ARTIGO ORIGINAL

O PAPEL DA LEGISLAÇÃO DE PROTEÇÃO DE DADOS PESSOAIS NA MITIGAÇÃO DE VIESES RACIAIS EM DECISÕES TOMADAS POR INTELIGÊNCIA ARTIFICIAL

THE ROLE OF DATA PROTECTION LEGISLATION IN MITIGATING RACIAL BIAS IN DECISIONS MADE BY ARTIFICIAL INTELLIGENCE

Gustavo Salema Marques[1]

RESUMO: Este artigo investiga o papel da legislação sobre proteção de dados pessoais na redução do viés racial em decisões automatizadas tomadas por sistemas de Inteligência Artificial (IA), como o *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS). O COMPAS é um sistema algorítmico utilizado pelo Sistema de Justiça Criminal dos EUA com o objetivo de avaliar a probabilidade de um infrator reincidir em um crime. O artigo apresenta as descobertas do ProPublica sobre as decisões racialmente tendenciosas tomadas pelo COMPAS e analisa as razões pelas quais os infratores negros são erroneamente classificados como indivíduos de alto risco pelo COMPAS, mesmo quando são menos violentos do que réus brancos que recebem avaliações de baixo risco. O estudo explora o potencial da legislação de proteção de dados pessoais, como a LGPD e o GDPR, na mitigação do viés racial em decisões automatizadas, tomadas por sistemas de inteligência artificial. Assim sendo, enfatizando a necessária conformidade com o princípio do *Data Protection by Design*, este artigo propõe o uso do legítimo interesse como base legal para a coleta de dados pessoais sensíveis relacionados à origem racial, com o intuito de permitir que os desenvolvedores de IA identifiquem e corrijam os vieses raciais presentes nos conjuntos de dados, para, dessa forma, promover decisões automatizadas racialmente imparciais, especialmente no âmbito de Sistemas de Justiça Criminal. Logo, buscou-se demonstrar a importância de desenvolver sistemas de inteligência artificial capazes de identificar vieses raciais em conjuntos de dados, evitando assim a natureza discriminatória dos conjuntos de dados aos quais esses sistemas estão expostos.

Palavras-chave: Proteção de dados; Legitimo interesse; Inteligência artificial.

[1] Mestrando em Direito da Regulação pela FGV Direito Rio. Advogado. Professor da disciplina de Direito e Inteligência Artificial na FGV Direito Rio. Visiting Student Researcher na Schulich School of Law (Dalhousie University). E-mail: gustavo.salema@hotmail.com. Orcid: https://orcid.org/0000-0002-3992-2550.

ABSTRACT: This article investigates the role of data protection legislation in reducing racial bias in automated decisions made by Artificial Intelligence (AI) systems, such as the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS). COMPAS is an algorithmic system used by the U.S. Criminal Justice System to assess the likelihood of an offender reoffending. The article presents findings from ProPublica on the racially biased decisions made by COMPAS and analyzes the reasons why Black offenders are incorrectly classified as high-risk by COMPAS, even when they are less violent than white defendants who receive low-risk assessments. The study explores the potential of data protection legislation, such as Brazil's LGPD and the EU's GDPR, in mitigating racial bias in automated decisions made by AI systems. Emphasizing the need for compliance with the principle of Data Protection by Design, this article proposes the use of legitimate interest as a legal basis for collecting sensitive personal data related to racial origin. The aim is to enable AI developers to identify and correct racial biases in datasets, thereby promoting racially impartial automated decisions, particularly within the scope of Criminal Justice Systems. Consequently, the article demonstrates the importance of developing AI systems capable of identifying racial biases in datasets, thus avoiding the discriminatory nature of the data to which these systems are exposed.

KEY-WORDS: Data protection; Legitimate interest; Artificial intelligence.

INTRODUÇÃO

Em 2016, o jornal investigativo ProPublica, nos EUA, publicou um artigo alegando que um *software* (programa de computador) estava sendo utilizado em todo o país com o objetivo de prever o potencial de reincidência de criminosos (SCHWARTING et al, 2023, p. 309). Assim sendo, o ProPublica, analisou as decisões automatizadas, que foram tomadas por meio de inteligência artificial (IA) e, com base em estatísticas que sustentavam a sua alegação, concluiu que esse aplicativo era tendencioso contra pessoas negras (SCHWARTING et al, 2023, p. 309).

O *software* analisado e criticado pelo ProPublica foi o *Correctional Offender Management Profiling for Alternative Sanctions* – COMPAS (Gestão Correicional de Perfis de Infratores para Sanções Alternativas), criado em 1998 pela Northpointe, com o objetivo de avaliar a probabilidade de um réu se tornar reincidente (SCHWARTING et al, 2023, p. 315). Logo, cabe ressaltar, que esta avaliação era realizada pelo referido sistema de IA, por meio da análise de 137 itens relacionados ao envolvimento criminal, estilo de vida, personalidade, atitudes e ambiente familiar e social dos detentos (CATALENA, 2020, p. 6).

Ocorre que o COMPAS é um sistema algorítmico que tem efeitos discriminatórios. Não porque utilize qualquer input étnico ou racial, mas porque ele aprendeu com um conjunto de dados repleto de decisões humanas discriminatórias (BORGESISUS, 2020, p. 5). Portanto, de acordo com o próprio desenvolvedor, tal sistema de IA funciona corretamente e não possui preconceito racial, sendo a desigualdade de tratamento racial natural, já que há mais negros encarcerados no sistema prisional dos EUA do que caucasianos (YEUNG et al, 2021, p. 4).

Todavia, para os fins deste ensaio, o viés racial ou étnico não pode ser considerado natural apenas porque o conjunto de dados está repleto de informações que podem influenciar as decisões tomadas pelo sistema de IA, tornando-as discriminatórias e, consequentemente, prejudicando um grupo racial em detrimento de outro.

Dito isso, o objetivo deste artigo é compreender como o COMPAS e outras inteligências artificiais, utilizadas como ferramenta decisória no ambiente do Sistema de Justiça Criminal, podem minimizar os vieses discriminatórios nas suas decisões e, assim, proporcionar um tratamento igualitário aos infratores, cujos dados são analisados pelo algoritmo da IA.

Portanto, considerando que o COMPAS pode utilizar dados dos réus, que podem ser indicadores de raça (GARRET, 2020, p. 3), para tomar decisões automatizadas com base nessas informações (CHELIOUDAKIS, 2020, p. 80 e 89), pode haver alguma incompatibilidade desse sistema de IA com a regulamentação relacionada à proteção de dados e ao direito à privacidade (BINNS et al, 2021, p. 319-320). Assim, este artigo pretende analisar de que forma as normas legais de proteção de dados pessoais podem mitigar o viés racial em decisões tomadas com o uso de sistemas de inteligência artificial.

Nesse sentido, considerando que o Brasil não está imune à adoção de um sistema de IA, no formato do COMPAS, pelo Poder Judiciário, este estudo conduzirá uma análise comparativa da legislação brasileira e europeia, para determinar se essas legislações podem ser ferramentas para reduzir os riscos que sistemas de IA, como o COMPAS, podem expor grupos socialmente marginalizados.

Logo, com base no exposto, este artigo tem como objetivo responder às seguintes perguntas: Por que o viés racial afeta as decisões automatizadas do COMPAS e quais são as consequências desse viés? Como a legislação de proteção de dados pessoais pode ser utilizada para reduzir o risco de decisões automatizadas racialmente viesadas?

Para responder a essas questões, o artigo será dividido em duas partes: na primeira, serão abordadas as questões relacionadas ao viés racial nas decisões tomadas pelo COMPAS. Já na segunda parte, será demonstrado como a legislação de proteção de dados pessoais podem minimizar o viés racial em decisões automatizadas.

1. O IMPACTO DO CONJUNTO DE DADOS NO PROCESSO DECISÓRIO DA IA

Como explicado no início deste artigo, o ProPublica analisou as decisões tomadas pelo COMPAS para verificar a existência de erros decisórios tendentes a prejudicar infratores negros por razões essencialmente raciais. Assim sendo, durante sua pesquisa, eles descobriram que as decisões do COMPAS estavam imbuídas de viés racial, conforme evidenciado nas informações que eles tornaram públicas e que serão apresentadas a seguir:

> - Os réus negros eram frequentemente previstos como apresentando um risco maior de reincidência do que realmente tinham. Nossa análise mostrou que réus negros que não reincidiram durante um período de dois anos tinham quase o dobro de chances de serem erroneamente classificados como de alto risco em comparação com seus equivalentes brancos (45% contra 23%).
> - Os réus brancos, por outro lado, eram frequentemente considerados menos arriscados do que realmente eram. Nossa análise revelou que réus brancos que reincidiram nos dois anos seguintes foram erroneamente rotulados como de baixo risco quase duas vezes mais frequentemente do que os réus negros reincidentes (48% contra 28%).
> - A análise também mostrou que, mesmo ao controlar por crimes anteriores, reincidência futura, idade e gênero, réus negros tinham 45% mais chances de serem classificados com pontuações de risco mais altas do que réus brancos.
> - Além disso, os réus negros tinham duas vezes mais chances do que os réus brancos de serem erroneamente classificados como sendo de maior risco de reincidência violenta. E os reincidentes violentos brancos tinham 63% mais chances de terem sido erroneamente classificados como de baixo risco de reincidência violenta, em comparação com os reincidentes violentos negros.
> - A análise sobre reincidência violenta também mostrou que, mesmo ao controlar por crimes anteriores, reincidência futura, idade e gênero, réus negros tinham 77% mais chances de serem classificados com pontuações de risco mais altas do que réus brancos. (LARSON et al., 2016)[2].

[2] Trecho retirado diretamente do site do ProPublica. Tradução realizada com o uso de ferramenta de inteligência artificial (ChatGPT).

Com base nos dados apresentados pelo ProPublica, é possível concluir que as decisões tomadas, a partir da avaliação de risco realizada pelo COMPAS, são racialmente viesadas e que possuem o potencial de prejudicar uma parcela da sociedade, somente em função da sua origem racial (LARSON et al., 2016). Dito isso, mesmo que o desenvolvedor alegue que este sistema de IA é racialmente neutro, é evidente que algo está errado quando a pesquisa empírica demonstra que os réus afro-americanos são rotulados como sendo de alto risco, mesmo quando são menos violentos do que um réu branco, a quem foi atribuída uma avaliação de baixo risco. (SULOCKI, 2020, p. 679).

Assim sendo, por um lado, ainda que seja possível aceitar que o cenário discriminatório percebido pelo ProPublica, não tenha sido causado pela existência de viés racial no algoritmo do COMPAS (GARRET, 2020, p. 3), por outro, a análise do ProPublica mostra que o conjunto de dados utilizados para treinar o sistema de IA está maculado por viés racial, significando que o COMPAS apenas aplica a seletividade que aprendeu com o conjunto de dados fornecidos pelo Sistema de Justiça Criminal. (SULOCKI, 2020, p. 679).

Logo, o fato do COMPAS ter sido treinado com um conjunto de dados racialmente viesado, resulta na perpetuação de um padrão decisório que tende a beneficiar infratores caucasianos (GARRET, 2020, p. 3), levando o sistema algorítmico a tomar decisões severas em relação a réus de grupos marginalizados, como os afro-americanos e outros grupos sociais discriminados pelo Sistema de Justiça Criminal (SULOCKI, 2020, p. 679).

Portanto, para propor uma possível solução para essa questão, este artigo sugere uma alteração na legislação sobre proteção de dados, com o intuito de viabilizar a utilização de dados sensíveis relacionados à raça, de forma a mitigar o preconceito existente no conjunto de dados. Em outras palavras, este artigo elucidará como o conhecimento prévio da raça do réu, pelo sistema de inteligência artificial, pode auxiliar o Judiciário na redução do viés racial em decisões automatizadas.

Embora isso possa parecer inusitado, este estudo está alinhado com a afirmação de que as organizações têm como objetivo prevenir o viés racial em decisões tomadas por sistemas de IA (BEKKUN et al., 2022, p. 6). No entanto, para fazer isso, uma organização precisa ter ciência de que o algoritmo está prejudicando determinada etnia, o que só pode ser feito se a instituição souber a raça da pessoa avaliada pelo sistema de IA (BEKKUN et al., 2022, p. 6).

Como essa proposta pode enfrentar barreiras legais na LGPD e no GDPR, é necessário primeiro entender se é possível para um desenvolvedor de IA adicionar esse fator

racial, com o objetivo de proteger o titular dos dados. Para tanto, a seguir, serão apresentadas as normas legais referentes a dados sensíveis na LGPD e no GDPR.

Artigos 5º e 11º da LGPD:

> Art. 5º Para os fins desta Lei, considera-se: (...) II - dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural; (...) Art. 11. O tratamento de dados pessoais sensíveis somente poderá ocorrer nas seguintes hipóteses: I - quando o titular ou seu responsável legal consentir, de forma específica e destacada, para finalidades específicas; II - sem fornecimento de consentimento do titular, nas hipóteses em que for indispensável para: a) cumprimento de obrigação legal ou regulatória pelo controlador; b) tratamento compartilhado de dados necessários à execução, pela administração pública, de políticas públicas previstas em leis ou regulamentos; c) realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais sensíveis; d) exercício regular de direitos, inclusive em contrato e em processo judicial, administrativo e arbitral, este último nos termos da Lei nº 9.307, de 23 de setembro de 1996 (Lei de Arbitragem); e) proteção da vida ou da incolumidade física do titular ou de terceiro; f) tutela da saúde, exclusivamente, em procedimento realizado por profissionais de saúde, serviços de saúde ou autoridade sanitária; ou (Redação dada pela Lei nº 13.853, de 2019) Vigência g) garantia da prevenção à fraude e à segurança do titular, nos processos de identificação e autenticação de cadastro em sistemas eletrônicos, resguardados os direitos mencionados no art. 9º desta Lei e exceto no caso de prevalecerem direitos e liberdades fundamentais do titular que exijam a proteção dos dados pessoais. (BRASIL, 2018, artigos 5º e 11º da LGPD)[3].

Artigo 9.º do GDPR:

> 1. O tratamento de dados pessoais que revelem origem racial ou étnica, opiniões políticas, crenças religiosas ou filosóficas, filiação sindical, bem como o tratamento de dados genéticos, dados biométricos para identificar de forma única uma pessoa natural, dados relativos à saúde ou dados relativos à vida sexual ou orientação sexual de uma pessoa natural, será proibido. 2. O parágrafo 1 não se aplicará se uma das seguintes condições for atendida: (a) o titular dos dados deu consentimento explícito para o tratamento desses dados pessoais para um ou mais propósitos específicos, exceto nos casos em que a lei da União ou dos Estados-Membros preveja que a proibição mencionada no parágrafo 1 não possa ser afastada pelo titular dos dados; (b) o tratamento for necessário para cumprir obrigações e exercer direitos específicos do controlador ou do titular dos dados no âmbito da legislação trabalhista, de segurança social e proteção social, desde que autorizado por lei da União ou dos Estados-Membros ou por um acordo coletivo conforme a legislação nacional, prevendo salvaguardas apropriadas para os direitos fundamentais e os interesses do titular dos dados; (c) o tratamento for necessário para proteger os interesses vitais do titular dos dados ou de outra pessoa natural, caso o titular dos dados esteja física ou legalmente incapaz de dar o seu consentimento; (d) o tratamento for realizado no âmbito das atividades legítimas de uma fundação, associação ou qualquer outra entidade sem fins lucrativos com

[3] Artigos extraídos do texto da Lei nº 13.709, de 14 de agosto de 2018 – Lei Geral de Proteção de Dados Pessoais **(**LGPD).

> objetivo político, filosófico, religioso ou sindical, com as devidas salvaguardas, e desde que o tratamento se refira exclusivamente aos membros ou ex-membros da entidade, ou a pessoas que mantenham contato regular com ela em conexão com seus objetivos, e que os dados pessoais não sejam divulgados fora da entidade sem o consentimento dos titulares dos dados; (e) o tratamento se refira a dados pessoais que foram manifestamente tornados públicos pelo titular dos dados; (f) o tratamento for necessário para o estabelecimento, exercício ou defesa de reivindicações legais ou sempre que os tribunais atuarem no exercício de sua função judicial; (g) o tratamento for necessário por razões de interesse público substancial, com base em legislação da União ou dos Estados-Membros, que deverá ser proporcional ao objetivo perseguido, respeitar a essência do direito à proteção de dados e prever medidas adequadas e específicas para proteger os direitos fundamentais e os interesses do titular dos dados; (h) o tratamento for necessário para fins de medicina preventiva ou ocupacional, para a avaliação da capacidade de trabalho do empregado, diagnóstico médico, prestação de cuidados ou tratamento de saúde ou social, ou para a gestão de sistemas e serviços de saúde ou cuidados sociais, com base em legislação da União ou dos Estados-Membros ou em contrato com um profissional de saúde, e sujeito às condições e salvaguardas mencionadas no parágrafo 3; (i) o tratamento for necessário por razões de interesse público na área da saúde pública, como a proteção contra ameaças graves transfronteiriças à saúde ou para garantir altos padrões de qualidade e segurança dos cuidados de saúde e de produtos ou dispositivos médicos, com base em legislação da União ou dos Estados-Membros, que preveja medidas adequadas e específicas para proteger os direitos e liberdades do titular dos dados, em particular o sigilo profissional; (j) o tratamento for necessário para fins de arquivamento no interesse público, para fins de pesquisa científica ou histórica ou para fins estatísticos, de acordo com o Artigo 89(1), com base na legislação da União ou dos Estados-Membros, que deverá ser proporcional ao objetivo perseguido, respeitar a essência do direito à proteção de dados e prever medidas adequadas e específicas para proteger os direitos fundamentais e os interesses do titular dos dados. 3. Os dados pessoais mencionados no parágrafo 1 podem ser tratados para os fins referidos no ponto (h) do parágrafo 2 quando esses dados forem tratados por ou sob a responsabilidade de um profissional sujeito à obrigação de sigilo profissional de acordo com a legislação da União ou dos Estados-Membros, ou regras estabelecidas por órgãos nacionais competentes, ou por outra pessoa também sujeita a uma obrigação de sigilo de acordo com a legislação da União ou dos Estados-Membros ou regras estabelecidas por órgãos nacionais competentes. 4. Os Estados-Membros podem manter ou introduzir condições adicionais, incluindo limitações, em relação ao tratamento de dados genéticos, dados biométricos ou dados relativos à saúde. (UNIÃO EUROPEIA, 2016, artigo 9 da GDPR)[4]

Como pode ser visto, ambas as legislações são restritivas quanto ao tratamento de dados sensíveis e possuem apenas algumas exceções para isso. Nesse sentido, ressalta-se que, nem a LGPD (BRASIL, 2018, artigos 5º e 11º da LGPD) e nem a GDPR (BEKKUN et al., 2022, p. 6) trazem uma exceção que permita o tratamento de dados sobre origem racial, para fins de mitigação de viés racial em decisões automatizadas.

Ocorre que, se uma empresa, como a Northpointe, quiser testar se o seu sistema de IA está sendo racialmente tendencioso, a organização não poderá verificá-lo e não saberá

[4] Artigo extraído do texto da Regulação (EU) 2016/679 do Parlamento e do Conselho Europeu (GDPR). Tradução realizada com o uso de ferramenta de inteligência artificial (ChatGPT).

que criou, acidentalmente, um sistema de inteligência artificial capaz de discriminar pessoas devido à sua raça, (BEKKUN et al., 2022, p. 3), especialmente quando esse preconceito provém do conjunto de dados utilizado durante o treinamento da IA e não do próprio algoritmo (YEUNG et al, 2021, p. 3).

Portanto, como os sistemas de IA são treinados com base em dados históricos, esses sistemas automatizados de tomada de decisão reproduzirão todo o viés racial presente nos dados sobre os quais foram treinados (YEUNG et al, 2021, p. 3). Assim sendo, ao aplicar esse conhecimento ao caso COMPAS, é possível concordar com a Northpointe quando essa organização afirma que, o viés racial apontado pela ProPublica não foi produzido pelo algoritmo, mas foi uma consequência de um conjunto de dados racialmente viesado (YEUNG et al, 2021, p. 4).

Por essa razão, este ensaio propõe que o legítimo interesse se torne uma base legal e uma exceção para o tratamento de dados de origem racial, com o único objetivo de prevenir a discriminação. Sendo assim, espera-se que, com o uso de dados pessoais relacionados à raça, as organizações consigam verificar se seus sistemas de IA são racialmente tendenciosos (BEKKUN et al., 2022, p. 9).

2. LEGÍTIMO INTERESSE E O DESENVOLVIMENTO DO SISTEMA DE IA

Antes de analisar o uso do legítimo interesse para a redução do viés racial em sistemas de IA, como o COMPAS, este artigo discorrerá sobre a importância da Proteção de Dados na regulamentação da inteligência artificial. Para tanto, será feita uma breve análise das legislações brasileira (LGPD) e europeia (GDPR) sobre Proteção de Dados, destacando como elas podem promover a adoção do *Data Protection by Design* (proteção de dados desde a concepção) no desenvolvimento de sistemas de inteligência artificial.

O *Data Protection by Design* é um princípio presente tanto na LGPD (BRASIL, 2018, artigo 46, §2 da LGPD) quanto no GDPR (UNIÃO EUROPEIA, 2016, artigo 25 da GDPR). Entretanto, para os fins desse artigo, será adotado o conceito fornecido pelo *UK Information Commissioner's Office – ICO* (Escritório do Comissariado de Informação do

Reino Unido) sobre *Data Protection by Design* (REINO UNIDO - ICO)[5], já que, na visão desse estudo, abrange os conceitos trazidos pela LGPD e pelo GDPR de forma mais didática.

Assim, conforme o ICO, o princípio do *Data Protection by Design* significa que a proteção de dados deve ser considerada desde a fase de concepção de um sistema e ao longo de todo o seu ciclo de vida (REINO UNIDO - ICO). Portanto, a aplicação do *Data Protection by Design* a sistemas de decisão automatizada, implica que o sistema de IA será desenvolvido e implementado considerando questões de proteção de dados e conformidade com as legislações pertinentes. (REINO UNIDO - ICO).

Assim sendo, no contexto do processamento de dados pessoais, por sistemas de IA, que podem produzir decisões automatizadas com viés racial, é importante notar que tanto a LGPD quanto o GDPR já possuem regras que promovem a não discriminação (BELLI et al., 2022, p. 22-23), proibindo, portanto, o tratamento de dados pessoais para fins discriminatórios (BRASIL, 2018, artigos 6º, IX da LGPD).

Todavia, ressalta-se que, mesmo já existindo normas que tem o potencial de mitigar os efeitos de decisões racialmente discriminatórias, não se vislumbra como elas podem evitar que o sistema de inteligência artificial assimile o viés racial existente no conjunto de dados utilizados no seu treinamento. Portanto, com o intuito de melhor explicar a proposta desse artigo, o próximo passo será explicar a importância dos dados no desenvolvimento de sistemas de IA similares ao COMPAS.

Assim sendo, as inteligências artificiais são sistemas algorítmicos que podem aprender com suas próprias experiências e resolver problemas complexos, sendo sustentados e viabilizados por meio de dados (AUTORIDADE NORUEGUESA DE PROTEÇÃO DE DADOS, 2018, p. 6-7). Em outras palavras, a IA precisa de um enorme volume de dados, que podem ou não ser dados pessoais, para alcançar seu objetivo de criação (AUTORIDADE NORUEGUESA DE PROTEÇÃO DE DADOS, 2018, p. 7 e 11). No entanto, para que uma inteligência artificial seja eficaz, o desenvolvedor do sistema de IA prezará mais pela qualidade desses dados do que pela quantidade deles (AUTORIDADE NORUEGUESA DE PROTEÇÃO DE DADOS, 2018, p. 11). Assim sendo, para que o conjunto de dados utilizado durante o treino da IA seja representativo da tarefa pretendida, é essencial que o sistema de IA seja treinado com dados abundantes e de alta qualidade (AUTORIDADE NORUEGUESA DE PROTEÇÃO DE DADOS, 2018, p. 11).

[5] Informação extraída diretamente do site do UK Information Commissioner's Office (ICO.)

Ocorre que, com o desenvolvimento dos sistemas de IA, passou-se a acreditar que as inteligências artificiais poderiam ter um desempenho superior ao dos seres humanos, uma vez que não seriam afetadas por questões relacionadas à natureza humana. Todavia, essa pode ser uma percepção equivocada, já que a inteligência artificial é tão objetiva quanto a pessoa que a desenvolveu e quanto os dados com os quais foi treinada (AUTORIDADE NORUEGUESA DE PROTEÇÃO DE DADOS, 2018, p. 16).

Tendo isso em mente, pode-se concluir que os sistemas de inteligência artificial e suas decisões automatizadas serão tão discriminatórias quanto os dados utilizados durante o seu treinamento (AUTORIDADE NORUEGUESA DE PROTEÇÃO DE DADOS, 2018, p. 16). Logo, como a base de dados do Sistema Judicial Criminal está repleta de preconceitos raciais contra afro-americanos (SULOCKI, 2020, p. 679), os sistemas de algoritmos que aprenderem com esse conjunto de dados também será viesado contra indivíduos negros e poderá tomar decisões com base nesse viés racial (AUTORIDADE NORUEGUESA DE PROTEÇÃO DE DADOS, 2018, p. 16).

No entanto, a simples adequação às regras de não discriminação impostas pelas leis de proteção de dados, não é suficiente para que os sistemas de IA estejam em conformidade com essas legislações (AUTORIDADE NORUEGUESA DE PROTEÇÃO DE DADOS, 2018, p. 16). Portanto, o modelo de inteligência artificial deve também ser treinado com dados corretos e de qualidade, com ênfase em dados que não levem a tratamentos ou decisões discriminatórias (AUTORIDADE NORUEGUESA DE PROTEÇÃO DE DADOS, 2018, p. 16).

O treinamento com dados pessoais de qualidade é essencial para o desenvolvimento de sistemas de IA imparciais e, consequentemente, a utilização de conjuntos de dados de baixa qualidade pode levar a IA a tomar decisões racialmente viesadas (KRAMCSÁK, 2022, p. 5). Entretanto, coletar dados pessoais sensíveis, como a origem racial, com o objetivo de criar um conjunto de dados de qualidade, não é uma tarefa simples, uma vez que o desenvolvedor precisará do consentimento do titular dos dados, para utilizar seus dados pessoais sensíveis como parte do conjunto de dados de treinamento (KRAMCSÁK, 2022, p. 7, 9 e 10).

Por essa razão, este ensaio propõe a criação de uma exceção para a coleta de dados pessoais sensíveis relacionados à raça. Logo, propõe-se que os legisladores considerem o legítimo interesse como uma isenção para viabilizar o tratamento de informações sobre a raça. Com isso, os sistemas de IA seriam capazes de identificar preconceitos raciais no

conjunto de dados e evitariam influências de dados discriminatórios (BEKKUN et al., 2022, p. 9).

Nesse sentido, cabe observar que o uso do legítimo interesse para a coleta de dados que não sejam sensíveis, não encontra barreiras na legislação que trata sobre proteção de dados pessoais (KRAMCSÁK, 2022, p. 11, 13, 15 e 17). Todavia, o mesmo não se aplica ao tratamento de dados pessoais sensíveis, pois tanto a LGPD (BRASIL, 2018, artigos 5º e 11º da LGPD) quanto o GDPR (UNIÃO EUROPEIA, 2016, artigo 9 da GDPR) são muito restritivas quanto ao uso desses dados, como é o caso da origem racial.

Entretanto, como a aplicação do legítimo interesse exige que o interesse do responsável pelo tratamento dados se sobreponha a um direito fundamental do titular dos dados pessoais (KRAMCSÁK, 2022, p. 13), pode-se discutir, para efeitos deste artigo, se a mitigação do viés racial em decisões automatizadas, não seria suficiente para permitir o uso de dados relacionados à raça no treinamento de sistemas de IA, com o intuito de capacitá-los a identificar preconceitos raciais inseridos no conjunto de dados utilizados em seu treinamento.

Dito isto, é possível argumentar que, caso a prevalência do legítimo interesse do controlador sobre os direitos e liberdades do titular dos dados, beneficie a sociedade como um todo, inclusive o titular dos dados sensíveis, é possível que os criadores de IA possam utilizar o legítimo interesse como base legal para justificar o tratamento de dados pessoais sensíveis, sem que precisem obter o consentimento de cada titular de dados. (KRAMCSÁK, 2022, p. 13).

Ressalta-se que este ensaio demonstrou que os conjuntos de dados podem conter vieses raciais, que podem levar os sistemas de inteligência artificial a prejudicarem grupos étnicos marginalizados. Em razão disso, buscou-se destacar a importância da qualidade e da quantidade de dados no desenvolvimento de um sistema de decisão automatizado que cumpra com o princípio do *Data Protection by Design*.

Dessa forma, este texto propôs uma solução que visa não apenas proteger a sociedade e o interesse do controlador, mas, especialmente, proteger os direitos fundamentais dos indivíduos de grupos étnicos marginalizados. Portanto, conclui-se que o uso do legítimo interesse é uma opção válida para viabilizar o processamento de dados relacionados à raça, uma vez que visa garantir que sistemas de IA, como o COMPAS, não tomem decisões automatizadas tendenciosas contra grupos socialmente vulneráveis, o que, segundo este

artigo, configura uma exceção legal que protegerá os titulares dos dados, contra conjuntos de dados discriminatórios.

CONCLUSÃO

Este artigo examinou o caso do COMPAS e concluiu que o viés racial no sistema de IA decorre da discriminação racial generalizada nas bases de dados do Sistema de Justiça, levando esse sistema a tomar decisões automatizadas que são discriminatórias contra afro-americanos e outros grupos étnicos marginalizados.

Este estudo também analisou como a legislação de proteção de dados pessoais pode ser utilizado para mitigar os riscos associados ao preconceito racial. Assim sendo, após compreender como as leis de proteção de dados podem regulamentar sistemas de IA, constatou-se que a qualidade do conjunto de dados influencia diretamente as decisões da IA tornando-as mais ou menos tendenciosas em termos raciais.

Por fim, com base nas discussões apresentadas, este ensaio propõe a utilização do legítimo interesse como base legal para a coleta de dados raciais, com o objetivo de capacitar os desenvolvedores de inteligência artificial a criar sistemas que reconheçam e desconsiderem o viés racial presente nos conjuntos de dados em que são treinados.

REFERÊNCIAS

AUTORIDADE NORUEGUESA DE PROTEÇÃO DE DADOS (January 2018) **Artificial intelligence and privacy report**. Disponível em: < https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf. > Acesso em: 05 dec 2023.

BEKKUN, Marvin van e BORGESIUS, Frederik Zuiderveen. **Using sensitive data to prevent discrimination by artificial intelligence: Does the GDPR needs a new exception?** Disponível em: < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4104823>. Acesso em: 17 dez 2023.

BELLI, Luca, GASPAR, Walter Britto e CURZI, Yasmin. **AI Regulation in Brazil: advancements, flows and need to learn from the data protection experience.** In Computer Law and Security Review: Special Issue on Artificial Intelligence and Data

Protection in Latin America. (2022). Disponível em: < https://www.sciencedirect.com/journal/computer-law-and-security-review/special-issue/10SD06FBTBZ>. Acesso em: 04 dez 2023.

BINNS, Reuben e VEALE, Michael. **Is that your final decision? multi-stage profiling, selective effects, and article 22 of the GDPR**. International Data privacy Law, 2021, Vol 11, nº 4. p. 319-332. Disponível em: < https://academic.oup.com/idpl/article/11/4/319/6403925>. Acesso em: 14 dez. 2023.

BORGESISUS, Frederik J Zuiderveen. **Strengthening legal protection against discrimination by algorithms and artificial**. Disponível em: < https://papers.ssrn.com/sol3/papers.cfm?abstract id=3561441>. Acesso em: 14 dez. 2023

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018 (LGPD)**. Disponível em: < https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm>. Acesso em: 16 dez. 2023.

CATALENA, Maria Stefania. **Humane artificial intelligence: the fragility of Human Rights Facing AI**. East-West Center (2020). Disponível em: < https://www.jstor.org/stable/resrep25514>. Acesso em: 12 dez 2023.

CHELIOUDAKIS, Eleftherios. **Risk assessment tools in Criminal Justice: Is there a need for such tools in Europe and would their use comply with European Data Protection?** Disponível em: < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3743757>. Acesso em: 18 dec. 2023.

GARRET, Brandon L. **Justice in forensic algorithms**. Havard Data Science Review-2020. Disponível em: < https://assets.pubpub.org/cc7ikcib/684fe49b-82b0-448a-8f33-92dfc42920e1.pdf>. Acesso em: 16 dec 2023.

KRAMCSÁK, Pablo Trigo. **Can legitimate interest be an appropriate lawful basis for processing Artificial Intelligence training datasets**? In Computer Law and Security Review: Special Issue on Artificial Intelligence and Data Protection in Latin America. (2022). Disponível em: < https://www.sciencedirect.com/journal/computer-law-and-security-review/special-issue/10SD06FBTBZ >. Acesso em: 05 dec 2023.

LARSON, Jeff; MATTU Surya, KIRCHNER, Lauren and ANGWIN, Julia. **How we analyze the COMPAS recidivism algorithm**. ProPublica – May 23, 23016. Disponível em: < https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm >. Acesso em: 15 dec. 2024.

REINO UNIDO - ICO. **Data protection by design and default**. Disponível em: < https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/guide-to-accountability-and-governance/accountability-and-governance/data-protection-by-de sign-and- default/#:~:text=Exte rnal%20link-,What%20is%20data%20protection%20by%20design%3F,and%20then%20throughout%20the%20lifecycle. Acesso em: 20 dec.2023.

SCHWARTING, Rena and ULBRICHT, Lena. **Why organizations matter in "algorithmic discrimination"**. Köln Z Soziol 74, (Suppl 1), 307-330 (2022). Disponível em: https://link.springer.com/article/10.1007/s11577-022-00838-3. Acesso em: 14 dec 2023.

SULOCKI, Victoria de. **Novas tecnologias, velhas discriminações: ou da falta de reflexão sobre o sistema de algoritmos na Justiça Criminal**. In: MELLO, Ana de Oliveira Frazão Vieira de e MULHOLLAND, Caitlin Sampaio (coord.). Inteligência Artificial e Direito: Ética, Regulação e Responsabilidade. 2 ed. São Paulo: Revista dos Tribunais, 2020.

UNIÃO EUROPEIA. **Regulation (EU) 2016/679 of the European Parliament and of the Council (GDPR)**. Disponível em < https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>. Acesso em: 16 dec. 2023.

YEUNG, Douglas; KHAN, Inez; KALRA, Nidhi e OSOBA, Osonde A. **Identifying systemic bias in the acquisition of machine learning decision aids for law enforcement applications**. Disponível em:< https://www.jstor.org/stable/resrep29576. Acesso em: 18 dec 2023.