mais concretas para impedir a discriminação algorítmica em tais sistemas. Além disso, uma abordagem de não discriminação *by design* elevaria os requisitos e parâmetros para que um tratamento discriminatório pudesse ser justificado. Dessa forma, tratamentos que configuram discriminação indireta (como na maior parte dos casos envolvendo discriminação algorítmica) não poderiam ser justificados de forma fácil, como ocorre atualmente (Pires, 2021, p.33).

Não bastasse, um princípio assim elevaria a pressão para que desenvolvedores e fornecedores se empenhassem em projetar novos sistemas de reconhecimento facial que não fossem enviesados. Consequentemente, a pressão exercida sobre a base da cadeia de fornecimento da TRF estimularia uma minimização geral do viés na TRF usada por autoridades policiais. Explica-se: um tratamento injusto, seja ele direto ou indireto, não pode ser justificado caso exista uma tecnologia tão precisa quanto a usada e menos enviesada. Logo, autoridades policiais, a fim de não violarem o princípio da não discriminação *by design*, teriam que abandonar a TRF antiga e enviesada, em favor de outra que atenda a padrões aceitáveis de precisão (Pires, 2021, p.34).

Também é interessante destacar que por representar uma abordagem regulatória específica, voltada exclusivamente para a questão da discriminação perpetrada por TRF usada por autoridades policiais, o princípio da não discriminação *by design* apresenta-se como um instrumento com maior campo de aplicação do que a legislação secundária antidiscriminação da UE. Não bastasse, tal princípio pode incentivar a produção de uma

legislação secundária que regule especificamente a não discriminação no âmbito da TRF operada por autoridades policiais, ao contrário do que seria feito em uma legislação geral de sistemas de IA. Dessa forma, haveria a possibilidade de que fossem criados requisitos mais claros e contextuais para que os atores envolvidos cumprissem o direito de não discriminação (à semelhança do que foi feito com a proteção de dados na LED, por exemplo) (Pires, 2021, p.33-34).

Dito isso, é preciso destacar que uma abordagem de regulamentação baseada em princípios opta por adotar normas mais amplas, flexíveis e menos detalhadas do que regras específicas (Black *et al.*, 2007, p.191). Logo, passíveis de serem aplicadas a uma série de situações diferentes. O foco está em atingir a essência do regulamento, não em regular cada situação possível (Ford, 2008, p.29). Além disso, um princípio de não discriminação *by design* deve estabelecer normas gerais claras, seguidas de outras medidas reguladoras mais específicas e concretas que orientarão os desenvolvedores de TRF (e demais atores) a traduzi-las em medidas técnicas práticas que impedirão o enviesamento da tecnologia (Black, 2010, p.12). Consequentemente, ao se abrir maior espaço para que os entes regulados (desenvolvedores, fornecedores e usuários) possam participar da concretização do princípio, aumenta-se também a responsabilidade dos dirigentes de tais entes para com a conformidade da tecnologia (Black *et al.*, 2007, p.193). E não só, pois a atuação conjunta entre reguladores e regulados faz com que o diálogo entre ambas as partes seja facilitado e encorajado, a fim de que conjuntamente clarifiquem os objetivos e os requisitos que devem ser cumpridos pelos regulados (Pires, 2021, p. 36-37).

Também é importante ressaltar que, por ser flexível, uma abordagem baseada em princípio é adaptável de forma rápida a mudanças que venham a ocorrer na sociedade e no ordenamento jurídico ao longo do tempo (Black *et al.*, 2007, p.195). Além disso, é uma escolha mais barata, pois é necessário menos conhecimento para redigir um princípio (Ford, 2008, p.38-40); e é a melhor opção para se atingir de fato a finalidade de uma norma, impedindo assim que os regulados cumpram a regra literal, mas não sua finalidade (Black *et al.* 2007 p.194).

Outro ponto relevante é que uma abordagem regulatória baseada no princípio da não discriminação *by design* leva em consideração o fato de que nenhuma tecnologia é neutra. Com efeito, ao serem desenvolvidas, as tecnologias (dentre elas a TRF) incorporam os valores e visões de mundo das pessoas e entidades que as desenvolvem e as utilizam (Friedman *et al.*, 2008, p. 85). Inclusive, a própria utilização de tais

tecnologias é orientada por essas subjetividades. Tendo em mente, portanto, que não há neutralidade, é essencial que o respeito aos direitos humanos fundamentais (como a não discriminação) seja incorporado a toda a cadeia de produção, desenvolvimento e uso da tecnologia, de forma a impedir a reprodução de padrões violadores de tais direitos (Pires, 2021, p. 40).

Como não poderia deixar de ser, a formulação de um princípio de não discriminação *by design* inspira-se em grande medida no já consagrado princípio de privacidade *by design*, previsto no art. 25(1) do GDPR, art. 20(1) da LED, e art. 10(2) e (3) da Convenção 108+ (Cavoukian, 2009, p.2-7). Isso significa que a implementação da não discriminação em todas as etapas de uma TRF demandará, assim como acontece com a privacidade *by design*, uma postura pró-ativa e preventiva por parte dos desenvolvedores, fornecedores e usuários, de forma a não se limitarem apenas ao mero cumprimento da legislação antidiscriminação em vigor, mas sim a irem além, concretizando de fato a não discriminação como uma prática padrão. Não bastasse, é preciso que tais atores assegurem máxima transparência em todo o processo de desenvolvimento e uso da TRF, bem como assegurem que a tecnologia possua configurações padrão mais amigáveis possíveis à não discriminação. É necessário que este princípio seja um dos componentes centrais no projeto da TRF (Pires, 2021, p.44).

Como foi dito, a concretização do princípio da não discriminação *by design* na TRF usada por autoridades policiais envolve uma série de medidas técnicas e organizacionais adequadas que devem ser implementadas durante a concepção, desenvolvimento e utilização da tecnologia. Tais medidas podem ser traduzidas em orientações regulamentares específicas de implementação que nortearão as políticas e práticas dos diferentes atores envolvidos (desenvolvedores, fornecedores e usuários), que de posse dos seus conhecimentos científicos especializados é quem elaborarão e implementarão medidas técnicas concretas no uso da tecnologia durante o seu dia a dia (Pires, 2021, p.50).

Sendo assim, um primeiro exemplo de tais orientações pode ser traduzido na necessidade de que a TRF usada por autoridades policiais seja transparente, explicável e inteligível (Zuiderveen Borgesius, 2018, p.34). Isso significa que as informações a respeito do sistema de reconhecimento facial devem ser abrangentes, concisas e compreensíveis, e disponibilizadas de maneira clara e em linguagem direta, simples e acessível a toda a sociedade (Pires, 2021, p.55). Busca-se, com isso, entender, interpretar e explicar como funciona a TRF e o seu processo de tomada de decisão, de forma a

minimizar a questão da caixa-preta dos algoritmos (*algorithm black box*) e possibilitar aos indivíduos (e à sociedade como um todo) entenderem como a TRF chegou a um determinado resultado após correlacionar e pesar diversas variáveis contidas nos dados (Magrani *et al.* 2021a, p.23).

A transparência, a explicabilidade e a inteligibilidade podem ser garantidas através de medidas como, por exemplo: (I) a informação à sociedade sobre o uso de TRF por autoridades policiais e os riscos de violação dos direitos fundamentais, bem como informação sobre as demais opções disponíveis para se atingir o mesmo fim pretendido (Comitê Consultivo da Convenção para a proteção dos indivíduos com relação ao processamento automático de dados pessoais, 2021, p.11); (II) publicação dos riscos e limitações potenciais que os desenvolvedores identificaram em relação ao direito de não discriminação (*Amnesty International e Access Now*, 2018, p.9-14); (III) publicação do código do algoritmo para que possa ser checada a existência de vieses. Apesar de existirem questões relacionadas a segredos comerciais e propriedade intelectual, que limitariam o acesso ao código, o interesse público envolvido possibilitaria que fossem criados meios para que especialistas independentes terceirizados e autoridades europeias (e nacionais) pudessem ter acesso ao código para assegurar a sua conformidade com o direito de não discriminação (Pires, 2021, p.54); (IV) a publicação periódica de todos os relatórios, testes e auditorias (internas ou independentes) relacionados à TRF usada por autoridades policiais (Garvie *et al.* 2016, p.62-68); (V) disponibilização de registros precisos a respeito dos conjuntos de dados de treinamento e de avaliação usados, bem como disponibilização dos processos de programação, treinamento, teste e avaliação da TRF; (VI) informação sobre o funcionamento da TRF de forma detalhada e compreensível para todos (Pires, 2021, p.54).

Outro exemplo de medida regulamentar específica que pode contribuir para a concretização do princípio da não discriminação *by design* é a implementação de diversidade na força de trabalho. Isso significa que tanto os desenvolvedores e fornecedores, por um lado, quanto as autoridades policiais, por outro, deverão focar na contratação de mão-de-obra diversificada, tanto no que diz respeito a uma maior representatividade de grupos demográficos distintos, quanto no que diz respeito a uma maior diversificação de especializações. E essa diversidade deverá estar presente em todas as etapas de concepção, desenvolvimento e implementação da TRF (Pires, 2021, p.52).

De fato, a existência de um corpo profissional de formação diversa e com experiências distintas contribuirá muito para a concretização da não discriminação *by design* na TRF. Isso significa que o desenvolvimento de uma IA responsável deve passar pelo debate e crivo de diversos grupos sociais e profissionais qualificados (Magrani, 2021b). De fato, a incorporação de mais mulheres e pessoas não-brancas ao corpo profissional das empresas e órgãos públicos permitirá que sejam dadas mais contribuições a partir de diferentes perspectivas, o que aumentará a conscientização e prevenirá riscos de enviesamento, já que mais pessoas (que inclusive também são vítimas de tecnologias enviesadas) estarão analisando as possíveis consequências de um projeto (Magrani *et al.* 2021a, p. 44; HLEGAI, 2019b, p.18). Possibilita-se, com isso, a criação de uma cultura compromissada em detectar e mitigar o viés de gênero e de raça (Pires, 2021, p.52), e que permitirá ao algoritmo aprender a reconhecer e processar melhor as características de grupos demográficos minoritários (Danks e London, 2017, p.4691).

Não bastasse, além de uma equipe demograficamente diversificada, é igualmente necessária uma equipe composta por profissionais técnicos altamente qualificados de diferentes áreas, tais como engenheiros e cientistas da computação, mas também de especialistas em direitos humanos, ética, ciências sociais, gerência de fluxo de trabalho e operações, consultores, advogados e encarregados pelo tratamento de dados pessoais (Magrani, 2021b; FRA, 2019b, p.34). Além disso, a diversificação também está relacionada com a contratação de profissionais externos responsáveis pela *compliance* da organização (Magrani *et al.*, 2021a, p.44), bem como com a realização de consultas e diálogos com terceiros interessados (ONGs, representantes de grupos afetados, entidades públicas e especialistas em direitos humanos) em todo o ciclo de vida da TRF (Pires, 2021, p.53). Isso é importante, pois tais atores externos terão maior capacidade e liberdade para apontar déficits e recomendar melhorias (Magrani *et al.*, 2021a, p.44).

No entanto, é preciso ressaltar que, mesmo com a diversificação das equipes de trabalho, também é importante que se invista na capacitação e treinamento adequados em direitos humanos, ética e aspectos jurídicos (Magrani *et al.*, 2021a, p.45), tanto para as equipes que desenvolverão a TRF (programadores e engenheiros de *software*), quanto para as equipes das autoridades policiais que utilizarão a TRF (Pires, 2021, p.53). Dessa forma, possibilita-se que tais atores compreendam os riscos dos vieses algorítmicos em relação ao direito de não discriminação e saibam como combatê-lo (Amnesty International e Access Now, 2018, p.10). Além disso, a exigência de um programa de treinamento estruturado e apropriado uniformiza o aprendizado para todos os integrantes

das equipes e possibilita que todos conjuntamente (e não apenas individualmente) reflitam sobre os aspectos legais, sociais e éticos da tecnologia (Pires, 2021, p.53).

Em relação especificamente aos agentes policiais, estes deverão ser altamente treinados para compreenderem e interpretarem de forma correta o funcionamento e os resultados da TRF, de forma a estarem atentos aos possíveis vieses e falsos positivos (Pires, 2021, p.53). É essencial que eles entendam as questões operacionais e sejam capazes de interpretar e agir adequadamente, tendo em vista as exceções geradas pelo sistema. Além disso, eles também devem ser assistidos com uma infraestrutura gerencial de apoio que permita lidar de forma rápida e eficaz com situações de ambiguidade. Tais medidas são cruciais para se crie uma maior confiança do público na tecnologia (Introna e Nissenbaum, 2010, p.41).

Outra medida que pode contribuir para a concretização da não discriminação *by design* é o uso de conjuntos de dados de alta qualidade e demograficamente diversificados (Pires, 2021, p.51). Essa alta qualidade e diversidade são requisitos essenciais em todos os conjuntos de dados usados na TRF, sejam eles de treinamento, de avaliação (visam garantir resultados confiáveis ao testar o viés) (Comitê Consultivo da Convenção para a proteção dos indivíduos com relação ao processamento automático de dados pessoais, 2021, p.9), ou até mesmo o conjunto de dados presente na galeria usada nas pesquisas da TRF (listas de vigilância) (Garvie *et al.*, 2016, p.56). É essencial que sejam criados grandes bancos de dados com inúmeras imagens de pessoas de diferentes grupos demográficos, configurando, assim, uma representatividade demográfica equilibrada em tais bancos (Pires, 2021, p.51). Além disso, estes bancos devem ser abertos ao público e à sociedade, a fim de que desenvolvedores e usuários possam usá-los para aperfeiçoar seus algoritmos e evitar, assim, falsos positivos em relação a grupos historicamente sub-representados (Buolamwini e Gebru, 2018, p.77).

É importante que os conjuntos de dados aqui discutidos reproduzam as condições reais de uso da uma TRF, dando especial atenção para os aspectos relacionados à qualidade e captura das imagens, tais como, por exemplo, iluminação, ângulo em que a imagem foi capturada, qualidade da câmera (Introna e Nissenbaum, 2010, p.18). Tais questões devem ser levadas em consideração, pois a TRF é uma tecnologia muito sensível e o menor dos detalhes pode impactar a acurácia do sistema. Em outras palavras, o comportamento da tecnologia depende muito da qualidade do conjunto de dados usado para treinar os algoritmos. Além disso, outra questão importante é a necessidade de que os desenvolvedores coletem informações sobre os próprios conjuntos de dados

(metadados) por eles usados para treinar e avaliar os algoritmos, a fim de descobrirem se tais conjuntos são adequados ou se precisam ser mudados (FRA, 2018b, p.11).

Outra medida que pode contribuir para a concretização da não discriminação *by design* é a realização de supervisão e auditorias internas e independentes (Pires, 2021, p.55). Com efeito, o uso de TRF por autoridades policiais requer uma atenção constante a fim de que se possa identificar possíveis riscos (como a discriminação) e agir de forma rápida para minimizá-los. Uma postura preventiva deve sempre ser adotada (Magrani *et al.*, 2021a, p.46). Diante disso, é preciso que desenvolvedores, fornecedores e usuários (autoridades policiais) garantam a existência de instrumentos internos de supervisão e teste para que seja possível identificar os vieses e minimizá-los rapidamente (Pires, 2021, p.55). Tais instrumentos devem ser previstos e usados em todas as fases de existência da TRF: desde a concepção pelos desenvolvedores, até a implementação pelas autoridades policiais (HLEGAI, 2019b, p.22).

Nesse contexto de supervisionamento e testagem preventivos, ganham especial importância as avaliações de impacto dos direitos fundamentais (*fundamental rights impact assessment – FRIA*) (Pires, 2021, p.56). Tais instrumentos podem ser aplicados antes e durante a implementação da TRF e devem ser conduzidos de forma transparente por uma equipe de profissionais composta por especialistas de diferentes áreas (advogados, programadores, engenheiros, especialistas em ética), que identificarão os potenciais riscos da tecnologia sobre os direitos fundamentais (dentre eles, o da não discriminação), e indicarão medidas adequadas para reduzir tais riscos. Tudo feito de forma pública e transparente. (Zuiderveen Borgesius, 2018, p.28-29; Pires, 2021, p.56). Durante a realização da FRIA, devem ser fornecidas aos avaliadores todas as informações possíveis sobre a TRF, tais como, por exemplo, sua finalidade, funcionamento, bancos de dados usados, consequências negativas, resultados dos testes de precisão (FRA, 2020, p.97).

Outro instrumento que ganha destaque no contexto da supervisão preventiva é a auditoria de algoritmos realizada por especialistas altamente qualificados. Tais auditorias podem ser realizadas de forma interna ou por especialistas externos independentes (órgão do governo ou profissional terceirizado contratado) (Magrani *et al.* 2021a, p.49), e têm como objetivo avaliar a consistência dos sistemas de reconhecimento facial relativamente às normas e princípios em vigor (como a não discriminação), com foco principal na revisão dos códigos-fonte e nos impactos dos resultados gerados pelos algoritmos (Kaufman, 2021, p.80). Em outras palavras, procura-se verificar se (e como) as regras e

princípios são aplicados nos sistemas e processos da TRF. Não bastasse, também busca analisar como o algoritmo foi desenvolvido e quais foram os resultados obtidos, a fim de assegurar que não sejam discriminatórios (Magrani *et al.* 2021a, p.49).

É fundamental que a auditoria seja realizada de forma independente e diga respeito a todos os elementos relacionados ao funcionamento da TRF de forma geral, condições de operação, dados de treinamento e da galeria, e código. Dessa forma, caso existam vieses no processamento algorítmico, a sua constatação e correção será facilitada (Pires, 2021, p.55). No entanto, é preciso destacar que a auditoria não deve ser encarada como o único (e nem o principal) instrumento de adequação jurídica da TRF, mas sim como mais uma das boas práticas que devem ser adotadas a fim de se atingir o uso ético responsável e confiável da TRF (Magrani *et al.*, 2021a, p.49).

Para que seja viável e eficaz, a auditoria deve: ser regular e contínua; fazer parte do sistema sociotécnico como um todo (e não algo isolado); não representar um processo estático e pré-concebido, mas sim um processo dialético; estar de acordo com os incentivos e as políticas organizacionais; fornecer *feedback* ativo a fim de que a interpretabilidade e a robustez sejam incorporadas à tecnologia desde o início (Mokander e Floridi, 2021, p.3).

Outra medida que pode contribuir para a concretização da não discriminação *by design* é a inclusão nos contratos públicos de requisitos e especificações técnicas a serem cumpridos pelos desenvolvedores, fornecedores e usuários (autoridades policiais) (Pires, 2021, p.56). São exemplos de requisitos que podem ser incluídos: requisitos específicos de precisão e não discriminação a serem alcançados pela TRF comprada; deveres e obrigações, tanto para os fornecedores quanto para os compradores, no sentido de respeitarem os direitos fundamentais (dentre eles, a não discriminação) (FRA, 2019b, p.34); garantia de que a TRF foi desenvolvida para um propósito e contexto específico planejado pela autoridade policial; possibilidade de que a TRF possa ser adaptada facilmente ao propósito e contexto planejados pela autoridade policial; obrigação de que os Estados adquirentes da TRF supervisionem o projeto da tecnologia e exijam que o fornecedor identifique e elimine os vieses algorítmicos; obrigação de que os desenvolvedores e fornecedores realizem teste regulares de precisão e parcialidade enquanto durar o contrato (Garvie *et al.*, 2016, p.68; Pires, 2021, p.57).

Por último, também pode ser destacada como medida para a concretização da não discriminação *by design* a elaboração de relatórios detalhados sobre a criação da TRF e quais os resultados obtidos. Com efeito, a elaboração e manutenção de relatórios e

documentos atualizados é essencial para o sucesso de futuras auditorias e para que sejam cumpridas as exigências de accountability (Magrani *et al.*, 2021a, p.51).

Diante de todo o exposto, é preciso destacar que a lista de medidas concretizadoras do princípio da não discriminação *by design* não é uma lista exaustiva. De fato, outras medidas concretas podem ser agregadas a fim de contribuir para que o direito de não discriminação seja respeitado em todo o ciclo de vida da TRF. Além disso, um princípio de não discriminação *by design* não pretende invalidar toda a legislação já existente na UE em relação à transparência, proteção de dados e proteção dos direitos fundamentais. Pelo contrário, o intuito é justamente complementar tais legislações, constituindo-se como um instrumento de salvaguarda especificamente desenvolvido para a proteção do direito fundamental de não discriminação na TRF usada por autoridades policiais (Pires, 2021, p.57).

3.3. ANÁLISE COMPARATIVA ENTRE AS ABORDAGENS REGULATÓRIAS DA PROPOSTA DE REGULAMENTO DE IA DA COMISSÃO EUROPEIA E O PRINCÍPIO DE NÃO DISCRIMINAÇÃO *BY DESIGN*

Analisando as particularidades de cada abordagem apresentada, é possível traçar pontos comuns e distintos quando se trata de combater vieses em sistemas de reconhecimento facial usados por autoridades policiais. Nesse sentido, um primeiro ponto de diferença encontrado diz respeito ao tipo de TRF regulada em cada abordagem. Enquanto a Proposta classifica como de alto risco apenas os sistemas de identificação biométrica remota, o princípio não faz essa diferenciação (classifica como de alto risco os sistemas remotos e presenciais). Ocorre, no entanto, que a TRF operada presencialmente por autoridades policiais apresenta os mesmos riscos de enviesamento de uma TRF operada remotamente, e pode ser tão lesiva quanto em relação aos direitos fundamentais. Por isso, é importante que todos os tipos de TRF (sejam eles operados remotamente ou presencialmente) sejam considerados como de alto risco e estejam submetidos a rigorosos requisitos de conformidade.

Apesar disso, como se percebe dos Considerandos 4, 11, 28, 33 e 50 da Proposta, esta também se preocupa em grande medida com a questão dos vieses discriminatórios em sistemas de IA. Tal preocupação é traduzida principalmente nos dispositivos que estabelecem os requisitos obrigatórios para os sistemas de IA de alto risco. Sendo assim, são pontos comuns entre a Proposta e o princípio de não discriminação *by design* as

disposições relativas à documentação técnica. Com efeito, o art. 11 da Proposta, assim como o princípio, impõe a obrigação de ser elaborada documentação técnica antes da IA de alto risco ser colocada no mercado, e os parágrafos do Anexo IV da Proposta estabelecem as informações que devem constar em tais documentos. São exemplos: descrição geral do sistema de IA; descrição pormenorizada dos elementos do sistema de IA e do respectivo processo de desenvolvimento e testes; informações sobre o acompanhamento, funcionamento e controle do sistema de IA, especialmente em relação às suas capacidades e limitações de desempenho e níveis de precisão; resultados previsíveis e fontes de risco para os direitos fundamentais; medidas de supervisão humana; dados usados.

Além disso, o art. 12 da Proposta, assim como o princípio, determina que as capacidades de registro de tais sistemas devem proporcionar, no mínimo, registro do período de cada utilização do sistema (data e hora de início e data e hora de fim de cada utilização); a base de dados de referência relativamente à qual os dados de entrada foram verificados pelo sistema; os dados de entrada cuja pesquisa conduziu a uma correspondência; a identificação das pessoas singulares envolvidas na verificação dos resultados.

Por sua vez, o art. 13 e o Considerando 47 da Proposta se aproximam da ideia de transparência do princípio de não discriminação *by design*, na medida em que determinam que a IA de alto risco deve ser concebida e desenvolvida de forma que o seu funcionamento seja suficientemente transparente para permitir aos utilizadores interpretar o resultado do sistema e utilizá-lo corretamente. Além disso, tais sistemas devem ser acompanhados de instruções de utilização que incluam informações concisas, completas, corretas e claras (inclusive sobre possíveis riscos para os direitos fundamentais), acessíveis e compreensíveis para os utilizadores.

Outro ponto comum entre a Proposta e o princípio de não discriminação *by design* é a realização de supervisão interna e independente durante todas as fases da TRF. Especificamente na Proposta, esta exigência aparece no art. 9, com a necessidade de ser estabelecido um sistema de gestão de riscos. De forma complementar, o art. 43(1) da Proposta permite que, em alguns casos, a avaliação de conformidade do sistema de identificação biométrica remota seja realizada de forma independente por terceiros.

Além disso, o art. 9(2) e (4) da Proposta, à semelhança da avaliação de impacto dos direitos fundamentais (prevista no princípio de não discriminação *by design*), prevê que o sistema de gestão de riscos deve consistir num processo iterativo contínuo,

executado ao longo de todo o ciclo de vida de um sistema de IA de risco elevado, o que requer atualizações regulares sistemáticas. Além disso, ao se identificar as medidas de gestão de riscos mais apropriadas, deve-se assegurar a eliminação ou redução de riscos tanto quanto possível; adoção de medidas de atenuação e controlo adequadas em relação a riscos que não possam ser eliminados; prestação de informações adequadas e, se for o caso, formação dos utilizadores.

O art. 61(2) da Proposta, reforçado pelo Considerando 78, prevê, também à semelhança da medida prevista no princípio de não discriminação *by design*, que o sistema de acompanhamento pós-comercialização deve recolher, documentar e analisar, de forma ativa e sistemática, os dados sobre o desempenho dos sistemas de IA de alto risco ao longo da sua vida útil, bem como permitir ao fornecedor avaliar de forma contínua a conformidade dos sistemas de IA com os requisitos obrigatórios.

Por sua vez, o art. 33 da Proposta, assim como no princípio de não discriminação *by design*, determina que na supervisão realizada por terceiros, é preciso que os organismos notificados sejam independentes, imparciais, e tenham competência para avaliar aspectos técnicos e legais. Além disso, de acordo com os parágrafos 4 e 5 do Anexo VII da Proposta, a fim de que o sistema de gestão de qualidade tenha a sua manutenção e aplicação assegurados, o terceiro independente deve auditar periodicamente o fornecedor.

Apesar das semelhanças acima destacadas, a Proposta e o princípio de não discriminação *by design* apresentam diferenças em relação ao fornecimento de informações para a sociedade como um todo. Explica-se: por um lado, o art. 13 da Proposta prevê a obrigação de transparência e de prestação de informações apenas aos utilizadores dos sistemas de IA de alto risco. Dessa forma, falha a Proposta ao não prever que as informações devam ser prestadas também a toda a sociedade, a fim de que todos possam ter acesso a informações claras e completas sobre os riscos decorrentes da utilização de TRF por autoridades policiais. Por outro lado, a não discriminação *by design* exige que toda a sociedade seja informada sobre os riscos relacionados ao uso de TRF.

Outra diferença entre a Proposta e o princípio relaciona-se ao treinamento dos utilizadores (autoridades policiais) da TRF. Com efeito, a Proposta não prevê a obrigação de treinamento dos indivíduos responsáveis pela verificação do sistema. Tão somente menciona de forma vaga o treinamento dos utilizadores no seu art. 9(4) (em relação ao sistema de gerenciamento de risco). Por sua vez, o princípio de não discriminação *by design* determina que é necessário que as autoridades policiais estejam plenamente

treinadas e preparadas (em relação a questões técnicas e jurídicas) para utilizar a TRF de forma segura, justa e responsável.

Também pode ser destacada como uma diferença entre a Proposta e o princípio a exigência de diversidade na força de trabalho de todos os atores envolvidos (desenvolvedores, fornecedores e utilizadores). Enquanto o art. 69(2) e o Considerando 81 da Proposta consideram a diversidade das equipes de desenvolvimento apenas como algo a ser cumprido de forma voluntária (sem qualquer obrigatoriedade), o princípio da não discriminação *by design* considera tal diversidade primordial para que os vieses discriminatórios sejam combatidos nos algoritmos.

Outra questão que diferencia a Proposta e o princípio de não discriminação *by design* é a que diz respeito ao requisito de precisão dos sistemas de IA de alto risco. Embora exista uma exigência de precisão no art. 15 da Proposta, tal exigência não especifica a necessidade de os testes de precisão serem contextualizados em relação ao conjunto de dados usados. Já o princípio da não discriminação *by design*, por sua vez, leva em consideração essa diferença de representatividade dos diferentes grupos demográficos (nos conjuntos de dados de treinamento), e exige que a TRF tenha um alto nível de precisão para todos os grupos demográficos.

Por fim, é possível destacar como ponto comum entre a Proposta e o princípio de não discriminação *by design* a questão relacionada à alta qualidade e diversidade dos dados. Com efeito, à semelhança do que prevê o princípio, o art. 10 da Proposta, reforçado pelo Considerando 44, estabelece que os sistemas de IA de alto risco que utilizem técnicas que envolvam o treino de modelos com dados devem ser desenvolvidos com base em conjuntos de dados de treino, validação e teste que cumpram rigorosos critérios de qualidade. Além disso, o dispositivo estabelece que conjuntos de dados de treino, validação e teste devem ser pertinentes, representativos, isentos de erros e completos. Devem ter as propriedades estatísticas adequadas, nomeadamente, quando aplicável, no tocante às pessoas ou grupos de pessoas em que o sistema de IA de alto risco se destina a ser utilizado.

O art. 10(4) da Proposta determina ainda a necessidade dos conjuntos de dados de treino, validação e teste levarem em consideração as características e elementos que são idiossincráticos do enquadramento geográfico, comportamental ou funcional específico no qual o sistema de IA de risco elevado se destina a ser utilizado. Já o art. 10(5) estabelece que, no intuito de assegurar o controle, a detecção e a correção de enviesamentos em relação a sistemas de IA de alto risco, os fornecedores desses sistemas

podem tratar categorias especiais de dados pessoais a que se refere o art. 9(1), GDPR, art. 10, LED e o art. 10(1), do Regulamento (UE) 2018/1725, desde que assegurem salvaguardas adequadas aos direitos fundamentais e liberdades das pessoas. Tais salvaguardas podem incluir impor limitações técnicas à reutilização e utilizar medidas de segurança e preservação da privacidade de última geração, tais como a pseudonimização ou a cifragem, nos casos em que a anonimização possa afetar significativamente a finalidade preconizada. Por sua vez, o art. 10(6) da Proposta estabelece a necessidade de aplicação de práticas adequadas de governação e gestão de dados ao desenvolvimento de sistemas de IA de risco elevado, mesmo que estes não utilizem técnicas de treinamento de modelos.

Diante de todo o exposto neste capítulo, é possível concluir que tanto a abordagem regulatória da Proposta, quanto a abordagem regulatória do princípio de não discriminação *by design*, apresentam aspectos importantes que contribuem para a minimização das tensões e lacunas existentes na legislação antidiscriminação da UE e, consequentemente, para a resolução eficaz do problema do enviesamento no uso de TRF por autoridades policiais.

Por um lado, a Proposta publicada pela Comissão representa um passo sólido em relação à correta regulação de diferentes sistemas de IA e da proteção de direitos fundamentais frente aos possíveis vieses algorítmicos. Embora contenha alguns elementos principiológicos, a Proposta opta por representar uma solução regulatória baseada em regras detalhadas. Se por um lado isso pode trazer desvantagens como, por exemplo, menor flexibilidade e durabilidade das normas, e a possibilidade dos regulados cumprirem a literalidade da norma, mas não a sua finalidade, por outro podem trazer claros benefícios, tais como, por exemplo, maior clareza e facilidade de aplicação uniforme, maior certeza e previsibilidade, redução da arbitrariedade dos reguladores, menor abertura à interpretação (por capacitarem os supervisores e os sistemas de conformidade), proibição de conformidade mínima por regulações mal-intencionadas. É possível perceber, portanto, que a Proposta aborda muitos dos pontos também abordados pelo princípio da não discriminação *by design* e, ao mesmo tempo, representa um instrumento apto a mitigar os pontos negativos deste princípio.

Por outro lado, tendo em vista que o princípio de não discriminação *by design* é desenvolvido para regular um ponto específico numa tecnologia específica, qual seja, os vieses discriminatórios nos algoritmos de TRF usada por autoridades policiais, sua utilização representa um instrumento eficiente para prever particularidades e proteger o

direito fundamental de não discriminação. Assim como a Proposta, a abordagem do princípio possui vantagens e desvantagens. Por um lado, pode causar menor certeza e previsibilidade (pela ausência de requisitos específicos e detalhados); e pode gerar maior arbitrariedade dos regulados, (o que pode causar interpretações muito conservadoras ou o cumprimento mínimo por parte dos regulados). Já por outro, os princípios são mais flexíveis, duráveis e inclusivos, além de forçarem mais os regulados a atingirem o verdadeiro propósito da norma.

É possível perceber, portanto, que as abordagens da Proposta e do princípio se complementam: uma acerta onde a outra deixa a desejar, e vice-versa. Dessa forma, é possível que o regulador europeu desenvolva um sistema regulatório no qual ambos estão presentes: o princípio da não discriminação *by design* estabelecendo as orientações regulamentares gerais, e as regras detalhadas da Proposta fazendo o complemento com normas claras e precisas. Ao dialogarem um com o outro, é possível mitigar as lacunas e tensões existentes na legislação europeia antidiscriminação de forma eficaz.

Nem mesmo o fato de tais opções regulatórias representarem níveis diferentes de regulamentação apresenta-se como um impeditivo para que atuem de forma complementar. Os pontos em comum expostos no tópico anterior demonstram que as duas abordagens trilham caminhos no mesmo sentido e podem contribuir para a melhoria recíproca. Além disso, nada impede que um princípio de não discriminação *by design* seja incorporado à Proposta (à semelhança do que foi feito com a privacidade *by design* no GDPR), ampliando a aplicação do princípio a todos os tipos de sistemas de IA regulados pela Proposta. De momento, no entanto, dado que a Proposta ainda não está em vigor, e dado que o princípio ainda se encontra no seu estágio inicial de concepção, é preciso que reguladores e regulados mantenham o pensamento aberto a fim de considerarem as duas abordagens como formas legítimas e complementares de se alcançar a mitigação de vieses discriminatórios de raça e de gênero no uso de TRF por autoridades policiais. Tais instrumentos, juntamente com outros já existentes, podem gerar um arcabouço jurídico eficaz no sentido de incutir uma cultura de proteção ao direito de não discriminação nas estruturas e políticas de todos os atores envolvidos na criação e no uso das TRFs. Dessa forma, é possível caminhar no sentido de tornar mais confiável a TRF usada para fins de manutenção da ordem pública.

CONCLUSÃO

O rápido desenvolvimento da inteligência artificial nos últimos anos (motivado pelo aprimoramento de modelos teóricos, melhoria da capacidade computacional e maior disponibilidade de dados), juntamente com a redução do custo de outras tecnologias digitais, fez com que a tecnologia de reconhecimento facial (TRF) entrasse nas sociedades sem um debate profundo e um verdadeiro controle democrático. Um de seus usos que mais vem ganhando destaque, e gerando preocupação, é o realizado pelas autoridades policiais para fins de manutenção da ordem pública, seja em tempo real ou em tempo diferido (para investigações criminais).

O grande problema que vem sendo apontado em relação ao uso da TRF por autoridades policiais diz respeito aos vieses de gênero e de raça presentes na tecnologia, que podem levar as autoridades policiais a identificarem incorretamente pessoas de determinados grupos demográficos e, com isso, violarem os direitos fundamentais (como o de não discriminação) de tais pessoas, tornando a tecnologia um instrumento de perpetuação de preconceitos e opressões históricas. Em outras palavras, ao invés de se configurar num meio de manutenção da ordem pública, a TRF nada mais seria do que um meio de manutenção do *status quo*, que aprofunda e reproduz em larga escala perseguições a grupos demográficos já historicamente marginalizados.

No entanto, apesar dos diferentes receios e posicionamentos existentes nas instituições e na sociedade civil, é possível perceber que o uso da TRF na UE avança a passos largos. Isso indica que a UE precisará realizar rapidamente uma abordagem regulatória harmonizada se quiser promover o seu desenvolvimento tecnológico sem comprometer os seus valores primordiais e direitos fundamentais (como a não discriminação). Caso não o faça, a UE arrisca permitir a proliferação de legislações nacionais diferentes e contraditórias, que por sua vez poderão impedir a livre circulação de bens e serviços relacionados à TRF; bem como impedir o estabelecimento de um elevado nível de proteção dos direitos fundamentais comum a todos os países membros da UE, no que diz respeito ao uso da TRF por autoridades policiais.

Sendo assim, a fim de definir qual seria a melhor abordagem regulatória para combater os vieses discriminatórios na TRF usada por autoridades policiais, o presente trabalho procurou, num primeiro momento, entender como funciona a TRF, quais são os seus principais usos pelas autoridades policiais e quais são os seus principais desafios. Como consequência, descobriu-se que: a precisão da TRF varia em relação a grupos

demográficos distintos; os grupos mais afetados são as mulheres e as pessoas negras (em especial, mulheres negras); e os vieses são causados por questões como a qualidade e diversidade (demográfica) dos dados usados, erros na codificação dos algoritmos e preconceitos dos desenvolvedores e usuários. Como solução, destaca-se a necessidade de haver maior diversidade nos dados usados para treinar e TRF, bem como que haja um controle sobre os designers e programadores, a fim de que eles hajam de forma ética e em favor dos direitos humanos desde a concepção da TRF.

Levantadas tais questões, foi feita uma análise da atual legislação antidiscriminação da UE e ficou constatado que tal legislação não é suficiente para lidar com sistemas de IA e suas formas sutis e não intuitivas de discriminação. Primeiro, porque a abordagem da justiça individual não é forte o suficiente para combater a discriminação algorítmica (visto que uma série de problemas impedem as pessoas de iniciar uma ação legal para proteger seu direito à igualdade). Segundo, porque a legislação secundária antidiscriminação não é clara (o que gera insegurança para os demandantes). Terceiro, porque a lei de não discriminação da UE permite justificação fácil de algumas formas de discriminação algorítmica (principalmente a discriminação indireta). E quarto, porque a legislação secundária antidiscriminação da UE apresenta âmbito de aplicação limitado e alguns requisitos com limiares desproporcionalmente altos para que a proteção seja conferida.

Ciente dos problemas existentes, o presente trabalho focou na análise de duas abordagens regulatórias específicas, a fim de descobrir qual delas é a melhor opção para minimizar as tensões e lacunas existentes na legislação antidiscriminação da UE, e resolver de forma eficaz o problema do enviesamento no uso de TRF por autoridades policiais.

Uma das abordagens é a Proposta de Regulamento de IA da Comissão Europeia, que adota uma regulação geral baseada no risco, aplicável à maioria dos sistemas de IA (dentre eles, a TRF). Já a outra abordagem é o “princípio de não discriminação *by design*” (ou igualdade *by design*), que estabelece orientações regulamentares que devem ser implementadas por desenvolvedores, fornecedores e usuários em todas as etapas de desenvolvimento de uma TRF usada por autoridades policiais, a fim de que sejam evitados ou minimizados vieses de gênero e de raça.

Tanto a abordagem regulatória da Proposta, quanto a abordagem regulatória do princípio, apresentam aspectos importantes que contribuem para a minimização das tensões e lacunas existentes na legislação antidiscriminação da UE e, consequentemente,

para a resolução eficaz do problema do enviesamento no uso de TRF por autoridades policiais.

Por um lado, a Proposta publicada pela Comissão representa um passo sólido na direção da correta regulação de diferentes sistemas de IA e da proteção de direitos fundamentais frente aos possíveis vieses algorítmicos. Embora contenha alguns elementos principiológicos, a Proposta opta por representar uma solução regulatória baseada em regras detalhadas. Se por um lado isso pode trazer desvantagens como, por exemplo, menor flexibilidade e durabilidade das normas, maior restrição à discricionariedade dos entes regulados, menor campo de aplicação, e a possibilidade dos regulados cumprirem a literalidade da norma, mas não a sua finalidade; por outro, podem trazer claros benefícios, tais como, por exemplo, maior clareza e facilidade de aplicação uniforme, maior certeza e previsibilidade, redução da arbitrariedade dos reguladores e regulados, menor abertura à interpretação (por capacitarem os supervisores e os sistemas de conformidade), proibição de conformidade mínima por regulações mal-intencionadas. É possível perceber, portanto, que a Proposta aborda muitos dos pontos também abordados pelo princípio da não discriminação *by design* e, ao mesmo tempo, representa um instrumento apto a mitigar os pontos negativos deste princípio.

Por outro lado, tendo em vista que o princípio de não discriminação *by design* é desenvolvido para regular um ponto específico numa tecnologia específica, qual seja, os vieses discriminatórios nos algoritmos de TRF usada por autoridades policiais, sua utilização representa um instrumento eficiente para prever particularidades e proteger o direito fundamental de não discriminação. Assim como a Proposta, a abordagem do princípio possui vantagens e desvantagens. Por um lado, pode causar menor certeza e previsibilidade (pela ausência de requisitos específicos e detalhados); e pode gerar maior arbitrariedade dos regulados, (o que pode causar interpretações muito conservadoras ou o cumprimento mínimo por parte dos regulados). Já por outro, os princípios são mais flexíveis, duráveis e inclusivos, além de forçarem mais os regulados a atingirem o verdadeiro propósito da norma.

É possível perceber, portanto, que as abordagens da Proposta e do princípio se complementam: um acerta onde o outro deixa a desejar, e vice-versa. Dessa forma, é possível que o regulador europeu use isso em seu benefício e desenvolva um sistema regulatório no qual ambos estão presentes: o princípio da não discriminação *by design* estabelecendo as orientações regulamentares gerais, e as regras detalhadas da Proposta fazendo o complemento com normas claras e precisas. Ao dialogarem um com o outro, é

possível mitigar as lacunas e tensões existentes na legislação europeia antidiscriminação de forma eficaz.

Nem mesmo o fato de tais opções regulatórias representarem níveis diferentes de regulamentação apresenta-se como um impeditivo para que atuem de forma complementar. Com efeito, os pontos em comum entre as duas abordagens demonstram que ambas caminham na mesma direção e podem contribuir para a melhoria recíproca. Além disso, nada impede que um princípio de não discriminação *by design* seja incorporado à Proposta (à semelhança do que foi feito com a privacidade *by design* no GDPR), ampliando a aplicação do princípio a todos os tipos de sistemas de IA regulados pela Proposta. De momento, no entanto, dado que a Proposta ainda não está em vigor, e dado que o princípio ainda se encontra no seu estágio inicial de concepção, é preciso que reguladores e regulados mantenham o pensamento aberto a fim de considerarem as duas abordagens como formas legítimas e complementares de se alcançar a mitigação de vieses discriminatórios de raça e de gênero no uso de TRF por autoridades policiais. Tais instrumentos, juntamente com outros já existentes, podem gerar um arcabouço jurídico eficaz no sentido de tornar mais confiável a TRF usada para fins de manutenção da ordem pública.

Uma abordagem regulamentar que aplicasse conjuntamente a Proposta e o princípio de não discriminação *by design*, se basearia em medidas como, por exemplo: diversificação dos conjuntos de dados e da força de trabalho; treinamento especial para equipes de desenvolvimento dos algoritmos e utilizadores finais; requisitos de transparência e compras públicas; consulta às partes interessadas; prestação de informações aos utilizadores e à sociedade como um todo; supervisão humana interna e independente; estabelecimento de um sistema de gestão de riscos iterativo e contínuo; alta qualidade dos dados e práticas adequadas de governança de dados; elaboração e atualização de documentação técnica detalhada; elaboração e manutenção de registros; alto nível de precisão, solidez e cibersegurança; realização de uma avaliação de conformidade; realização de acompanhamento pós-comercialização pelos fornecedores.

Ao se aplicar uma solução regulatória que combine as duas abordagens, todos saem ganhando: sociedade, indivíduos, desenvolvedores, fornecedores e utilizadores. O presente trabalho entende nesse sentido e visa contribuir para um debate vibrante e democrático de modo a encontrar a melhor solução regulatória. De forma alguma se tentou esgotar o tema. Sendo assim, se uma nova tecnologia traz benefícios para o ser humano, é preciso que o seu uso seja considerado, debatido e avaliado de forma

responsável e democrática. Isto significa que o uso de tal tecnologia não pode ser deixado aos caprichos de indivíduos e organizações específicas com interesses particulares, e muitas vezes obscuros. Deve haver um amplo e profundo debate com indivíduos de vários setores a fim de se chegar a um consenso sobre os princípios norteadores do desenvolvimento dessa tecnologia.

Nessa esteira, é preciso destacar que qualquer tecnologia que surja deve ser utilizada para o benefício e desenvolvimento do ser humano, seja individual ou coletivo. Sendo assim, o desenvolvimento da TRF deve ser um vetor de sociedades equitativas, colocando-se a serviço dos seres humanos e dos direitos fundamentais, sem restringir ou desviar a autonomia humana. O desenvolvimento adequado da TRF é uma responsabilidade de nossa geração para consigo mesma e, acima de tudo, para com as gerações futuras.

REFERÊNCIAS

ADEE, Sally - *Controversial software claims to tell personality from your face* (2016). Disponível em: [more: https://www.newscientist.com/article/2090656-controversial-software-claims-to-tell-personality-from-your-face/](https://www.newscientist.com/article/2090656-controversial-software-claims-to-tell-personality-from-your-face/). Acesso em: 01 de maio de 2022.

ADLER, Andy; SCHUCKERS, Michael E. - *Comparing Human and Automatic Face Recognition Performance*. IEEE Transactions on Systems, Man, and Cybernetics. v.37, n.5, p. 1248-1255 (2007).

AMNESTY INTERNATIONAL; ACCESS NOW - *The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems* (2018).

ARTICLE 29 DATA PROTECTION WORKING PARTY - *Opinion 02/2012 on facial recognition in online and mobile services* (WP 192: 2012).

ARTICLE 29 DATA PROTECTION WORKING PARTY - *Opinion 3/2012 on developments in biometric technologies* (WP 193: 2012).

AWS - *Amazon Rekognition Developer Guide* (2020). Disponível em: <https://docs.aws.amazon.com/rekognition/latest/dg/rekognition-dg.pdf>. Acesso em: 29 de abril de 2022.

BENNETT, Kanya A - *Can Facial Recognition Technology Be Used to Fight the New Way against Terrorism: Examining the Constitutionality of Facial Recognition Surveillance Systems*. NC JL & Tech. v. 3, p. 151-174 (2001).

BLACK, Julia; HOPPER, Martyn; BAND, Christa - *Making a success of principles-based regulation*. Law and financial markets review, v. 1, n. 3, p. 191-206 (2007).

BLACK, Julia - *The Rise, Fall and Fate of Principles Based Regulation*. LSE Law, Society and Economy Working Paper. v. 17 (2010). Disponível em: <http://ssrn.com/abstract=1712862>. Acesso em: 18 de maio de 2022.

BUOLAMWINI, Joy; GEBRU, Timnit - *Gender shades: Intersectional accuracy disparities in commercial gender classification*. Conference on fairness, accountability and transparency (2018). Disponível em: <https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>. Acesso em: 30 de abril de 2022.

BUOLAMWINI, Joy - *Response: Racial and Gender bias in Amazon Rekognition – Commercial AI System for Analyzing Faces*. Medium (2019). Disponível em: <https://medium.com/@Joy.Buolamwini/response-racial-and-gender-bias-in-amazon-rekognition-commercial-ai-system-for-analyzing-faces-a289222eeced>. Acesso em: 29 de abril de 2022.

BURGESS, Matt - *Europe Is Building a Huge International Facial Recognition System*. Wired (2022). Disponível em: <https://www.wired.com/story/europe-police-facial-recognition-prum/>.

CATH, Corinne [et al.] - *Artificial Intelligence and the ‘Good Society’: the US, EU, and UK approach*. Science and Engineering Ethics. v. 24, p. 505–528 (2018). Disponível em: <https://doi.org/10.1007/s11948-017-9901-7>. Acesso em: 01 de maio de 2022.

CAVOUKIAN, Ann - *Privacy by Design’: The 7 foundational principles*. Information and Privacy Commissioner of Ontario, Canada. v. 5 (2009).

CHESNEY, Bobby; CITRON, Danielle - *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*. California Law Review, v. 107, p.1753-1820 (2019). Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954. Acesso em: 01 de maio de 2022.

CHRISTAKIS, Theodore; BECUYWE, Mathias - *Pre-Market Requirements, Prior Authorisation and Lex Specialis: Novelties and Logic in the Facial Recognition-Related Provisions of the Draft AI Regulation*. European Law Blog (2021). Disponível em: <https://europeanlawblog.eu/2021/05/04/pre-market-requirements-prior-authorisation-and-lex-specialis-novelties-and-logic-in-the-facial-recognition-related-provisions-of-the-draft-ai-regulation/>. Acesso em: 17 de maio de 2022.

CNIL – Commission Nationale de l'Informatique et des Libertés - *Facial Recognition - for a Debate Living up to the Challenges*, 2019.

COMISSÃO EUROPEIA - *Inteligência Artificial para Europa*. COM. p. 237 (2018).

COMISSÃO EUROPEIA - *White Paper on Artificial Intelligence - A European approach to excellence and trust*. COM. p. 65 (2020).

COMISSÃO EUROPEIA - *Police Cooperation Code: Questions and Answers* (2021). Disponível em: https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_6646. Acesso em: 02 de maio de 2022.

CONSELHO DA EUROPA - *Comitê Consultivo da Convenção para a proteção dos indivíduos com relação ao processamento automático de dados pessoais: Diretrizes sobre Reconhecimento Facial* (2021). Disponível em: <https://edoc.coe.int/en/artificial-intelligence/9753-guidelines-on-facial-recognition.html>. Acesso em: 03 de maio de 2022.

CONSELHO DA EUROPA (2018) - *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (entered into force 1 October 1985) ETS no 108, as amended by Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*. CETS no 223 (Convention 108+).

CONSELHO DA EUROPA (1950) - *Convenção para a Proteção dos Direitos Humanos e Liberdades Fundamentais* (entrada em vigor em 3 de setembro de 1953) ETS No.005 (ECHR).

CONSELHO DA EUROPA (2005) - *PROTOCOLO Nº 12 da Convenção para a Proteção dos Direitos Humanos e Liberdades Fundamentais* (entrou em vigor em 1 de abril de 2005) ETS No.177.

CONSELHO DA EUROPA (2018) - *The Practical Guide on the Use of Personal Data in the Police Sector*. Disponível em: <https://rm.coe.int/tpd-201-01-practical-guide-on-the-use-of-personal-data-in-the-police-/16807927d5>. Acesso em: 04 de maio de 2022.

CONSELHO DA EUROPA (2000) - *Diretiva 2000/43/CE do Conselho, de 29 de junho de 2000, que aplica o princípio da igualdade de tratamento entre as pessoas, sem distinção de origem racial ou étnica*. Jornal Oficial nº L 180 de 19/07/2000 p. 0022 – 0026. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32000L0043&from=EN>. Acesso em 10 de maio de 2022.

CONSELHO DA UNIÃO EUROPEIA (2020) - *Conclusões da Presidência - A Carta dos Direitos Fundamentais no contexto da Inteligência Artificial e da Mudança Digital* 11481/20. Disponível em: <https://www.consilium.europa.eu/media/46496/st11481-en20.pdf>. Acesso em: 12 de maio de 2022.

CRENSHAW, Kimberlé - *Documento para o encontro de especialistas em aspectos da discriminação racial relativos ao gênero*. Revista de Estudos Feministas. v. 7, n. 12, p. 171-188 (2002). Disponível em: <https://www.scielo.br/j/ref/a/mbTpP4SFXPnJZ397j8fSBQQ/?format=pdf&lang=pt>. Acesso em: 01 de maio de 2022.

DANKS, David; LONDON, Alex John - *Algorithmic Bias in Autonomous Systems*. IJCAI, v. 17, p. 4691-4697 (2017). Disponível em: <https://www.cmu.edu/dietrich/philosophy/docs/london/IJCAI17-AlgorithmicBias-Distrib.pdf>. Acesso em: 30 de abril de 2022.

DROZDOWSKI, Pawel [et al.] - *Demographic Bias in Biometrics: A Survey on an Emerging Challenge*. IEEE Transactions on Technology and Society, v. 1, n. 2, p. 89-103 (2020).

EDPB – European Data Protection Board. *Guidelines 4/2019 on Article 25 Data Protection by Design and by Default* (2020a).

EDPB – European Data Protection Board. *Guidelines 3/2019 on processing of personal data through video devices* (2020b).

EDPB – European Data Protection Board. *EDPB & EDPS Call for Ban on Use of AI for Automated Recognition of Human Features in Publicly Accessible Spaces, and Some Other Uses of AI That Can Lead to Unfair Discrimination* | European Data Protection Board (2021). Disponível em: <https://edpb.europa.eu/news/news/2021/edpb-edps-call>

ban-use-ai-automated-recognition-human-features-publicly-accessible_en. Acesso em: 03 de maio de 2022.

EDPS – European Data Protection Supervisor. *Facial recognition: A solution in search of a problem?* (2019). Disponível em: https://edps.europa.eu/press-publications/press-news/blog/facial-recognition-solution-search-problem_en. Acesso em: 11 de maio de 2022.

EDPS – European Data Protection Supervisor. *Shaping a Safer Digital Future: a New Strategy for a New Decade* (2020). Disponível em: https://edps.europa.eu/press-publications/publications/strategy/shaping-safer-digital-future_en. Acesso em: 03 de maio de 2022.

EDRI. - *Ban Biometric Mass Surveillance!* (2020). Disponível em: <https://edri.org/our-work/blog-ban-biometric-mass-surveillance/>.

FERNANDEZ, Valérie [et al.] - *Facial Recognition: Embodying European Values*. Paris: Renaissance Numérique (2020). Disponível em: https://www.renaissancenumerique.org/ckeditor_assets/attachments/548/report_facial_recognition.pdf. Acesso em: 04 de maio de 2022.

FIEUX-CASTAGNET, Geneviève; SANTUCCI, Gerald - *An Ethical Facial Recognition: An Oxymoron?* (2020). Disponível em: <https://www.theinternetofthings.eu/sites/default/files/docs/AN%20ETHICAL%20FACIAL%20RECOGNITION%2021.04.20.pdf>. Acesso em: 25 de abril de 2022.

FONTES, Ana Catarina; LÜTGE, Cristoph - *Vigilância e Relações de Poder – O Uso de Tecnologias de Reconhecimento Facial e Identificação Biométrica a Distância em Espaço Público e Impactos na Vida Pública*. *Direito Público*. [S. l.], v. 18, n. 100 (2022). doi: 10.11117/rdp.v18i100.6203. Disponível em: <https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/6203>. Acesso em: 22 de abril de 2022.

FORD, Cristie L - *New Governance, Compliance and Principles-Based Securities Regulation*. *American Business Law Journal*. v. 45, n. 1, p.1-60 (2008).

FRA – European Union Agency for Fundamental Rights. *Fundamental rights and the interoperability of EU information systems: borders and security* (2017).

FRA – European Union Agency For Fundamental Rights. *European Union Agency for Fundamental Rights, Handbook on European non-discrimination law* (2018a).

FRA – European Union Agency For Fundamental Rights. *#BigData: Discrimination in data-supported decision making* (2018b).

FRA – European Union Agency For Fundamental Rights. *Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights* (2019a).

FRA – European Union Agency For Fundamental Rights. *Facial recognition technology: fundamental rights considerations in the context of law enforcement*. ISBN 978-92-9474-758-7, doi: 10.2811/52597 (2019b).

FRA – European Union Agency For Fundamental Rights. *Getting the future right - Artificial intelligence and fundamental rights* (2020).

FRANTZIOU, Eleni - *The Horizontal Effect of the Charter*. Cambridge Yearbook of European Legal Studies. v. 22, p. 208-232 (2020).

FRAZÃO, Ana - *Fundamentos da proteção de dados pessoais. Noções introdutórias para a compreensão da importância da Lei Geral de Proteção de Dados*. In: TEPEDINO, Gustavo; Frazão, Ana; Oliva, Milena D. (coord.). *Lei Geral de Proteção de Dados Pessoais e suas repercussões no Direito Brasileiro*. 1. ed. São Paulo: Ed. RT, 2019.

FRIEDMAN, Batya; NISSENBAUM, Helen - *Bias in computer systems*. ACM Transactions on Information Systems (TOIS). v. 14, n. 3, p. 330-347 (1996).

FRIEDMAN, Batya; KAHN, Peter H; BORNING, Alan - *Value Sensitive Design and Information Systems' in Himma K e Tavani H* (eds). *The Handbook of Information and Computer Ethics*. p. 69-101 (2008).

FUSSEY, Pete; MURRAY, Daragh - *Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology*. The Human Rights, Big Data and Technology Project, Londres: [s. l.] 2019. Disponível em: <http://repository.essex.ac.uk/24946/1/London-Met-Police-Trial-of-Facial-Recognition-Tech-Report-2.pdf>.

GARVIE, Clare - *The Untold Number of People Implicated in Crimes They Didn't Commit Because of Face Recognition*. ACLU, [s.l.] (2020). Acesso em: 28 de abril de 2022.

GARVIE, Clare; BEDOYA, Alvaro; FRANKLE, Jonathan - *The Perpetual Line-up: Unregulated Police Face Recognition in America*. Georgetown Law Center on Privacy & Technology (2016).

GERARDS, Janneke; XENIDIS, Raphaële - *Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law*. European Commission (2021).

GÉRON, Aurélien - *Mãos à Obra Aprendizado de Máquina com Scikit-Learn & TensorFlow: Conceitos, Ferramentas e Técnicas Para a Construção de Sistemas Inteligentes*; traduzido por Rafael Contatori. Rio de Janeiro: Alta Books, 2019.

GILLIOM, John; MONAHAN, Torin - *Supervision: An introduction to the surveillance society*. Chicago: The University of Chicago Press, 2013.

GONZALEZ FUSTER, Gloria - *Artificial Intelligence and Law Enforcement. Impact on Fundamental Rights* (PE 656.295). Brussels: European Parliament, 2020. [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU\(2020\)656295_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295_EN.pdf).

GORDON, Barrie J - *Automated Facial Recognition in Law Enforcement: The Queen (On Application of Edward Bridges) v The Chief Constable of South Wales Police*. PER / PELJ. v. 24 (2021). doi: <http://dx.doi.org/10.17159/1727-3781/2021/v24i0a8923>, 2021.

GROTHER, Patrick J.; QUINN, George W.; PHILLIPS, P. Jonathon - *Report on the Evaluation of 2D Still-Image Face Recognition Algorithms, NIST Interagency/Internal*

Report (NISTIR). National Institute of Standards and Technology, Gaithersburg, MD. 2010. Disponível em: <https://doi.org/10.6028/NIST.IR.7709>. Acesso em: 28 de abril de 2022.

GROTHER, Patrick J.; NGAN, Mei L.; HANAOKA, Kayee K - *Face Recognition Vendor Test Part 3: Demographic Effects, NIST Interagency/Internal Report (NISTIR)*. National Institute of Standards and Technology, Gaithersburg, MD. 2019. Disponível em: <https://doi.org/10.6028/NIST.IR.8280>. Acesso em: 28 de abril de 2022.

GUO, Guodong; ZHANG, Na - *A survey on deep learning based face recognition* Computer Vision and Image Understanding. v. 189 (2019).

HACKER, Philipp - *Teaching fairness to artificial intelligence*. Common Market Law Review. v. 55, n. 4, p. 1143-1185 (2018).

HAO, Karen - *This is how AI bias really happens – and why it’s so hard to fix*. MIT Technology Review (2019a). Disponível em: <https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/>. Acesso em: 30 de abril de 2022.

HAO, Karen - *An AI app that ‘undressed’ women shows how deepfakes harm the most vulnerable*. MIT Technology Review (2019b). Disponível em: <https://www.technologyreview.com/2019/06/28/134352/an-ai-app-that-undressed-women-shows-how-deepfakes-harm-the-most-vulnerable/>. Acesso em: 01 de maio de 2022.

HERN, Alex - *Google’s solution to accidental algorithmic racism: ban gorillas*. (2019). Disponível em: <https://www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorilla-black-people>. Acesso em: 01 de maio de 2022.

HLEGAI – High Level Expert Group on Artificial Intelligence - *A definition of AI: main capabilities and disciplines* (2019a).

HLEGAI – High Level Expert Group on Artificial Intelligence - *Ethics Guidelines for Trustworthy AI* (2019b).

HONG, Sun-há - *Technologies of Speculation: The Limits of Knowledge in a Data-Driven Society*. Nova York: NYU Press, 2020.

INTRONA, Lucas; NISSENBAUM, Helen - *Facial Recognition Technology a survey of policy and implementation issues*. Center for Catastrophe Preparedness and Response, New York University (2010). Disponível em: <https://ssrn.com/abstract=1437730>.

KAUFMAN, Dora - *Inteligência Artificial e os desafios éticos: a restrita aplicabilidade dos princípios gerais para nortear o ecossistema de IA*. PAULUS: Revista de Comunicação da FAPCOM. v. 5, n. 9, p. 73-84 (2021).

KIND, Carly - *Containing the canary in the AI coalmine – the EU's efforts to regulate biometrics*. Ada Lovelace Institute, 2021. Disponível em: <https://www.adalovelaceinstitute.org/blog/canary-ai-coalmine-eu-regulate-biometrics/>. Acesso em: 05 de maio de 2022.

KLARE, Brendan F [et al.] *Face Recognition Performance: Role of Demographic Information*. IEEE Transactions on Information Forensics and Security. v. 7, n. 6, p. 1789-1801 (2012). doi: 10.1109/TIFS.2012.2214212.

LAPIN – Laboratório de Políticas Públicas e Internet - *Vigilância automatizada: uso de reconhecimento facial pela Administração Pública*, 2021. Disponível em: <https://lapin.org.br/2021/07/07/vigilancia-automatizada-uso-de-reconhecimento-facial-pela-administracao-publica-no-brasil/>. Acesso em: 27 de abril de 2022.

LECUN, Yan, BENGIO, Yoshua; HINTON, Geoffrey E. - *Deep Learning*. Nature. v. 521, p. 436-444 (2015).

LEHR, David; OHM, Paul - *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*. Davis Law Review, University of California. v. 53, p. 653-717 (2017).

LESLIE, David - *Understanding bias in facial recognition technologies*. The Alan Turing Institute, 2020. Disponível em:

<https://www.turing.ac.uk/research/publications/understanding-bias-facial-recognition-technologies>.

MADIEGA, Tambiama; MILDEBRATH, Hendrik - *Regulating facial recognition in the EU*. European Parliamentary Research Service – EPRS, 2021.

MAGRANI, Eduardo - *Entre dados e robôs: ética e privacidade na era da hiperconectividade*. 2. ed. Porto Alegre: Arquipélago Editorial, 2019.

MAGRANI, Eduardo; OLIVEIRA, Renan Medeiros de; CAMPELLO, Tatiana - *Guia de boas práticas em inteligência artificial*. Demarest, 2021a.

MAGRANI, Eduardo - *Inteligência Artificial responsável. O que é isto?* Estadão, 2021. Disponível em: <https://politica.estadao.com.br/blogs/fausto-macedo/inteligencia-artificial-responsavel-o-que-e-isto/>. Acesso em: 21 de maio de 2022.

MCCARTHY, John - *What is artificial intelligence?* Computer Science Department Stanford University, 2000. Disponível em: <http://www-formal.stanford.edu/jmc/whatisai.pdf>. Acesso em: 23 de abril 2022.

MCCARTHY, John [et al.] - *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. AI Magazine. v. 27, n. 4, p. 12-14, (2006). doi: 10.1609/aimag.v27i4.1904. Disponível em: <https://ojs.aaai.org/index.php/aimagazine/article/view/1904>. Acesso em: 23 de abril 2022.

MCGREGOR, Lorna; MURRAY, Daragh; NG, Vivian - *International Human Rights Law as a Framework for Algorithmic Accountability*. ICLQ. v. 68, p.309-343 (2019). Disponível em: https://www.researchgate.net/publication/332457262_International_human_rights_law_as_a_framework_for_algorithmic_accountability. Acesso em: 30 de abril de 2022.

MENEGUS, Bryan - *Defense of Amazon's Face Recognition Tool Undermined by Its Only Known Police Client* Gizmodo. GIZMODO, 2019. Disponível em: <https://gizmodo.com/defense-of-amazons-face-recognition-tool-undermined-by-1832238149>. Acesso em: 29 de abril de 2022.

MOKANDER, Jacob; FLORIDI, Luciano - *Ethics-Based Auditing to Develop Trustworthy AI. Minds and Machines*. Online (2021). Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3788841. Acesso em: 22 de maio de 2022.

NEWSROOM - *Shaping Europe's Digital Future: What You Need to Know*. Modern Diplomacy, 2020. Disponível em: <https://moderndiplomacy.eu/2020/02/22/shaping-europes-digital-future-what-you-need-to-know/>. Acesso em: 04 de maio de 2022.

NORRIS, Clive - *From personal to digital CCTV, the panopticon, and the technological mediation of suspicion and social control*. In: LYON, David - *Surveillance as social sorting: privacy, risk, and digital discrimination*. Routledge: New York, 2003. p. 247-281.

NORRIS, Clive; ARMSTRONG, Gary. 1999 - *The Maximum Surveillance Society: The Rise of CCTV*. Oxford: Berg, 1999.

OTTERLO, Martijn Van - *A machine learning view on profiling*. In: HILDEBRANDT, M.; DE VRIES, K. (Eds.) - *Privacy, due process and the computational turn: philosophers of law meet philosophers of technology*. Abingdon: Routledge, 2013 p. 41-64.

PARLAMENTO EUROPEU (2017) - *Resolução do Parlamento Europeu, de 14 de março de 2017, sobre as implicações dos grandes volumes de dados nos direitos fundamentais: privacidade, proteção de dados, não discriminação, segurança e aplicação da lei* [2016/2225(INI)]. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-8-2017-0076_PT.html. Acesso em: 12 de maio de 2022.

PARLAMENTO EUROPEU (2021) - *Resolução do Parlamento Europeu, de 20 de janeiro de 2021, sobre a inteligência artificial: questões de interpretação e de aplicação do direito internacional na medida em que a UE é afetada nos domínios da utilização civil e militar e da autoridade do Estado fora do âmbito da justiça penal* [2020/2013(INI)]. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2021-0009_PT.html. Acesso em: 03 de maio de 2022.

PHILLIPS, P. Jonathon [et al.] - *Face recognition vendor test 2002*. IEEE International SOI Conference. Proceedings (Cat. No.03CH37443). p. 44 (2003). doi: 10.1109/AMFG.2003.1240822.

PHILLIPS, P. Jonathon [et al.] - *An Other-Race Effect for Face Recognition Algorithms*, *ACM Transactions on Applied Perception*. Online (2009). Disponível em: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=906254. Acesso em: 28 de abril de 2022.

PIRES, Inês - *How they see us: a non-discrimination-by-design regulatory solution to the gender and racial bias in law enforcement facial recognition technologies*. Tilburg: [S.n.], 2021. Dissertação de mestrado.

POLLI, Frida - *The Dark Side Of Artificial Intelligence*. Forbes, [s.l.], 2017. Disponível em: <https://www.forbes.com/sites/fridapolli/2017/12/05/the-dark-side-of-artificial-intelligence/#285d55212614>. Acesso em: 28 de abril de 2022.

RAGAZZI, Francesco [et al.] - *Biometric & Behavioural Mass Surveillance*. EU Member States. Report for the Greens/EFA in the European Parliament (2021). Disponível em: <https://www.greens-efa.eu/en/article/document/biometric-and-behavioural-mass-surveillance-in-eu-member-states>. Acesso em: 03 de maio de 2022.

RAJI, Inioluwa Deborah; BUOLAMWINI, Joy - *Actionable auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products*. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, p. 429-435 (2019). Disponível em: https://dam-prod.media.mit.edu/x/2019/01/24/AIES-19_paper_223.pdf. Acesso em: 30 de abril de 2022.

RAMOS, Silvia - *Retratos da Violência – Cinco meses de monitoramento, análises e descobertas*. Rio de Janeiro: Rede de Observatórios da Segurança/CESeC, 2019. Disponível em: <https://cesecseguranca.com.br/textodownload/retratos-da-violencia-cinco-meses-de-monitoramento-analises-e-descobertas/>. Acesso em: 01 de maio de 2022.

ROSE, Adam - *Are Face-Detection Cameras Racist?* TIME, 2010. Disponível em: <http://content.time.com/time/business/article/0,8599,1954643,00.html>. Acesso em: 01 de maio de 2022.

SCHICK, Nina - *Deepfakes: The Coming Infocalypse*. Nova York: Twelve, 2020.

SHEPLEY, Andrew Jason. - *Deep learning for face recognition: a critical analysis*. Casuarina: [S.n.], 2019.

SIMONITE, Tom - *Best Algorithms Struggle to Recognize Black Faces Equally*. WIRED, 2019. Disponível em: <https://www.wired.com/story/best-algorithms-struggle-recognize-black-faces-equally/>. Acesso em: 01 de maio de 2022.

SNOW, Jacob - *Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots*. ACLU, 2018. Disponível em: <https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28>. Acesso em: 30 de abril de 2022.

SURDEN, Harry - *Artificial intelligence and law: An overview*. Georgia State University Law Review. v. 35, p. 1305-1337 (2019).

TELEFI Project - *Summary Report of the project "Towards the European Level Exchange of Facial Images"* (2021). Disponível em: https://www.telefi-project.eu/sites/default/files/TELEFI_SummaryReport.pdf.

TURING, Alan - *Computing Machinery and Intelligence*. Mind, v. LIX, n. 236, p. 433-460 (1950).

UNIÃO EUROPEIA (2016) - *Carta Dos Direitos Fundamentais Da União Europeia* (2016/C 202/02).

UNIÃO EUROPEIA (2016) - *Diretiva (UE) 2016/680 do Parlamento Europeu e do Conselho de 27 de abril de 2016 relativa à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais pelas autoridades competentes para efeitos de prevenção, investigação, deteção ou repressão de infrações penais ou execução de sanções penais, e à livre circulação desses dados, e que revoga a Decisão-Quadro 2008/977/JAI do Conselho*. JO L 119/89 (Law Enforcement Directive).

UNIÃO EUROPEIA (2016) - *Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016 relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE* (Regulamento Geral sobre a Proteção de Dados).

UNIÃO EUROPEIA (2012) - *Versão consolidada do Tratado da União Europeia*. JO C326/13 (TEU).

UNIÃO EUROPEIA (2012) - *Versão consolidada do Tratado sobre o Funcionamento da União Europeia*. JO C326/47 (TFEU).

UNITED KINGDOM (2020) - *Court of Appeal (Civil Division) - R (Bridges) v CC South Wales & ors*. EWCA Civ, 1058.

UNITED NATIONS (1979) - *Convention on the Elimination of Discrimination Against Women* (adopted 18 December 1979, entered into force 3 September 1981). 1249 UNTS 13 (CEDAW).

UNITED NATIONS (1965) - *International Convention on the Elimination of All Forms of Racial Discrimination* (adopted 21 December 1965, entered into force 4 January 1969). 660 UNTS 195 (ICERD).

UNITED NATIONS (1966) - *International Covenant on Civil and Political Rights* (adopted 16 December 1966, entered into force 23 March 1976). 999 UNTS 171 (ICCPR).

UNITED NATIONS (1948) - *Universal Declaration of Human Rights (UDHR)*. [adopted and proclaimed by the United Nations General Assembly (resolution 217 A III) on 10 December 1948].

XENIDIS, Raphaële; SENDEN, Linda - *EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination* in Ulf Bernitz et al (eds) - *General Principles of EU law and the EU Digital Order*. Kluwer Law International (2020).

WACHTER, Sandra; MITTELSTADT, Brent; RUSSELL, Chris - *Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI*. Oxford Internet Institute, 2020.

WARD, Angela - *The Impact of the EU Charter of Fundamental Rights on Anti-Discrimination Law: More a Whimper than a Bang*. Cambridge Yearbook of European Legal Studies. v. 20, p. 32-60 (2018).

WECHSLER, Harry - *Reliable face recognition methods: system design, implementation and evaluation*. Springer, 2007.

WOODWARD, John D [et al.] - *Army Biometric Application: Identifying and Addressing Sociocultural Concerns*. RAND Corporation, 2001.

ZIZI, Martin - *The Flaws and Dangers of Facial Recognition*. Security Today, 2019. Disponível em: <https://securitytoday.com/articles/2019/03/01/the-flaws-and-dangers-of-facial-recognition.aspx>.

ZUIDERVEEN BORGESIU, Frederik - *Discrimination, artificial intelligence, and algorithmic decision-making*. Directorate General of Democracy, 2018.