

**INSTITUTO BRASILEIRO DE ENSINO, DESENVOLVIMENTO E PESQUISA
PROGRAMA DE PÓS-GRADUAÇÃO STRICTO SENSU EM DIREITO
MESTRADO ACADÊMICO EM DIREITO CONSTITUCIONAL**

PEDRO SOUSA

**DIREITO PENAL NOS TEMPOS DA INTELIGÊNCIA ARTIFICIAL:
UMA ANÁLISE DA RESPONSABILIDADE DOS AGENTES ENVOLVIDOS NO
DESENVOLVIMENTO E NA OPERAÇÃO DE ALGORITMOS DE SELEÇÃO E
RECRUTAMENTO EM RELAÇÃO AO CRIME DE RACISMO PREVISTO NO ART. 4º DA
LEI 7.716/1989**

BRASÍLIA

2022

PEDRO SOUSA

**DIREITO PENAL NOS TEMPOS DA INTELIGÊNCIA ARTIFICIAL:
UMA ANÁLISE DA RESPONSABILIDADE DOS AGENTES ENVOLVIDOS NO
DESENVOLVIMENTO E NA OPERAÇÃO DE ALGORITMOS DE SELEÇÃO E
RECRUTAMENTO EM RELAÇÃO AO CRIME DE RACISMO PREVISTO NO ART. 4º DA
LEI 7.716/1989**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação Stricto Sensu em Direito Constitucional do IDP como requisito parcial para obtenção do título de Mestre em Direito Constitucional.

BRASÍLIA

2022

PEDRO SOUSA

DIREITO PENAL NOS TEMPOS DA INTELIGÊNCIA ARTIFICIAL

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação Stricto Sensu em Direito Constitucional do IDP como requisito parcial para obtenção do título de Mestre em Direito Constitucional.

19 de dezembro de 2022

BANCA EXAMINADORA

Prof.^a Dr.^a Laura Schertel Ferreira Mendes

Orientadora

Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa

Prof.^a Dr.^a Carolina Costa Ferreira

Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa

Membro Interno

Prof. Dr. Evandro Charles Piza Duarte

Universidade de Brasília

Membro Externo

BRASÍLIA
2022

Aos meus pais, pelo suporte e amor em meio a este mundo turbulento,
À Larah, pelo amor e pelo carinho.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus, que me deu forças para enfrentar esse desafio. Agradeço, também a meus amigos de trincheira Thales Cassiano, Ana Ribeiro, Anna Balzacchi, Alexandre Satyro, Thiago Bouza, pelos debates, pela troca de ideias e por me ajudarem a pensar e tentar entender o mundo. À professora Fernanda Lage, cujas aulas me levaram aos questionamentos que fizeram um penalista se interessar por temas de Inteligência Artificial e à professora Marilda Silveira, cujas lições de metodologia foram fundamentais para tentar desenvolver a pesquisa de forma mais adequada. Agradeço, também, aos professores Evandro Piza e Carolina Costa Ferreira, cujos apontamentos – em banca de qualificação – foram fundamentais para delimitar com mais precisão os rumos para finalização deste trabalho.

*“Desliga o telefone, rádio da cozinha
Iphone, celular, computador, motor
Ipad, som, luz e televisão
E liga cabeça, sintoniza o presente
Pra que aconteça simplesmente você.”*

João Suplicy

Resumo

O presente trabalho buscou analisar a responsabilidade penal decorrente de algoritmos de seleção e recrutamento, especificamente em relação ao crime de racismo, previsto no art. 4º da Lei n.º 7.716/1989. Para isso, foi feita análise da literatura existente sobre IA de seleção e recrutamento e sobre responsabilidade penal decorrente de IA. A hipótese apresentada foi de que, dadas as limitações existentes no direito penal brasileiro (como a regra de responsabilidade subjetiva), seria difícil a caracterização do crime de racismo no caso algoritmos de seleção e recrutamento. Após análise de casos, a hipótese foi parcialmente refutada, concluindo-se pela possibilidade de caracterização do crime de racismo por dolo eventual quando os agentes envolvidos na operação e no desenvolvimento de algoritmos possuem conhecimento de seu caráter discriminatório sob uma perspectiva racial. Sobre as dificuldades de responsabilização em face do racismo algorítmico, defende-se que a resposta estatal deve ser regulatória, e não penal, criando formas de evitar a discriminação – o que também é importante pela dificuldade de individualizar não somente os autores, mas também as vítimas do racismo algorítmico nas atividades de seleção e recrutamento.

Palavras-chave: Direito penal; Constitucionalismo Digital; Inteligência Artificial; seleção e recrutamento; Direito regulatório.

Abstract

This work sought to analyze criminal liability resulting from selection and recruitment algorithms, specifically in relation to the crime of racism, enacted in art. 4th of the Law 7.716/1989. For this, was made an analysis of the existing literature on selection and recruitment AI and on criminal liability resulting from AI. The hypothesis presented was that, given the existing limitations in Brazilian criminal law (such as the rule of subjective liability), it would be difficult to characterize the crime of racism in the case of selection and recruitment algorithms. After the analysis of the cases, the hypothesis was partially refuted, concluding that it is possible to characterize the crime of racism by eventual intent when the agents involved in the operation and development of algorithms are aware of its discriminatory nature from a racial perspective. Regarding the difficulties of accountability in the face of algorithmic racism, it is argued that the state response should be regulatory, and not criminal, creating ways to avoid discrimination - which is also important due to the difficulty of individualizing not only the perpetrators, but also the victims of algorithmic racism in selection and recruitment activities.

Keywords: Criminal law; Digital constitutionalism; Artificial Intelligence; selection and recruitment; Regulatory law.

Sumário

1. INTRODUÇÃO	10
2. ALGORITMOS DE SELEÇÃO E RECRUTAMENTO, DISCRIMINAÇÃO E MODELOS DE RESPONSABILIDADE PENAL	14
2.1. O MUNDO DIGITAL E A REGULAÇÃO JURÍDICA	17
2.1.1. Algoritmos de seleção e recrutamento: estado da arte	25
2.1.2. Estado da arte no Brasil.....	26
2.1.3. Estado da arte em língua inglesa	29
2.2. RESPONSABILIDADE PENAL POR INTELIGÊNCIA ARTIFICIAL: estado da arte	36
2.2.1. Estado da arte no Brasil.....	36
2.2.2. Estado da arte no mundo	41
2.3. MODELOS DE RESPONSABILIDADE PENAL NO DIREITO BRASILEIRO E IA	46
2.3.1. Responsabilidade penal da pessoa jurídica no direito penal brasileiro	47
2.3.2. O direito penal brasileiro, as teorias da ação e o dolo no tipo penal	49
2.3.3. Autoria no direito penal brasileiro.....	54
2.3.4. Culpabilidade no direito penal brasileiro	56
2.3.5. Legalidade no direito penal brasileiro	57
2.4. CONSIDERAÇÕES FINAIS	57
3. RACISMO ALGORÍTMICO: DEFINIÇÃO DO CONCEITO E ESTUDO DE CASOS	59
3.1. O RACISMO ENQUANTO FENÔMENO SOCIAL E JURÍDICO NO BRASIL	59
3.1.1. O cenário antes da Lei 7.716/1989 e da Constituição Federal de 1988	62
3.1.2. A lei de racismo, seu contexto e seus tipos penais	63
3.1.3. O Racismo algorítmico e a discriminação algorítmica: as dificuldades e dilemas de um mundo tecnológico	68
3.2. ESTUDO DE CASOS	74
3.2.1. Caso 1: formas tradicionais de contratação e de racismo.....	75
3.2.2. Caso 2: a filtragem de currículos por meio de IA	75
3.2.3. Caso 3: os anúncios de emprego desiguais.....	81
3.2.4. Caso 4: os algoritmos de recrutamento passivo	82
3.3. CONSIDERAÇÕES FINAIS	83
4. PONTE PARA O FUTURO: QUESTÕES DE POLÍTICA CRIMINAL E REGULATÓRIA PARA OS ALGORÍTMOS DE SELEÇÃO E RECRUTAMENTO	85
4.1. MODELOS PENAIS NA SOCIEDADE CONTEMPORÂNEA E A IA DE SELEÇÃO E RECRUTAMENTO.....	86

4.2. FUGINDO DA PRISÃO? AS POSSIBILIDADES REGULATÓRIAS E O COMBATE AO RACISMO ALGORÍTMICO	94
4.3. CONSIDERAÇÕES FINAIS	100
5. CONCLUSÕES	101

1. INTRODUÇÃO

A inteligência artificial (IA) possui cada vez mais importância na vida das pessoas, funcionando como meio para auxiliar decisões humanas ou para tomar decisões. Ferramentas de inteligência artificial são usadas para pilotar aviões e carros, para determinar se – e em que condições – uma pessoa conseguirá crédito, quem vai conseguir um emprego ou ser promovido. As possibilidades são cada vez maiores e influenciam diretamente a vida das pessoas. No entanto, se as potencialidades do uso de IA ampliam as capacidades sociais, também representam diversos riscos para a vida das pessoas, o que gera desafios para o direito e, mais especificamente, para o direito penal.

Por um lado, condutas tipificadas como crimes podem ser praticadas por meio de IA: um carro autônomo pode atropelar pessoas (fato tipificado como crime contra a vida ou incolumidade física), um algoritmo de atendimento ao público pode xingar e ofender (fato tipificado como crime contra a honra). Por outro lado, há a dificuldade de atribuição de conduta, ação ou omissão ao algoritmo, e de individualizar os autores, de modo que as atuais regras de responsabilidade penal parecem insuficientes para resolver questões relacionadas aos algoritmos, seja pela dificuldade de individualizar os agentes ou mesmo de determinar quais são as condutas praticadas por ou por meio de IA¹.

As dificuldades de pesquisa do tema são grandes, uma vez que há pouca literatura sobre responsabilidade penal e inteligência artificial, sendo que a maior parte dessa literatura trata dos carros autônomos e dos riscos à vida enquanto bem jurídico tutelado. Em relação a crimes decorrentes de atitudes discriminatórias e o uso da inteligência artificial, ainda deve haver desenvolvimento teórico acerca o tema, notadamente sob a ótica do direito penal. Dentro desse contexto de possibilidades discriminatórias, um dos usos mais perigosos é justamente o de algoritmos de seleção de emprego, que, no contexto da União Europeia, podem ser considerados como aplicações de alto risco, como proposto no projeto do *Artificial Intelligence Act*², apresentado pela Comissão Europeia, e que está tramitando no Parlamento Europeu. No contexto brasileiro, não há regramento legal da IA, mas tão somente projeto em tramitação sobre o tema (Projeto de Lei 21/2020 da Câmara dos Deputados³), já aprovado em sua casa iniciadora, e que se encontra em fase de produção de

1 Este debate será detalhado de forma mais adequada em momento posterior deste trabalho (2.2).

2 Tema abordado de forma mais precisa em momento posterior (2.1).

3 O projeto mencionado possui proposta inicial de relatoria da Dep. Luiza Canziani e, em sua redação atual, estabelece a responsabilidade civil subjetiva em decorrência de IA. No entanto, tem ocorrido – no contexto brasileiro – intenso debate em relação às possibilidades regulatórias e estruturas de responsabilidade em relação a algoritmos, não havendo definição em relação às regras que eventual legislação que regulará o tema irá estabelecer. Além deste projeto de lei, a tramitação

anteprojeto pela comissão de juristas instituída pelo Senado Federal. No entanto, não há qualquer indicativo, até o presente momento, de que o projeto tratará de responsabilidade penal em decorrência de IA, mas das possibilidades de responsabilidade civil, regras de governança e outras normas gerais relacionadas ao tema.

É nesse cenário de indefinição regulatória e de pouca produção acadêmica que se desenvolve o presente trabalho. O tema é a responsabilidade penal relacionada à discriminação ocorrida no âmbito de IA aplicada a seleções de emprego. Deste tema, surge a seguinte pergunta: os desenvolvedores e os operadores de algoritmos de seleção e recrutamento podem ser penalmente responsabilizados pelo crime de racismo tipificado no art. 4º da Lei 7.716/1989?

A hipótese que se estabelece é que somente em casos excepcionais – em que houver caracterização de dolo dos desenvolvedores e operadores de algoritmos no sentido de excluir candidatos negros – incidirá o direito penal.

Os objetivos da presente pesquisa são compreender: (i) a existência de discriminação de determinados grupos por meio de algoritmos de seleção e recrutamento; (ii) as possibilidades de cometimento do crime de racismo por operadores de algoritmos ou seus superiores hierárquicos – na modalidade comissiva ou omissiva imprópria – por meio de algoritmos de seleção e recrutamento.

O método empregado para a realização do presente trabalho será o hipotético-dedutivo, de modo que, ao longo do trabalho, tentar-se-á atribuir responsabilidade penal a desenvolvedores e operadores de algoritmos nas situações trabalhadas, e compreender em quais casos estes poderão ser penalmente responsabilizados segundo as regras de imputação vigentes no direito penal brasileiro. Além disso, será feita pesquisa bibliográfica e revisão de literatura em relação a temas importantes para o desenvolvimento do trabalho, como: (i) algoritmos de seleção e recrutamento e (ii) responsabilidade penal decorrente de Inteligência Artificial.

Para isso, o trabalho fará uso de conceitos do direito digital para compreender sob uma perspectiva jurídica o funcionamento da inteligência artificial. Além disso, a teoria do direito penal será aplicável para que seja possível compreender as possibilidades de criminalização de condutas praticadas por meio de algoritmos. Neste contexto, a teoria da imputação objetiva será usada para estabelecer tipicidade das condutas vinculadas à inteligência artificial. Além do interesse inerente ao caso específico dos algoritmos de seleção e recrutamento e as possíveis consequências destes para o direito penal, este trabalho busca a ampliação do debate em relação às intersecções entre o direito

no Senado Federal reuniu com este outros dois projetos originados no próprio Senado, quais sejam: 5.051/2019 e 872/2021.

penal e a IA e o direito digital, o que poderá possibilitar o desenvolvimento de pesquisas futuras acerca de temas relacionados a este último.

Para possibilitar a análise, serão estabelecidos os casos referência para o presente trabalho. Os casos que serão apresentados neste trabalho se baseiam no que foi descoberto no estado da arte em relação aos algoritmos de seleção e recrutamento (2.2) e busca descrever as principais práticas de seleção e recrutamento por meio de inteligência artificial, além de um caso clássico de racismo na seleção de emprego, para permitir que se compare o cenário anterior (em que pessoas tomavam as decisões diretamente) com o cenário atual (marcado pela crescente aplicação de IA e a consequente redução do papel humano na tomada de decisões). A seguir, enumera-se a descrição dos casos:

Caso 1 – Lucas, chefe de RH da Empresa Alfa, deixa de contratar José por este último ser negro. Como se verá adiante, esse é o caso clássico de discriminação direta para o qual a Lei 7.716/1989 foi pensada.

Caso 2 – Lucas, chefe de RH da Empresa Alfa, contrata a Empresa Beta para filtrar somente os 3% melhores currículos, que serão chamados para entrevistas, nem José nem qualquer outra pessoa negra são chamadas para as entrevistas.

Caso 2 (Variação 1) - Lucas, chefe de RH da Empresa Alfa, contrata a Empresa Beta para filtrar somente os 3% melhores currículos, que serão chamados para entrevistas, José não é chamado para as entrevistas e são chamadas somente 3 pessoas negras dentro de 100 entrevistados, metade da quantidade habitual⁴, de acordo com o histórico de seleções anteriores. Ao final das entrevistas, nenhuma pessoa negra é contratada por Lucas.

Caso 2 (Variação 2) - Lucas, chefe de RH da Empresa Alfa, contrata a Empresa Beta (que aplica IA de seleção e recrutamento) para filtrar somente os 3% melhores currículos, que serão chamados para entrevistas, incluindo José, são chamadas somente 10 pessoas negras dentro de 100 entrevistados, mais do que a quantidade habitual, de acordo com o histórico de seleções anteriores. Ao final das entrevistas, nenhuma pessoa negra é contratada.

Caso 3 – Lucas, chefe de RH da Empresa Alfa, contrata a Empresa Ômega para fazer anúncios de emprego para uma nova vaga que abriu. Os anúncios são mostrados prioritariamente para pessoas brancas, chegando a 80% de pessoas brancas e 20% de negras, em um espectro amostral de 50% de brancos e negros. Em decorrência disso, José não é chamado para as entrevistas e são chamadas somente 3 pessoas negras dentro de 100 entrevistados, metade da quantidade habitual, de acordo com o histórico de seleções anteriores.

⁴ Esse número não se ampara em qualquer dado da realidade, mas em uma situação hipotética. Para fins deste trabalho, o que importa do percentual é pensar na redução proporcional de pessoas negras nas etapas seguintes da seleção de emprego, independentemente de quais números foram apresentados especificamente.

Caso 4 – Lucas, chefe de RH da Empresa Alfa, contrata a Empresa Pi para fazer recrutamento passivo, abordando diretamente os candidatos que combinam com o perfil da vaga em sua plataforma digital de seleção e recrutamento. Em decorrência disso, nenhum candidato negro é contactado ou contratado, de modo que o contato ocorreu somente em relação ao candidato (branco) que foi efetivamente contratado. Ninguém além do candidato selecionado sequer tomou conhecimento da vaga.

Além disso, nos casos 2 a 4, existe a figura do programador Marcos, que não toma as decisões estratégicas, mas elabora os códigos pelos quais os algoritmos dotados de aprendizagem de máquina funcionam. Ainda, nestes mesmos casos, há a figura do operador Erick, que analisa os dados dos algoritmos envolvidos e que pode suspender sua execução ou requerer alterações ao programador.

Dessa forma, no Capítulo 2, o trabalho tratará do uso de Inteligência Artificial e suas formas de regulação, e, especialmente em relação às tecnologias empregadas em seleção e recrutamento, será feito o estado da arte. Além disso, buscar-se-á compreender as formas de responsabilidade penal relacionadas à IA, seja por meio da apresentação do estado da arte acerca do tema, seja pela apresentação das regras de imputação a atribuição de responsabilidade existentes no direito penal brasileiro.

O Capítulo 3, por sua vez, trata das regras vigentes em relação ao crime de racismo no direito brasileiro, das formas de racismo existentes e dos mecanismos de seu combate e, à luz desses elementos, busca-se analisar os casos delimitados, no sentido de compreender se, e em que medida, o dispositivo do art. 4º da Lei 7.716/1989 pode ser aplicado em relação ao racismo algorítmico existente em IA de seleção e recrutamento.

No Capítulo 4, serão analisadas as questões de política criminal de combate ao racismo no ambiente de trabalho, buscando tão somente compreender se as normas penais vigentes são suficientes para o combate a esta modalidade discriminatória, sem adentrar nos instrumentos pedagógicos trabalhistas⁵. Além disso, em caso de resposta negativa no contexto do direito penal, serão apresentadas considerações no sentido de se buscar um modelo diverso que responsabilize adequadamente essas distorções discriminatórias e repare adequadamente os danos causados.

No Capítulo 5 serão apresentadas as considerações finais relativas ao presente trabalho e à pesquisa realizada.

⁵ Reconhece-se aqui a existência de normas no direito do trabalho nacionais, bem como internacionais, como a Convenção 111 da OIT, além de jurisprudência do Tribunal Superior do Trabalho, destinados a inibição de práticas discriminatórias, inclusive no que tange à discriminação racial e durante processos seletivos para vagas de emprego. Contudo, o que se pretende é tão somente a pesquisa sobre a responsabilização penal, motivo pelo qual não haverá profundamente acerca das normas trabalhistas atinentes ao tema.

2. ALGORITMOS DE SELEÇÃO E RECRUTAMENTO, DISCRIMINAÇÃO E MODELOS DE RESPONSABILIDADE PENAL

Consoante Celeste (2019, p. 1-2), as tecnologias digitais possuem crescente impacto na sociedade contemporânea, o que amplia as possibilidades de exercício de direitos fundamentais no ambiente digital, mas também cria riscos para esses direitos fundamentais. Neste contexto, o autor aponta que há uma alteração do equilíbrio constitucional em face desse fenômeno, o que demanda uma resposta normativa por parte dos sistemas constitucionais. De uma forma mais ampla, autores como Hoffmann-Riem (2022, p. 29) apontam as crescentes possibilidades e riscos decorrentes da transformação digital, defendendo a necessidade de atuação do direito, dos estados e dos agentes econômicos e inovadores individuais no sentido de determinar “se e como as oportunidades oferecidas pela digitalização podem ser exploradas e os riscos minimizados”.

O’Neil (2016, p. 10) afirma que algumas dessas formas de IA se baseiam em modelos matemáticos que a autora define como Algoritmos de Destruição em Massa⁶. Segundo a autora, os algoritmos podem ser usados para ampliar desigualdades, criar distorções sociais, estando, boa parte das vezes, acima do julgamento humano, de modo que a IA pode deixar de ser usada como uma ferramenta de evolução social e gerar distorções sociais ainda maiores. Ainda, segundo Pasquale (2016, p. 509), os crescentes usos de IA baseados em *big data*, análises preditivas e aprendizagem de máquina (ML) ameaçam a reputação e a liberdade das pessoas.

Floridi et al. (2018, p. 690-694) defendem que a inteligência artificial possui capacidade de auxiliar os humanos ampliando as possibilidades de autorrealização, de tomada de decisão das pessoas⁷, de aumento das capacidades sociais e de promoção da coesão social. No entanto, segundo os autores, o uso inadequado ou exagerado da IA pode ter efeito contrário, atrapalhando a realização dessas oportunidades. Neste contexto, os autores entendem que o uso ético da IA funciona como uma forma de ampliar os benefícios e reduzir os riscos associados às aplicações de inteligência artificial (FLORIDI et al., 2018, p. 706).

Assim, enquanto a IA amplia as possibilidades de transmissão da informação, aumentando as possibilidades de propagação de informação, de comunicação e de expressão, também há risco de 1) “bloquear ou limitar tal transmissão, 2) monitorar o conteúdo das informações transmitidas e 3)

⁶ *Weapons of Math Destruction* no original.

⁷ Segundo os autores: “AI is providing a growing reservoir of ‘smart agency’. Put at the service of human intelligence, such a resource can hugely enhance human agency. We can do more, better, and faster, thanks to the support provided by AI. In this sense of ‘Augmented Intelligence’, AI could be compared to the impact that engines have had on our lives. The larger the number of people who will enjoy the opportunities and benefits of such a reservoir of smart agency ‘on tap’, the better our societies will be” (FLORIDI et al., 2018, p. 692).

registrar outras informações relacionadas aos indivíduos envolvidos na transmissão”⁸. (CELESTE, 2019, p. 3)

Existem perigos ainda maiores decorrentes do uso descontrolado da IA. Esses perigos são abordados em filmes de ficção científica, como na série *O Exterminador do Futuro*, em que uma forma de IA chamada *Skynet* (aplicada para fins militares e de defesa) possui capacidade de processar grandes quantidades de dados e de aprender com os dados processados, características que se encaixam nos conceitos de, respectivamente, *big data*⁹ e *machine learning*¹⁰. Essa tecnologia de IA toma consciência e se vê ameaçada pelos seres humanos e, em decorrência disso, faz uma série de ataques nucleares, gerando o quase extermínio da humanidade, no que é apresentado como o Dia do Juízo Final (*Judgement Day*).

Saindo da ficção, neste ano foi noticiado que uma forma de IA criada pelo Google, chamada LaMDA – uma tecnologia usada em aplicações de diálogo com humanos – teria se tornado autoconsciente, segundo um dos engenheiros de software da própria empresa. Após a revelação, este engenheiro teria sido suspenso e, posteriormente, demitido por violação às políticas de privacidade da empresa¹¹. Dessa forma, o que parecia uma visão alarmista e ficcional passa a ter alguma correspondência com a realidade, caso confirmada essa alegação.

No entanto, o presente trabalho busca compreender em que medida algumas aplicações práticas de inteligência artificial e suas consequências para as pessoas, em especial no que trata de seleção e recrutamento de empregados pode configurar o crime de racismo previsto no art. 4º da Lei 7.716/1989 e se essa proibição à discriminação é uma política criminal adequada e suficiente para combater o racismo no mercado de trabalho.

No livro *Weapons of Math Destruction*, O’Neil (2016, n.p., Cap. 6) dedica um capítulo a seleções de emprego, mostrando como pessoas portadoras de necessidades especiais, mulheres,

⁸ Tradução nossa. No original: “blocking or limiting such a transmission, 2) monitoring the content of the transmitted information, and 3) registering other information related to the individuals involved in the transmission”.

⁹ Consoante De Mauro, Greco e Grimaldi (2016, p. 130), pode-se definir *Big Data* como “the Information asset characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value”. Ainda que, devido às diferentes perspectivas em que se pode descrever *Big Data*, a apresentada pelos autores considera suas diferentes dimensões enquanto características dos dados, capacidades tecnológicas e técnicas de análise dos dados.

¹⁰ Segundo Scheaus, Arsene e Popescu (2018, p. 12): “Machine Learning can be defined as a subset of artificial intelligence and includes predictive algorithms with reference to output values, resulted as a combination of decision trees, association rules, logistic regression etc.”

¹¹ É possível ver notícia sobre a revelação em artigo divulgado na *webpage* da Revista Exame em 12 de junho de 2022. Disponível em <<https://exame.com/pop/engenheiro-do-google-diz-que-inteligencia-artificial-da-empresa-ganhou-vida/>>. Acesso em 22 de julho de 2022.

Em 23 de julho de 2022, foi publicada notícia confirmando a demissão do engenheiro por violação à política de privacidade da empresa. Disponível em: “<https://epocanegocios.globo.com/Empresa/noticia/2022/07/google-demite-engenheiro-que-disse-que-ia-da-empresa-era-autoconsciente.html>” Acesso em 27 de julho de 2022.

negros e outros grupos minoritários podem ser prejudicados por IA de seleção e recrutamento. Segundo a autora, ainda que a ciência aplicada nessas tecnologias não passe, muitas das vezes, de presunções não testadas, elas podem legitimar a exclusão de determinados grupos no mercado de trabalho, como pode ocorrer com negros e estrangeiros.

Mais especificamente em relação ao direito penal, Silva Sanchez e Urbina Gimeno (2014, p. 27) afirmam que a teoria geral do crime tem como paradigma o “homicídio consumado, comissivo e doloso, cometido de modo direto por um autor individual e isolado¹²”. Assim, a teoria geral do crime se refere a uma forma de criminalidade específica, vinculada ao direito penal clássico, e que não é, muitas das vezes, capaz de alcançar a complexidade das estruturas sociais contemporâneas e de um mundo crescentemente digitalizado.

Considerando este novo contexto social e a evolução do direito penal, Estellita e Leite (2019, p. 32) apontam a ampliação das interações entre pessoas e IA, de modo que se “coloca a questão de quem (e a que título) deverá ser responsabilizado quando o resultado típico é causado pela interposição de um sistema autônomo, ou mesmo híbrido”. Neste contexto, diversos agentes (pesquisadores, programadores, operadores) envolvidos na operação desses sistemas da IA podem ser potencialmente responsabilizados por eventual crime, de modo que “a multiplicidade de atores sugere que os complexos problemas de autoria e participação, já conhecidos no âmbito da criminalidade de empresa, surgirão, também, neste contexto” (ESTELLITA; LEITE, 2019, p. 32).

Ainda sobre a responsabilidade penal e IA, Susana Ayres de Sousa (2020, p. 70) afirma que os resultados dos algoritmos podem “concretizar-se numa opção que causa danos a interesses jurídicos protegidos pelo direito penal”. Quando isso ocorre, há uma pluralidade de pessoas envolvidas, dentre as quais “os programadores da IA, os que recolhem dados e treinam os algoritmos, aqueles que produzem as máquinas, os seus proprietários, aqueles que as introduzem no processo produtivo” (SOUSA, 2020, p. 78-79), restando saber quem pode – se é que alguém – ser penalmente responsabilizado por essa violação penalmente relevante.

Assim, quando há violação de bens jurídicos penalmente relevantes envolvendo-se a aplicação de IA, existem diversas dificuldades de compreender se (e como) aplicar o direito penal. Isso passa pela insuficiência das categorias tradicionais de um direito penal clássico. Além disso, as regras de imputação de autoria e a necessidade de existência de elemento subjetivo geram dificuldades à atribuição de responsabilidade penal nesses casos.

¹² Tradução nossa. No original: “homicidio consumado, activo y doloso, cometido de modo directo por um autor individual aislado”.

Essas dificuldades se relacionam diretamente ao objeto deste estudo, uma vez que é impossível debater a responsabilidade penal por IA de seleção e recrutamento sem compreender, anteriormente, as regras de aplicação do direito penal – e mais especificamente as vigentes no Brasil – e como elas afetarão os agentes relacionados ao desenvolvimento e operação destes algoritmos.

A partir da inquietação e das dificuldades apresentadas, passa-se a apresentar as formas de regulação de tecnologia, seus regramentos e princípios, e a seguir, a apresentar as tecnologias de IA para seleção e recrutamento e seus vieses discriminatórios (2.1). Em seguida, apresentam-se, ainda, o estado da arte em relação à responsabilidade penal decorrente de IA no Brasil e no mundo e as regras de direito penal existentes em relação às formas de IA.

2.1. O MUNDO DIGITAL E A REGULAÇÃO JURÍDICA

No contexto da regulação das novas tecnologias no mundo digital, surgem teorias de normatividade e regulação. A selecionada para guiar o presente trabalho é o constitucionalismo digital, que compreende “limitar o poder do governo e inclui, entre outros valores fundantes, a democracia, a proteção dos direitos humanos e o estado de direito¹³” (CELESTE, 2019, p.12)¹⁴. Consoante Celeste (2019, p. 17), o constitucionalismo digital consiste na adaptação dos valores do constitucionalismo às particularidades de uma sociedade digital, buscando limitar a atuação e o poder de agentes públicos e privados.

De acordo com Gill, Redeker e Gasser (2018, p. 2), o constitucionalismo digital compreende as diferentes iniciativas com objetivo de desenvolver um conjunto de “direitos políticos, normas de governança e limitações ao exercício do poder na Internet¹⁵”, de modo que o conteúdo deste constitucionalismo está ligado à extensão dos direitos civis ao contexto da esfera digital (GILL; REDEKER; GASSER, 2018, p. 3).

Autores como Suzor (2018, p. 9) defendem o controle sobre os entes privados envolvidos no ambiente digital como o principal desafio do constitucionalismo digital, que deve ser usado para “encontrar maneiras de melhorar a legitimidade da governança das plataformas, tanto por meio de normas legais quanto de obrigações sociais¹⁶”. No entanto, há divergência nas posições apresentadas

¹³ Consoante Celeste (2019, p. 12), constitucionalismo compreende “limiting the power of government, and includes, among its foundational values, democracy, the protection of human rights and the rule of law”.

¹⁴ Não se desconhece o amplo debate entre as diversas concepções de constitucionalismo, ou mesmo da tensão existente entre constitucionalismo e democracia. No entanto, para o escopo do presente trabalho, a definição de constitucionalismo apresentada é satisfatória.

¹⁵ Tradução nossa. No original: “political rights, governance norms, and limitations on the exercise of power on the Internet.

¹⁶ Tradução nossa. No original: “[...] finding ways to improve the legitimacy of Plataforma governance, through both legal rules and social obligations, is the key challenge and opportunity of digital constitutionalism”.

por autores como Gill, Redecker e Gesser, de um lado, e, de outro lado, Suzor. Embora ambos defendam a necessidade de regulação do ambiente digital, sua forma e extensão é variada. Neste contexto, Celeste (2019, p. 7) defende que:

Gill, Redeker e Gasser argumentaram que sua concepção de constitucionalismo digital não é tão restrita quanto a de Suzor, que se referiu à limitação do poder apenas no contexto das comunidades virtuais. Na realidade, ao invés de mais ampla, sua interpretação aparenta ser diferente. Para eles, o constitucionalismo digital objetiva a limitação do poder público, enquanto para Suzor a limitação do poder privado (ver Yilma 2017). Além disso, Gill, Redeker e Gasser identificaram tal função limitante nos documentos da declaração de direitos da Internet, enquanto Suzor no direito privado estatal¹⁷.

No entanto, o presente trabalho adota a definição mais ampla de Celeste (2019, p. 17), que define o constitucionalismo digital como “a ideologia que adapta os valores do constitucionalismo contemporâneo às peculiaridades da sociedade digital. Neste particular, tem sido defendido que o constitucionalismo digital não visa exclusivamente limitar o poder de atores públicos ou privados, como afirmam partes da academia, mas de ambos¹⁸”.

No cenário brasileiro, Mendes e Fernandes (2020, p. 4) definem o constitucionalismo digital como “um guarda-chuva que abrange as mais diversas iniciativas jurídicas e políticas, estatais e não estatais, voltadas à afirmação de direitos fundamentais na internet” (MENDES; FERNANDES, 2020, p. 4). A perspectiva empregada pelos autores é, portanto, a de reconhecimento da internet como espaço para acesso a direitos. Nesse contexto, surgem novas regulações no sentido de: i) reconhecer novos direitos fundamentais, como o acesso à internet; ii) limitar as violações a direitos fundamentais na internet; iii) criar novas formas de controle social em face das instituições, “como o dever de transparência das informações controladas por governos e entidades privadas” (MENDES; FERNANDES, 2020, p. 8)¹⁹.

Dessa forma, para o presente trabalho, o constitucionalismo digital compreende a aplicação dos valores constitucionais contemporâneos ao ambiente digital, podendo limitar ou determinar a atuação do Estado ou de agentes privados. Ainda que não seja formalmente constitucional de acordo

¹⁷ Tradução nossa. No original: “Gill, Redeker and Gasser argued that their conception of digital constitutionalism is not as narrow as that of Suzor, who referred to the limitation of power only within the context of virtual communities. In reality, rather than broader, their interpretation appears to be different. For them digital constitutionalism aims at the limitation of public power, while for Suzor at that of private power (see Yilma 2017). Moreover, Gill, Redeker and Gasser identified such a limitative function in the documents of Internet bill of rights, while Suzor in the private law of the state.”

¹⁸ Tradução nossa. No original: “[...] the ideology that adapts the values of contemporary constitutionalism to the peculiarities of the digital society. In particular, it has been contended that digital constitutionalism does not exclusively aim to limit the power of either public or private actors, as affirmed by part of the existing scholarship, but both of them.”

¹⁹ Os autores, debatendo o papel da jurisdição constitucional em face do constitucionalismo digital, apontam que – a despeito de um pensamento inicial de que o direito privado seria o campo adequado para a regulação das complexas relações da sociedade interconectada – os novos contextos sociais demandam a atuação das Cortes Constitucionais no sentido de aplicar “princípios e valores do constitucionalismo digital na jurisdição constitucional” (MENDES; FERNANDES, 2020, p. 11).

com as regras do direito brasileiro, o constitucionalismo digital, além dos princípios constitucionais que constituem outros direitos que podem ser afetados ou fruídos, também, em contexto digital, é marcado por ampla existência de documentos, convenções e produção doutrinária.

Além disso, alguns ordenamentos jurídicos passaram a prever de forma constitucionalmente expressa princípios relacionados ao constitucionalismo digital. No direito brasileiro, por exemplo, a Emenda Constitucional nº. 115/2022 incluiu a proteção de dados pessoais no rol de direitos fundamentais do artigo 5º da Constituição Federal²⁰. No contexto não estatal, diversas organizações apresentaram documentos com o que entendiam como os princípios que deveriam nortear a IA.

Ainda em 2018, FLORIDI *et al.* (2018, p. 696-697) já apontavam seis documentos com “princípios para políticas propostas, melhores práticas e recomendações para novas estratégias²¹”, que totalizavam 47 princípios. Os documentos referidos são: i) Asilomar AI Principles, desenvolvidos na conferência de Asilomar, promovida pelo Future of Life Institute; ii) Montreal Declaration for Responsible AI, desenvolvidos no Forum on the Socially Responsible Development of AI; iii) The General Principles, presentes na segunda edição do Ethically Aligned Design; iv) Ethical Principles, presentes no Statement on Artificial Intelligence, Robotics and ‘Autonomous’ Systems, da Comissão Europeia do Grupo Europeu em Ética na Ciência e Novas Tecnologias (EGE); v) Five Overreaching Principles for an AI Code, presentes no relatório AIUK, do Comitê de Inteligência Artificial do Senado do Reino Unido; vi) Tenants on the Partnership in AI, promovido por diversos agentes da sociedade civil.

A despeito da profusão de documentos e de princípios, existe uma tendência de repetição dos princípios e, da criação de princípios diversos com conteúdo semelhante. Em decorrência disso, e para simplificar a presente análise, fizemos um recorte de alguns princípios mais relacionados ao objeto desta dissertação. Para o presente trabalho, os princípios serão divididos em três recortes maiores: (i) os princípios que estabelecem limites éticos às aplicações de IA; (ii) os princípios relacionados à explicabilidade e à democracia no contexto algorítmico; e (iii) os princípios que buscam estabelecer responsabilidade dos agentes envolvidos em violações de direitos praticadas por meio de algoritmos.

Dentro deste primeiro grupo, um princípio apresentado é o da segurança, presente nos *Asilomar AI principles*, segundo o qual os sistemas de IA devem ser seguros e verificáveis ao longo de seu tempo de vida. Além disso, o mencionado documento ainda estabelece como princípios os valores humanos e o controle humano, sendo que este estabelece que a delegação de uma decisão ou

²⁰ Segundo o dispositivo do inciso LXXIX do art. 5º da Constituição Federal: “é assegurado, nos termos da lei, o direito à proteção dos dados pessoais, inclusive nos meios digitais”.

²¹ Tradução nossa. No original: “principles to proposed policies, best practices, and recommendations for new strategies”.

atividade a robôs deve ser uma decisão humana e aquele que os sistemas de IA devem ser compatíveis com valores humanos, como a dignidade humana, os direitos, a liberdade e a diversidade cultural²².

Ainda no grupo dos limites éticos à IA, os *Ethical Principles* da EGE estabelecem o princípio da justiça, equidade e solidariedade, princípio amplo, que, entre outras coisas, estabelece uma lógica de justiça global e acesso igual às ferramentas tecnológicas, bem como a prevenção ou correção de vieses discriminatórios nos dados usados pelas formas de IA e a vigilância em relação ao uso de dados pessoais²³.

Na Declaração de Montreal, estão previstos alguns outros princípios éticos para o uso de IA. O primeiro deles é o respeito à autonomia, segundo o qual as formas de IA não devem ser usadas para moldar/determinar um único estilo de vida, nem para criar mecanismos e vigilância e avaliação opressivos para as pessoas. Além disso, o documento prevê o princípio da solidariedade, segundo o qual as formas de IA devem ser aplicadas para gestão e distribuição equitativa dos riscos ao longo das sociedades. Há, ainda, o princípio da equidade, que determina que as formas de IA não podem ser criadas para reforçar ou reproduzir discriminações étnicas, sociais e de gênero, entre outras, e ainda, que a IA deve ser aplicada no sentido de reduzir as formas de discriminação e de vulnerabilidade social²⁴.

Sobre o segundo grupo de princípios, há ampla defesa da necessidade da explicabilidade das formas de IA, de modo que “há referências acerca da necessidade de compreender e responsabilizar pelo processo de tomada de decisão da IA²⁵” (FLORIDI *et al.*, 2018, p. 700). Nos *Asilomar AI Principles* está previsto o princípio da transparência em caso de falha, segundo o qual as falhas das formas de IA tem que ter as razões divulgadas. Na Declaração de Montreal, há o princípio da participação democrática, que estabelece que decisões sobre IA que impactam a vida das pessoas devem ser explicadas em linguagem compreensível pelas pessoas afetadas, que os códigos de IA que afetam a vida das pessoas devem ser compartilhadas com as autoridades e *stakeholders* relevantes e, ainda, que deve haver liberação dos códigos de IA usadas por autoridades públicas, exceto quando houver alto risco de mau uso das pessoas pela publicação destes códigos. Já nos EGE *Ethical Principles*, há o princípio da democracia, que estabelece a necessidade de debate público em relação às principais decisões tomadas por algoritmos.

²² Os *Asilomar AI principles* foram criados em 2017 na Conferência de Asilomar, que foi organizada pelo *Future of Life Institute*. Atualmente, o documento conta com mais de 5 mil assinaturas.

²³ Este documento foi elaborado pelo Grupo Europeu de Ética em Ciência e Novas Tecnologias da Comissão Europeia.

²⁴ Este documento foi anunciado em 2017 ao final do Fórum de Desenvolvimento Social e Responsável da IA, por iniciativa da Universidade de Montreal.

²⁵ Tradução nossa. No original: “reference is made to the need to understand and hold to account the decision-making processes of AI”.

Dentro do terceiro grupo, os três documentos citados possuem um princípio com o mesmo nome: princípio da responsabilidade. De acordo com o recorte dos Asilomar AI Principles, os desenvolvedores são responsáveis pelo uso, mau uso e ações das formas de IA, sendo os responsáveis por determinar as implicações dessa ferramenta tecnológica. Nos *Ethical Principles* da EGE, este princípio estabelece que os desenvolvedores são responsáveis pelos riscos decorrentes das formas de IA, que não podem gerar certos danos ou comprometer autonomia e liberdade humana. Este documento prevê ainda o princípio do estado de direito e *accountability*, que determina que os sistemas jurídicos devem estabelecer regras de responsabilidade decorrente de aplicações de IA.

Retornando-se ao princípio da responsabilidade, a Declaração de Montreal busca algumas formas de densificá-lo. O documento aponta que decisões que afetam a vida das pessoas devem – sempre que possível – ser tomadas por humanos e que decisões humanas com base em recomendações de IA são de responsabilidade das pessoas. O documento afirma, ainda, que pessoas que autorizam algoritmos a cometer crimes ou ofensas são por estes responsáveis, exceto quando comprovada a confiabilidade da IA e seu uso adequado, não podendo haver responsabilização de pessoas mesmo se houver eventual falha de sua operação neste último caso.

Além desses grupos de princípios, a Declaração de Montreal estabelece, ainda, o princípio da prudência. O objetivo deste princípio é garantir a segurança na aplicação das formas de IA. Em decorrência deste princípio, deve haver previsão dos riscos para as pessoas e as aplicações algorítmicas devem ser testadas antes de seu lançamento no mercado. Segundo o referido documento, a prudência determina, ainda, que falhas algorítmicas em setores que colocam em risco a segurança ou organização social devem ser compartilhadas com autoridades públicas e empresas, com o objetivo de prevenir novos eventos semelhantes.

Dessa forma, a despeito da existência de diversos princípios do constitucionalismo digital os que possuem papel central no presente trabalho são os relacionados às limitações éticas e à discriminação, à explicabilidade, à responsabilidade e à prudência.

Além das questões constitucional e principiológica existentes no constitucionalismo digital, há extenso debate sobre as possibilidades e meios para a regulação da Inteligência Artificial. Sobre o tema, Floridi (2021, p. 620) aponta que, inicialmente, a autorregulação dos agentes privados atuantes para desenvolvimento das novas tecnologias parecia ser ferramenta suficiente para seu controle ético, entendendo-se que as empresas poderiam criar e controlar os padrões éticos que deveriam seguir, “sem a necessidade de controles ou imposições externas²⁶”. No entanto, segundo o autor:

²⁶ Tradução nossa. No original: “the digital industry could formulate its own ethical codes and standards as well as request and monitor adherence to them, without the need for external controls or impositions”.

Em última análise, as empresas pareciam relutantes ou incapazes de resolver seus problemas éticos, não necessariamente em termos de recursos, lobby e relações públicas, mas em termos de estratégia de alto nível, no nível executivo, para melhorar a mentalidade e os comportamentos errados que estavam muito profundamente enraizados²⁷. (FLORIDI, 2021, p. 621)

Segue o autor, afirmando que, em consequência disso, houve a criação de diversos regramentos no âmbito da União Europeia (UE), afastadas as possibilidades de um cenário unicamente autorregulado, citando normas como a GDPR, de 2016, e projetos de normas, como o caso do Digital Markets Act, do Digital Services Act e do Artificial Intelligence Act (AIA), ainda pendentes de aprovação definitiva pelos países-membros da UE.

Dessas normas europeias, a que trata mais diretamente do tema da regulação de inteligência artificial é o AIA. Nesta proposta legislativa, há o estabelecimento de critérios distintos de risco para as formas de inteligência artificial, dividindo-as em 3 categorias: i) risco inaceitável; ii) alto risco; iii) risco baixo ou mínimo²⁸ (EUROPEAN COMMISSION, 2021, p. 12). A despeito de debate mais amplo sobre as classificações previstas, interessa ao presente trabalho os algoritmos de seleção e recrutamento constarem no rol das aplicações de alto risco.

Neste contexto, o anexo III do Artificial Intelligence Act define como aplicações de alto risco as relacionadas a:

4. Emprego, gestão de empregados e acesso a emprego:
 - (a) sistemas de Inteligência Artificial aplicados para recrutamento e seleção de pessoas naturais, especialmente para publicação de vagas, triagem de candidaturas, avaliação de candidatos em entrevistas ou testes;
 - (b) IA destinada para tomada decisões sobre promoção e rescisão de relações contratuais relacionadas ao trabalho, para definição de tarefas e para monitorar e avaliar o desempenho e o comportamento das pessoas nessas relações.²⁹ (EUROPEAN COMMISSION, 2021, p. 5)

Em relação a essas aplicações tidas como de alto risco, o AIA estabelece um regramento específico, criando um sistema de gerenciamento de riscos³⁰ que, entre outras providências, estipula:

²⁷ Tradução nossa. No original: “Ultimately, companies appeared to be reluctant or unable to solve their ethical problems, not necessarily in terms of resources, lobbying, and public relations, but in terms of top-level strategy, at the C-suite level, to improve mentality and wrong behaviours that were just too deeply rooted”.

²⁸ O Título II do referido documento apresenta as formas proibidas de inteligência artificial, enquanto o Título III mostra as aplicações de alto risco.

²⁹ Tradução nossa. No original: “4. Employment, workers management and access to self-employment: (a) AI systems intended to be used for recruitment or selection of natural persons, notably for advertising vacancies, screening or filtering applications, evaluating candidates in the course of interviews or tests; (b) AI intended to be used for making decisions on promotion and termination of work-related contractual relationships, for task allocation and for monitoring and evaluating performance and behavior of persons in such relationships.”

³⁰ Art. 9. O inciso 2 do artigo aponta que as aplicações de IA de alto risco devem seguir os seguintes requisitos: “(a) identification and analysis of the known and foreseeable risks associated with each high-risk AI system; (b) estimation and evaluation of the risks that may emerge when the high-risk AI system is used in accordance with its intended purpose and under conditions of reasonably foreseeable misuse; (c) evaluation of other possibly arising risks based on the analysis

i) necessidade de testagem dos sistemas³¹; ii) criação de regras de governança para treinar e validar formas de IA³²; iii) necessidade de documentação técnica prévia ao lançamento dos sistemas³³; iv) necessidade de transparência da operação do sistema para os usuários³⁴; v) necessidade de que o sistema forneça diversas informações³⁵; vi) necessidade de supervisão humana nos sistemas de IA de alto risco³⁶; vii) preocupações com a precisão dos sistemas e redução das possibilidades de falhas³⁷; viii) entre outras obrigações que devem ser seguidas pelos provedores de IA classificadas como de alto risco³⁸.

Dessa forma, saindo-se de uma regra geral de liberdade na criação de sistemas de IA, no caso das aplicações de alto risco, como nos algoritmos de recrutamento e seleção, o AIA busca impor, no contexto da União Europeia, uma série de restrições, limites e controles à operação de algoritmos³⁹. A despeito dessas restrições, contudo, não há previsão de normas penais nessa proposta legislativa.

No contexto brasileiro, a regulação de temas atinentes ao direito digital também tem caminhado. Neste contexto, em 2015, foi editado o Marco Civil da Internet, no sentido de regular o espaço digital e de proteção da privacidade dos usuários. Em 2018, foi editada a Lei Geral de Proteção de Dados Pessoais, que, segundo Oliveira e Lopes (2020, p. 71), buscou sistematizar as normas

of data gathered from the post-market monitoring system referred to in Article 61; (d) adoption of suitable risk management”.

³¹ De acordo com o art. 9.7: “The testing of the high-risk AI systems shall be performed, as appropriate, at any point in time throughout the development process, and, in any event, prior to the placing on the market or the putting into service. Testing shall be made against preliminarily defined metrics and probabilistic thresholds that are appropriate to the intended purpose of the high-risk AI system.”

³² Segundo estabelecido no art. 10.2 do AIA, deve haver: “Training, validation and testing data sets shall be subject to appropriate data governance and management practices. Those practices shall concern in particular, (a) the relevant design choices; (b) data collection; (c) relevant data preparation processing operations, such as annotation, labelling, cleaning, enrichment and aggregation; (d) the formulation of relevant assumptions, notably with respect to the information that the data are supposed to measure and represent; (e) a prior assessment of the availability, quantity and suitability of the data sets that are needed; (f) examination in view of possible biases; (g) the identification of any possible data gaps or shortcomings, and how those gaps and shortcomings can be addressed.”

³³ Art. 11.

³⁴ Art. 12.

³⁵ Art. 13.

³⁶ O art. 14,2 determina que o sistema deve ser organizado de uma forma em que a pessoa natural responsável por sua supervisão deve ser capaz de: “(a) fully understand the capacities and limitations of the high-risk AI system and be able to duly monitor its operation, so that signs of anomalies, dysfunctions and unexpected performance can be detected and addressed as soon as possible; (b) remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (‘automation bias’), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons; (c) be able to correctly interpret the high-risk AI system’s output, taking into account in particular the characteristics of the system and the interpretation tools and methods available; (d) be able to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse the output of the high-risk AI system; (e) be able to intervene on the operation of the high-risk AI system or interrupt the system through a “stop” button or a similar procedure.”

³⁷ Art. 15.

³⁸ Art. 16.

³⁹ Sobre o cenário estadunidense, *vide* 2.3.3.

anteriores que tratavam do tema, promovendo “uma orientação para organização formal desses dispositivos”.

Atualmente, tramita projeto de criação de lei mais específica para a IA, o Marco Regulatório da Inteligência Artificial. Neste contexto, a despeito da existência do Projeto de Lei 21/2020, aprovado pela Câmara dos Deputados, com previsão de responsabilidade civil subjetiva para a inteligência artificial⁴⁰, a comissão de juristas instituída pelo Senado Federal tem debatido questões como o modelo de responsabilidade civil e as normas de regulação da inteligência artificial no país. Desta forma, após a elaboração de Anteprojeto do Marco Regulatório da IA pela comissão, certamente haverá uma nova rodada de debates legislativos sobre o tema, sendo o cenário vigente marcado por pouca regulação sobre o tema no cenário nacional.

No entanto, considerando-se este cenário atual, o AIA, no contexto europeu, inova ao prever a proibição de determinadas formas de IA e, em relação a outras, como as de seleção e recrutamento, determina a necessidade da existência de controles prévios à liberação destas tecnologias, e também de alguns controles posteriores. No cenário brasileiro, é importante que se trave debate direcionado a endereçar essas questões.

Uma questão fundamental – neste contexto de controle e regulação de IA – é a transparência em relação aos algoritmos, especialmente em face de ideias como a de segredo industrial. Em relação ao tema, Pasquale (2015, p. 6-9 e *passim.*) defende que os algoritmos (especialmente das *Big Techs*⁴¹) funcionam como verdadeiras caixas-pretas cujo modo de funcionamento e operação se desconhece. Neste contexto, as *Big Techs* se defendem de ter que entregar seus algoritmos para análise amparados na defesa de um segredo industrial em face de seus competidores.

Por outro lado, Kroll (2018, *passim.*), sobre a inescrutabilidade dos algoritmos, afirma que, como estes refletem propostas, escolhas e premissas humanas, e, além disso, são aplicados em usos determinados e específicos, nenhum sistema é completamente inescrutável. No entanto, segundo o autor, uma verdadeira governança de algoritmos deveria passar pela análise da engenharia e *design* dos sistemas, permitindo que se conheça as escolhas, premissas e propostas que determinaram a criação de determinada forma de IA. Assim, a responsabilização decorrente de IA passa pela premissa de que algoritmos refletem escolhas humanas que determinaram as possibilidades de sua

⁴⁰ O presente trabalho não se debruçará sobre este projeto de lei, uma vez que há grandes chances de o anteprojeto da comissão de juristas alterar substancialmente o projeto, razão pela qual foi feita a opção por apenas apresentar a existência do projeto de lei e sua tramitação, mas sem detalhamento de seu conteúdo. Cumpre, aqui, outro esclarecimento. Em relação ao AIA, a despeito de ainda não ter sido aprovado, há previsão de regulação mais extensiva e detalhada das formas de inteligência artificial, razão pela qual entendeu-se por incluí-lo no escopo da presente pesquisa. Além disso, a importância dessa proposta legislativa é ampliada pela possibilidade de regulações europeias virarem o padrão mundial, segundo o que Bradford (2012, *passim.*) definiu como Efeito Bruxelas.

⁴¹ Nome convencionalmente dado às grandes empresas da área de tecnologia.

programação e de sua atuação, devendo as pessoas ser responsáveis pelos resultados decorrentes da operação destes artefatos (KROLL, 2018, p. 7).

Dado esse cenário em relação aos algoritmos e suas formas de regulação, passa-se a apresentar o cenário vigente em relação aos algoritmos destinados especificamente às atividades de seleção e recrutamento.

2.1.1. Algoritmos de seleção e recrutamento: estado da arte

Uma vez que o objetivo central deste trabalho é a compreensão da responsabilidade penal decorrente de discriminação existente em algoritmos de seleção e recrutamento, é necessário que se faça revisão de literatura sobre esta aplicação de Inteligência Artificial. Para isso, será apresentado o estado da arte no Brasil e no mundo em relação ao tema.

No contexto brasileiro, o recorte dos textos foi feito considerando os artigos de periódicos científicos segundo o mecanismo de busca da Biblioteca do Senado⁴², que indexa diversas bibliotecas e periódicos. Neste mecanismo de busca, foram feitas 4 combinações de palavras: i) “algoritmos” + “recrutamento”; ii) “algoritmo” + “recrutamento”; iii) “algoritmo” + “emprego”; iv) “algoritmos” + “emprego”. Com base nesse recorte, foram encontrados 6 artigos científicos, que serão analisados em tópico específico (2.1.2.).

No caso dos textos brasileiros, não houve necessidade de um recorte temporal, uma vez que todos os textos encontrados são posteriores a 2018. No entanto, seja para os textos nacionais, seja para os estrangeiros, por uma questão de limitação de tempo para a realização da pesquisa e necessidade de dedicação aos demais capítulos da pesquisa, o estado da arte adota como fim de seu recorte temporal o dia 12 de julho de 2022, data em que foi finalizada a seleção dos textos analisados.

Em relação aos textos de língua estrangeira, foi usada a JSTOR⁴³ como ferramenta de pesquisa e seleção dos textos, também por indexar diversos periódicos científicos, possibilitando uma busca ampla. Para os mecanismos de busca, foi feita a seguinte combinação de elementos de pesquisa: “*artificial intelligence’ job selection*”. Foi feito um recorte temporal a partir de 2018 e selecionados somente artigos de periódicos científicos relacionados às seguintes áreas do conhecimento: *Law, Criminology & Criminal Justice, e Labor & Employment*. Com base nos critérios estabelecidos, foram encontrados 10 artigos científicos, que serão apresentados no tópico indicado⁴⁴ (2.1.3).

42 Disponível em: <<https://www12.senado.leg.br/institucional/biblioteca>>. Acesso em 12 de julho de 2022.

43 Disponível em: <<https://www.jstor.org/>>. Acesso em 12 de julho de 2022.

44 Existiam outras possibilidades de critérios de busca, mas, a despeito de uma profusão de artigos científicos relacionados ao tema de IA e mercado de trabalho, muitos destes tratavam de temas outros, como as relações de trabalho e o emprego de IA, mas sem relação com seleção e recrutamento, por exemplo. Desta forma, os termos de busca e

2.1.2. Estado da arte no Brasil

Inicialmente, cumpre esclarecer que, dos 6 artigos encontrados, 2 não se relacionam diretamente ao tema do presente trabalho⁴⁵. Em relação aos demais artigos científicos, é possível perceber que todos se preocupam com as possibilidades de discriminação praticada por meio de algoritmos de seleção e recrutamento, apontando a necessidade de mais transparência e mecanismos de mitigação de eventuais violações a direitos.

Em seu texto, Parreira (2020, *passim*.) aponta que são crescentes os usos de IA no mercado de trabalho, especialmente em relação a algoritmos de recrutamento e seleção. Dessa forma, a autora defende a necessidade de transparência em relação a estes algoritmos como forma de combate à discriminação no mercado de trabalho.

O segundo texto apresenta o crescimento da IA no direito e suas crescentes influências nas relações de trabalho. Apresenta a questão da discriminação de gênero, com dados ao final apontando, inclusive, a necessidade de pesquisas futuras sobre o tema. A pergunta que se buscou responder no artigo foi a seguinte: “os algoritmos usados para o recrutamento de trabalhadores pela Amazon comportam-se de maneira discriminatória em relação a gênero?” (BORGES; FILÓ, 2021, p. 220).

Os autores apresentam um caso em que algoritmos promoveram discriminação de mulheres na seleção de empregos e na análise de desempenho da empresa Amazon, apontando que:

Em 2014, a gigante Amazon.com, por meio de um centro de engenharia, objetivou construir uma inteligência artificial capaz de selecionar e localizar os melhores candidatos a vagas de empregos rapidamente por meio de seleção baseada em dados anteriores. Os algoritmos foram habilitados para reconhecer aproximadamente 50 mil termos que apareciam nos resumos de candidatos dos 10 anos anteriores, conectando habilidades com funções específicas de determinado emprego. Assim, desenvolveu uma ferramenta de inteligência artificial capaz de avaliar os candidatos ao trabalho dando-lhes avaliações ranqueadas de 1 a 5 estrelas, sendo posteriormente comparada pelo público com as avaliações de produtos da empresa. (BORGES; FILÓ, 2021, p. 233-234)

No entanto, uma vez que os critérios de seleção se basearam em dados das seleções anteriores, a dominância masculina nos padrões de seleção foi mantida, de modo que os algoritmos entenderam que “homens eram preferíveis, suscitando em avaliações melhores ao gênero masculino”,

parâmetros estabelecidos são uma forma de buscar minimizar a incidência de artigos que tangenciem os temas, mas não se relacionem ao recorte que se busca desenvolver. Ainda assim, o recorte proposto encontrou artigos relacionados a outros temas.

⁴⁵ O artigo de Pires Fincato e Wünsch (2020) não se relaciona ao tema deste trabalho, pois busca compreender as relações de trabalho em face da ampliação da IA, mas não especificamente de recrutamento e seleção, tema sobre o qual se faz este estado da arte. Guimarães (2019), por sua vez, não trata de mercado de trabalho, mas da utilização de algoritmos preditivos no contexto de decisões judiciais, motivo pelo qual seu texto foi excluído desta análise.

de modo que “os currículos que continham a palavra ‘mulher’ eram penalizados com avaliações baixas.” (BORGES; FILÓ, 2021, p. 234)

Em 2015, após a empresa reconhecer o viés discriminatório de gênero existente, o algoritmo foi alterado para buscar a neutralidade de gênero e, posteriormente, cancelado, “já que a simples edição dos termos para um padrão neutro, não deleta a possibilidade de a discriminação continuar presente por meio do aprendizado automatizado” (BORGES; FILÓ, 2021, p. 234). Segundo os autores, este caso é um exemplo de que, a despeito do uso de tecnologia para reduzir a influência de vieses e preconceitos humanos, “a IA é um produto das decisões humanas passadas, que são suscetíveis ao viés humano, podendo ser distorcidas por preconceitos institucionais e históricos” (BORGES; FILÓ, 2021, p. 235).

Borges e Filó (2021, p. 237-238) apontam que diversos direitos podem ser violados neste contexto de acesso desigual ao mercado de trabalho, incluindo-se a livre escolha. No artigo, ainda é apresentada outra forma de IA para determinação dos valores de salário pelos quais um candidato selecionado estaria disposto a aceitar o emprego, criando possibilidades de aumento da disparidade salarial entre as pessoas, o que pode prejudicar grupos socialmente menos favorecidos, como as mulheres, que em 60% dos casos recebem salários menores que os homens.

Em decorrência disso, o artigo – antes de apontar necessidade de pesquisas futuras sobre o tema – conclui que “a problemática do uso das inteligências artificiais nos recursos humanos pode se tornar um problema social cada vez maior, perpetuando a discriminação de gênero contra as mulheres nas relações de trabalho, o que ficou comprovadamente claro no caso de processo de contratação de trabalhadores pela Amazon” (BORGES; FILÓ, 2021, p. 239).

O terceiro artigo aborda a regulação de novas tecnologias, tratando da aplicação de IA em relação a recrutamento. Inicialmente, descreve o caso de Maria: uma mulher que teve que trabalhar durante sua graduação em Direito, não podendo estagiar ou incrementar seu currículo para seleções de emprego. Dessa forma, mesmo que possuísse uma história de vida impactante, o que poderia ser relevante em uma entrevista de emprego, a seleção automatizada de seu currículo impedia seu acesso às entrevistas de emprego, dificultando sua entrada nesse novo mercado de trabalho especializado (BRUNETTA; LEITÃO; DIAS, 2022, p. 165) como advogada.

Assim, a despeito da busca da imparcialidade pelo algoritmo, as pessoas que o elaboram definem quais critérios de seleção são aplicáveis. Dessa forma, é possível que os vieses das pessoas estejam presentes na IA justamente neste momento de definição de seus parâmetros de aplicabilidade (BRUNETTA; LEITÃO; DIAS, 2022, p. 168).

Baseados nas ideias de O'NEIL⁴⁶, os autores descrevem o caso da IA criada no contexto do sistema Impact de avaliação de professores. Segundo apresentado, esse sistema de avaliação fez com que 205 professores fossem demitidos em virtude de notas baixas nessas avaliações. No entanto, sequer houve a divulgação dos parâmetros usados para estabelecer o modelo, indicando que uma forma de IA sem critérios claros era usada para definir quem manteria seu emprego ou seria demitido (BRUNETTA; LEITÃO; DIAS, 2022, p. 170-172).

Em seguida, os autores apresentam a necessidade de *feedback* dos sistemas estatísticos, para que suas respostas sejam adequadas, apontando dois modelos de aprendizagem de máquina:

De fato, o *machine learning* – cerne da IA e dos modelos algorítmicos dinâmicos – necessariamente envolve dois tipos de técnicas básicas de “aprendizado”: a supervisionada (reflexiva), que treina o modelo com dados conhecidos para que ele possa prever resultados; e a não supervisionada, que envolve a busca de padrões ocultos e estruturas implícitas nos dados recebidos.

Deixando de reconhecer que o aprendizado supervisionado é relevante para seus modelos decisórios automatizados, o homem passa a utilizar o progresso apenas em seu viés de perpetuação de preconceitos e preferências – sobretudo se o mecanismo de IA em uso é dotado de capacidade de autoaprendizagem, por meio da qual pode vir a reprogramar seus critérios e métricas com esteio numa ideia muito geral de eficiência ou de melhora da produtividade no trabalho. (BRUNETTA; LEITÃO; DIAS, 2022, p. 172)

Em decorrência disso, os autores concluem pela necessidade de uma abordagem voltada para a concretização dos direitos humanos, com supervisão e controle de IA usada para determinar acesso ao mercado de trabalho, de modo que estes algoritmos “devam estar disponíveis aos órgãos de fiscalização – com a reserva de sigilo, se for o caso” (BRUNETTA; LEITÃO; DIAS, 2022, p. 176).

Marques e Martinez Neto (2022), por sua vez, apontam as possibilidades de distorção e discriminação em algoritmos de seleção e recrutamento, apresentando casos de ocorrência deste fenômeno. Um dos casos é o da seleção da faculdade de medicina do St. George's Hospital, sediado em Londres, que, para reduzir custos e ampliar a eficiência da seleção, criou um algoritmo para filtrar os candidatos, mas este mecanismo passou a replicar os padrões discriminatórios das seleções anteriores – usadas como parâmetros para a IA de recrutamento – especialmente em relação a etnia e gênero. Neste sentido, os autores apontam uma tendência crescente de aplicação de IA inclusive para fases posteriores de seleção, como nas entrevistas, ampliando ainda mais os riscos discriminatórios nas seleções de emprego (MARQUES; MARTINEZ NETO, 2022, p. 208-209).

Além disso, segundo os autores, os algoritmos de seleção de emprego conseguem rastrear os históricos dos candidatos e seus comportamentos em redes sociais, ampliando riscos aos trabalhadores, decorrentes de sua vigilância excessiva (MARQUES; MARTINEZ NETO, 2022, p.

⁴⁶ Como o texto de O'Neil (2016) não é artigo de periódico científico, ficou fora do escopo da análise realizada. No entanto, suas ideias estão presentes nos artigos de outros autores e são citadas em outros momentos desta dissertação.

210-213). Marques e Martinez Neto também apontam para a necessidade de que os sindicatos usem as negociações coletivas de trabalho como formas de regular a aplicação de IA nas seleções de emprego, a fim de evitar discriminação e fiscalização excessiva da vida privada (MARQUES; MARTINEZ NETO, 2022, p. 213-217).

Com base nos textos analisados, é possível compreender a descrição geral de um problema, e, conseqüentemente, a necessidade geral de uma solução. Esse problema se caracteriza pelas possibilidades de discriminação provenientes dos vieses presentes em algoritmos de recrutamento. A solução, por sua vez, ainda que não definida com precisão, parece passar pela existência de controles maiores em relação a estes algoritmos e à necessidade de transparência, compreensão e supervisão humana em relação a sua operação.

2.1.3. Estado da arte em língua inglesa

Em relação aos artigos de periódicos científicos em língua inglesa, dos 10 encontrados, somente 2 se relacionam diretamente ao tema da pesquisa, motivo pelo qual os 8 restantes não serão analisados. Passa-se, portanto, a apresentar os artigos pertencentes ao recorte deste trabalho.

No primeiro texto analisado, Gay e Kagan (2018, p. 192) apresentam suas definições de (i) IA e de (ii) aprendizagem de máquina (*machine learning*), que significam, respectivamente: i) quando um computador realiza tarefas atribuíveis à inteligência humana; ii) algoritmos programados para aprender sozinhos. Neste sentido, apontam que *big data* se configura pelas informações cruas abundantemente disponíveis e que as técnicas de ML servem para tratar e categorizar esses dados crus (GAY; KAGAN, 2018, p. 197).

Sobre aplicação de IA no mercado de trabalho, apontam um uso positivo, que é para prevenir acidentes de empregados. Essa aplicação ocorre ao se levantar os dados dos acidentes passados, buscando compreender em que condições ocorreram os acidentes anteriores, a fim de evitar que se reduzam as chances de repetição de uma dessas situações, apresentando-se uma aplicação algorítmica no mercado de trabalho em benefício dos empregados (GAY; KAGAN, 2018, p. 198-199).

Por outro lado, apontando os perigos do uso de ML e análise preditiva, os autores apontam o papel humano para gerar os códigos iniciais dessas tecnologias, o que pode contaminar – com preconceitos e vieses implícitos ou propositais – o mecanismo de operação do algoritmo. Além disso, os engenheiros e programadores, em alguns casos, não conhecem as características dos sistemas, de modo que pode haver dados incorretos nos algoritmos, gerando distorções em sua operação. Além disso, os parâmetros de seleção só funcionariam adequadamente se não houvesse vieses, o que é

difícil de controlar ante o elevado volume de dados existentes, tornando difícil a tarefa de controlar as aplicações de IA em seleção e recrutamento (GAY; KAGAN, 2018, p. 202-206).

Sobre as consequências jurídicas da seleção de empregos por IA, os autores apontam que não há regras claras de governança e que, portanto, a adequação das decisões deve-se basear nos parâmetros dos antigos precedentes acerca de litígios trabalhistas. Assim, os autores apontam que as antigas formas de regulação são insuficientes para endereçar essa nova demanda das sociedades contemporâneas (GAY; KAGAN, 2018, p. 207-208). Em decorrência disso, Gay e Kagan (2018, p. 209) concluem que, a despeito das possibilidades e oportunidades de emprego que podem ser geradas pela aplicação de novas tecnologias no mercado de trabalho, bem como do melhoramento das práticas trabalhistas, deve haver constante vigilância dessas ferramentas para que não haja danos aos trabalhadores e violações de seus direitos.

No segundo artigo analisado, aponta-se o crescimento do uso de redes sociais para a seleção de talentos (recrutamento), especialmente pelo emprego de plataformas como ZipRecruiter, LinkedIn e Monster. Acadêmicos comentam dos riscos de discriminação algorítmica e para a privacidade das pessoas. Assim, o artigo busca compreender o funcionamento das empresas de tecnologia na seleção de empregos e as consequências dessas novas condições de acesso ao mercado de trabalho (KIM, 2020, p. 870-872)

Neste sentido, a autora aponta que o uso de IA e análise preditiva pode funcionar por meio de diversos mecanismos, como: anúncios de emprego, *job-matching platforms* (em que se busca a correspondência entre perfil do candidato e os interesses da empresa), sistemas de recrutamento passivo (KIM, 2020, p. 875-881).

Em seguida, a autora aponta possibilidades de discriminação por meio dessas aplicações algorítmicas. No caso dos anúncios de emprego, a empresa pode escolher para quais grupos anunciar (ex.: só homens, só pessoas de um determinado local), bem como fazendo recorte com base em interesses e comportamento, outro mecanismo para selecionar determinados grupos e reproduzir vieses de seleção. A seleção geográfica, por exemplo, pode ser usada para segregar, ante os padrões étnico-geográficos existentes. Esses riscos são ainda maiores em casos como do Facebook, que permitia a anunciantes, incluindo empregadores, excluir público-alvo pertencente a determinados grupos, como afro-americanos, asiáticos americanos e hispânicos (KIM, 2020, p. 881-882). Além disso, empresas como Google e LinkedIn possibilitam a seleção de alvos por idades, e foram encontrados casos de anúncios separados por gênero no Facebook.

Ao ser questionado, o Facebook declarou que seus critérios de seleção são neutros e que os anunciantes são responsáveis pelos critérios empregados. Para encerrar os processos que recebeu em decorrência dessas práticas, o Facebook criou um mecanismo diferenciado para anúncios de emprego,

moradia e crédito, com a proibição do uso de parâmetros de seleção como idade, gênero ou CEP (KIM, 2020, p. 882-884).

No mercado digital, é comum que os anúncios sejam direcionados a atingir de forma mais direta os interesses de cada grupo (o chamado *microtargeting*), bem como para maximizar os lucros da empresa intermediária dos anúncios. Ainda que a discriminação explícita – com determinação de exclusão de determinados grupos do mercado de trabalho – seja mais fácil de condenar, a mera distribuição desigual das informações sobre oportunidades de trabalho pode gerar efeitos semelhantes. Isso porque há formas indiretas de discriminar entre gênero e raça, por exemplo, sem que haja uma definição expressa nesse sentido, e, mesmo com critérios neutros de distribuição determinados pelos anunciantes, as empresas intermediárias podem distribuir desigualmente entre diversos grupos com recortes de raça, gênero e outras características (KIM, 2020, p. 885-886).

Dessa forma, a autora cita estudos que apontaram diferença de acesso a anúncios, sendo que os de empregos que pagam maiores salários são mostrados em maior proporção para homens⁴⁷. Outro estudo citado⁴⁸ aponta que, em cargos de ciência, tecnologia, engenharia e matemática, mesmo com anúncios neutros em relação a gênero, os anúncios foram mostrados 20% a mais a homens do que a mulheres. Tal discrepância, segundo a autora, não está relacionada ao maior interesse dos homens nos cargos, já que as mulheres que viram os anúncios clicaram neles mais que os homens (KIM, 2020, p. 887-888).

Outro estudo citado busca compreender as razões para a distribuição desigual, entre os grupos recortados, de anúncios de emprego neutros em relação a gênero e raça, atribuindo essa responsabilidade a questões de mercado e ao algoritmo do Facebook⁴⁹. Ao testar a questão da valorização das mulheres no mercado de anúncios, o estudo concluiu que, quanto menor o orçamento, menor a proporção de mulheres no público atingindo, confirmando a influência desta causa de distinção (KIM, 2020, p. 888). Dessa forma, o público feminino pode ser prejudicado por ser mais valorizado no mercado publicitário.

Ainda em relação à plataforma Facebook, percebeu-se que os algoritmos das plataformas também influenciam essa disparidade de distribuição das informações, de modo que anúncios para empregos de lenhador foram mostrados para uma proporção de 90% de homens e 70% de brancos, enquanto de zelador foram mostrados para 65% de mulheres e 70% de negros. Assim, se por um lado o acordo do Facebook retirou dos anunciantes a possibilidade de intencionalmente discriminar em

⁴⁷ Para aprofundamento no tema, *vide*: (DATTA et al., 2018).

⁴⁸ Para uma visão mais detalhada, *vide*: (LAMBRECHT; TUCKER, 2019)

⁴⁹ Para aprofundamento no tema, *vide*: (ALI et al., 2019).

relação à seleção de empregos, os algoritmos da própria empresa ainda podem fazer esses recortes (KIM, 2020, p. 888-889).

Em seguida, o texto passa a explicar o caso Amazon, em que, ao criar um algoritmo para encontrar candidatos para recrutar possíveis funcionários, foram usados os dados das contratações anteriores, majoritariamente de homens, de modo que o algoritmo aprendeu a reduzir a importância dos currículos femininos pelo simples histórico da empresa (KIM, 2020, p. 891).

Outro caso semelhante apresentado é o da empresa ZipRecruiter, que busca encontrar correspondência (*match*) entre empresas e possíveis trabalhadores. Segundo sua operação, se o empregador seleciona uma determinada pessoa, o algoritmo vai enviar pessoas com perfis parecidos. Após a segunda pessoa ser selecionada, o algoritmo entende esse perfil como uma ótima correspondência (*great match*), gerando um novo padrão de contratação (KIM, 2020, p. 891). No entanto, a autora aponta a dificuldade de compreender se o padrão selecionou o candidato, ou se, por outro lado, o candidato gerou o padrão de contratação. Além disso, aponta que o fato de determinados empregos mais associados a um grupo, como enfermagem para mulheres, ainda que parta de um dado da realidade, atrapalha a alteração desses quadros sociais (KIM, 2020, p. 891-892).

Assim, o papel das plataformas digitais é determinante para a distribuição de oportunidades de emprego, ante o padrão vigente de seu uso para recrutamento e seleção. Nas exatas palavras de KIM (2020, p. 892):

As próprias plataformas desempenham um papel importante em relação à distribuição de oportunidades de emprego, porque eles criam os algoritmos de *targeting* e *matching* que controlam os fluxos de informação. Por essa razão, os efeitos de suas escolhas quando fazem esses sistemas demandam exame minucioso⁵⁰.

A autora afirma que os algoritmos de análise preditiva determinam os fluxos de informações relativas, dentre outros temas, ao mercado de trabalho. Desse modo, estes algoritmos “possuem o potencial de manipular⁵¹ os mercados de oportunidade⁵²” (KIM, 2020, p. 892). Essa forma de manipulação, segue a autora, gera preocupações em relação aos riscos discriminatórios que dela possam emergir. Assim, as empresas de tecnologia não são meros intermediários passivos, mas acabam por definir e organizar os padrões de exibição de anúncios de acordo com seus interesses – e não somente de acordo com o determinado por seus clientes ou usuários (KIM, 2020, p. 893-894).

⁵⁰ Tradução nossa. No original: “The platforms themselves play an important role in how job opportunities are distributed because they design the targeting or matching algorithms that control information flows. For that reason, the effects of their choices when building these systems warrant closer scrutiny.”

⁵¹ Ainda que haja uma dificuldade de definir manipulação, a autora define como o exercício de influência velada sobre os outros (KIM, 2020, p. 893).

⁵² Tradução nossa. No original: “have the potential for manipulating opportunity markets”.

Além disso, a autora apresenta a contradição existente nos algoritmos preditivos, apontando que sua forma de operação é também geradora das possibilidades discriminatórias, afirmando que (KIM, 2020, p. 895):

Algoritmos preditivos são construídos para descobrir padrões existentes nos dados, e um dos padrões contínuos na vida econômica americana é a distribuição desigual de oportunidades em recortes distintos de raça, gênero e outras características pessoais. Como resultado, o próprio recurso que torna os algoritmos tão poderosos – sua capacidade de reconhecer padrões e prever comportamentos – pode fazer com que eles reproduzam padrões existentes de segregação e exclusão⁵³.

A autora, ainda que não defenda que as empresas de tecnologia que intermediam os processos de anúncio e contratações de emprego estejam discriminando grupos sociais deliberadamente, afirma que há diversos riscos discriminatórios decorrentes das práticas vigentes dessas empresas, que “arriscam reproduzir os padrões de segregação⁵⁴” que as legislações sobre direitos civis buscam coibir (KIM, 2020, p. 898)⁵⁵.

Em seguida, a autora passa a tratar dos argumentos em defesa das empresas de tecnologia que atuam neste contexto de recrutamento e seleção, e dessa aplicação prática de IA. Se por um lado argumenta-se que já havia discriminação (possivelmente mais frequente) antes da atuação da tecnologia para auxiliar recrutamento e seleção, por outro, os agentes responsáveis pela discriminação – como sindicatos e agências de emprego – poderiam ser responsabilizados em caso de sua ocorrência. Assim, ainda que pudesse haver riscos maiores, havia como controlar e punir a concretização desses riscos e havia como identificar as pessoas responsáveis por eventual discriminação, o que não ocorre no cenário atual de IA de recrutamento e seleção (KIM, 2020, p. 898-899).

Sobre os algoritmos de *job matching*, a defesa é que eles funcionam como forma de otimizar a compatibilidade entre empregado e empresa contratante. No entanto, a autora aponta que, uma vez

⁵³ Tradução nossa. No original: “Predictive algorithms are built to uncover existing patterns in the data, and one of the enduring patterns in American economic life is the unequal distribution of opportunities along the lines of race, gender and other personal characteristics. As a result, the very feature that makes algorithms so powerful – their ability to recognize patterns and predict behaviors – may cause them to reproduce existing patterns of segregation and exclusion”.

⁵⁴ Tradução nossa. No original: “[they] risk reproducing the patterns of segregation”.

⁵⁵ Ainda que a existência destes dados sensíveis possa representar um risco de discriminação, a *datificação*, consistente no armazenamento dos dados permite maior transparência em relação à verificação da existência ou não da discriminação. Neste sentido, Barrocas e Selbst (2016, p. 718) afirmam que: “Any attempt to correct for collection bias immediately confronts the problem of whether or not the employer recognizes the specific type of bias that is producing disparate results. Then, in order to correct for it, an employer must have access to the underlying data and often an ability to collect more. Where more data is clearly not accessible, data miners can proactively compensate for some of the bias by oversampling underrepresented communities.”

Ainda sobre o tema, Anupam Chandler, em palestra dada no Seminário Internacional sobre IA, organizado pela comissão de juristas instituída pelo Senado Federal para elaborar anteprojeto do Marco Regulatório da IA no Brasil, defendeu que para que se possa comprovar efetivamente a existência de discriminação, é necessário que se tenha os dados em relação aos grupos cuja discriminação se busca verificar, como, por exemplo, raça e gênero.

que esses algoritmos analisam o perfil, mas não a capacidade de trabalho das pessoas, estes algoritmos “estão mais suscetíveis a incorporar julgamentos implicitamente enviesados⁵⁶” (KIM, 2020, p. 901-902).

Além disso, em relação ao argumento de que há personalização e seleção de anúncios, a autora aponta que as empresas selecionam quem vai ter acesso a que conteúdo como uma forma de maximizar seus interesses, como taxa de clique e aumento de receitas. Em decorrência disso, aponta-se a necessidade de desenvolvimento de mecanismos para reduzir discriminação. Um exemplo desses mecanismos é o da plataforma LinkedIn, que indica mulheres para as fases seguintes na proporção em que elas aplicaram ao emprego, e não somente o que os algoritmos preditivos entendem por mais relevante, como forma de tentar mitigar os riscos de eventual discriminação que possa ocorrer (KIM, 2020, p. 904).

Outra questão relevante é saber, antes de determinar se candidatos estão sendo discriminados em seleções de emprego, quem é candidato. Essa questão é mais complexa no caso dos algoritmos de recrutamento passivo, em que não é aberta uma concorrência, mas as empresas vão à busca de potenciais candidatos, de modo que a informação da vaga só ficará disponível aos selecionados. Outro caso é o do candidato que nem tenta participar da seleção de emprego, por saber que não será selecionado, caso em que a Suprema Corte dos Estados Unidos da América, em *International Brotherhood of Teamsters v. United States* reconheceu o potencial discriminatório desta situação (KIM, 2020, p. 907-910).

No caso, a Suprema Corte estadunidense reconheceu que políticas discriminatórias “certamente podem impedir aplicantes a empregos que estão cientes disso [políticas discriminatórias] e não desejam se submeter à humilhação da rejeição explícita e certa”⁵⁷ (UNITED STATES, 1977, n.p.). Ou seja, mesmo quem não ficou sabendo da existência da vaga de emprego ou não tentou acessar por já saber que seria excluído da seleção pode ser vítima de discriminação, pode ser equiparado a um candidato para esse fim.

A autora segue, apontando que o título VII do *Civil Rights Act* estadunidense se aplica a determinados atores, dentre eles, as agências de emprego. Entre as proibições estabelecidas está justamente a seleção discriminatória. No entanto, para que as plataformas digitais estejam no escopo regulatório desta legislação, é necessário que estas sejam reconhecidas enquanto agências de emprego (KIM, 2020, p. 911-914).

⁵⁶ Tradução nossa. No original: “they are more likely to incorporate implicit biased judgements”.

⁵⁷ Tradução nossa. No original: “can surely deter job applications from those who are aware of it and are unwilling to subject themselves to the humiliation of explicit and certain rejection”.

Neste contexto, há diferenciação entre as agências de emprego – que se engajam ativamente no recrutamento e na seleção – em face dos meros intermediários – que somente postam passivamente oportunidades de emprego, sem envolvimento nas demais etapas da seleção dos candidatos. Para a autora, a maior parte das plataformas de emprego atuais atuam como agências de emprego, uma vez que vão muito além de anunciá-las⁵⁸. Mesmo no caso do Facebook, que somente faz anúncios de emprego, eles limitam o acesso ao emprego por meio de suas predições, apresentando determinados anúncios especificamente a determinados grupos, o que também pode caracterizar a empresa como uma agência de emprego (KIM, 2020, P. 914-917).

Outro debate existente é sobre a liberdade de expressão, sob a tutela da seção 230 do *Communications Decency Act* estadunidense, que protege as editoras pela publicação de conteúdo de terceiros. No entanto, a conclusão da autora, amparada em decisões da justiça federal estadunidense⁵⁹, é no sentido de que – como as plataformas digitais não funcionam como meros veículos de publicação, mas selecionam que conteúdo deve chegar a que pessoas – esta atividade de seleção feita pelas empresas não se enquadra na liberdade de expressão em face de conteúdo de terceiros, mas como conteúdo próprio, diferentemente do que ocorre nos casos de mera reprodução de informação ou de anúncios (KIM, 2020, p. 918-929).

No entanto, a autora apresenta dificuldades para aplicação dessas normas. A primeira delas é que as leis antidiscriminação tinham como objeto uma vítima individualizada, diferentemente do que ocorre com a discriminação algorítmica, em que uma coletividade é afetada, de modo que deve haver uma abordagem regulatória, muitas vezes preventiva, em face da abordagem tradicional de busca de responsabilidade em cada caso concreto, como é o caso da GDPR europeia, que determina que as pessoas envolvidas tenham informação sobre esses sistemas e sejam informados sobre a lógica de sua operação (KIM, 2020, p. 929-934).

Dessa forma, após a análise dos textos, é possível perceber os riscos discriminatórios decorrentes da aplicação de IA para seleção e recrutamento, e, além disso, a necessidade de mecanismos de transparência, controle e limitação dessas aplicações algorítmicas para mitigar a discriminação e permitir o controle das práticas das empresas de tecnologia que passaram a intermediar a contratação de trabalhadores.

Outra questão que deve ser pensada posteriormente no estudo dos casos é que os algoritmos de seleção e recrutamento não operam por meio de um único ato, mas de um processo de várias

⁵⁸ Ainda que este seja um debate estadunidense focado no regramento estadunidense, ele é interessante por esclarecer o papel ativo das plataformas digitais no mercado de trabalho, excluindo pessoas das seleções de emprego.

⁵⁹ Trata-se de decisões como do caso *HomeAway.com v. City of Santa Monica* (2019) e *Fair Hous. Council v. Roommates.com* (2008), ambos julgados pela United States Courts for the Ninth Circuit.

pequenas decisões encadeadas e que se repete ao longo de tempo, devendo essas aplicações de IA ser compreendidas como algo contínuo e cuja revisão deve acontecer a todo momento, e não como um ato isolado que se encerra após a operação de uma dessas etapas.

2.2. RESPONSABILIDADE PENAL POR INTELIGÊNCIA ARTIFICIAL: estado da arte

Assim, uma vez que diversas formas de discriminação e, em um cenário mais amplo, diversos outros fatos penalmente relevantes podem decorrer do uso de IA, entendemos por necessário compreender o cenário vigente de responsabilidade penal com o emprego destas tecnologias. Para a análise do estado da arte em relação à responsabilidade penal por IA, é necessário que se estabeleça, inicialmente, o critério adotado para a seleção dos textos. Em seguida, a análise será dividida em dois momentos: antes, os textos encontrados em bases de dados nacionais (2.2.1) e, após, a produção encontrada em bases de dados internacionais (2.2.2).

Conforme estabelecido em momento anterior (2.1), as pesquisas no cenário brasileiro serão feitas pelo mecanismo de busca da Biblioteca do Senado, com recorte temporal a partir de 2016 e recorte final até o dia 01 de agosto de 2022. Por este mecanismo, foi feito o uso dos termos de busca “direito penal inteligência artificial”. Com base nesses parâmetros de pesquisa, foram encontrados 5 artigos de periódicos científicos relacionados aos temas tratados neste trabalho.

Em relação aos textos de bases de dados estrangeiras, foi usada a JSTOR para sua seleção, mantido o mesmo recorte temporal dos textos brasileiros. Para os mecanismos de busca, foi feita a seguinte combinação de elementos de pesquisa: “‘*artificial intelligence*’ *criminal liability*”. Em relação ao recorte temático, somente artigos de periódicos científicos relacionados às seguintes áreas do conhecimento foram selecionados: *Law* e *Criminology & Criminal Justice*. Com base nos critérios estabelecidos, foram encontrados 6 artigos científicos relacionados aos temas deste trabalho.

2.2.1. Estado da arte no Brasil

No recorte dos textos de bases de dados nacionais, o primeiro texto analisado é o de Barbosa (2020), que trata de crimes praticados por empresas de tecnologia e como a IA pode ajudar a prever fraudes, mas não especificamente de responsabilidade penal por meio de Inteligência Artificial.

No segundo texto, Divino (2020) busca compreender como se dá a responsabilidade penal em face de atos praticados por IA. Inicialmente, adotou-se a definição de IA fraca, definida como um programa de computador, uma aplicação, para um determinado fim específico. Segundo defende o

autor, IA não é inteligente no sentido da inteligência humana, podendo tão somente simular o pensamento e a inteligência humana (DIVINO, 2020, p. 154-162).

O autor, amparado nas ideias de Gabriel Hallevy, passa a descrever modelos de responsabilidade penal para IA. Inicialmente, aponta os elementos que constituem o ato criminoso: (i) *actus reus* e (ii) *mens rea*. Esses conceitos designam, respectivamente, (i) o ato em si e (ii) o elemento mental, “designado pelo conhecimento ou intenção *vis-à-vis* (em face) criminal” (DIVINO, 2020, p. 162). O *mens rea* pode variar em distintos níveis, desde os mais baixos - “que entendemos como *culpa sticto sensu*” - ao mais avançado, que “é o *conhecimento*” (DIVINO, 2020, p. 163).

Para compreender a responsabilidade penal decorrente de IA, o autor apresenta três modelos propostos por Hallevy. O primeiro desses modelos é o chamado *the perpetration-by-another liability model*, segundo o qual as capacidades de IA “não geram indícios de *actus reus* e *mens rea*”, de modo que IA não pode ser autora de crimes (DIVINO, 2020, p. 164). Neste modelo, fatos típicos praticados por IA seriam de responsabilidade das pessoas envolvidas em sua produção e/ou operação, quando for o caso (DIVINO, 2020, p. 165).

O segundo modelo é o *the natural probable-consequence liability model*, segundo o qual “uma pessoa pode ser responsabilizada criminalmente se o ilícito cometido é uma consequência natural e provável da conduta inicialmente praticada” (DIVINO, 2020, p. 166). Neste modelo, portanto, não é necessário que haja o conhecimento dos atos praticados pela IA, mas os operadores de algoritmos estariam “obrigados a saber que esse delito é uma consequência natural e provável” de seus atos (DIVINO, 2020, p. 166). Neste modelo, tanto a IA quanto as pessoas envolvidas em sua criação e operação são puníveis (DIVINO, 2020, p. 167).

O terceiro modelo apresentado é o *the direct liability model*, que separa a IA de seus desenvolvedores e usuários, não sendo impossível que os elementos de imputação penal “possam ser preenchidos por uma IA” (DIVINO, 2020, p. 167). Neste sentido, através de uma concepção de conhecimento hallevyana de “recepção e compreensão sensorial de dados factuais” (DIVINO, 2020, p. 167), tecnologias como *machine learning* e *deep learning* podem ser consideradas formas de conhecimento, possibilitando até mesmo, de acordo com Hallevy, “uma IA defender sua existência perante as excludentes de ilicitude, como legítima defesa” (DIVINO, 2020, p. 169). Por outro lado, autores como Abbott e Sarch defendem a impossibilidade de “criminalização das condutas dos entes inteligentes artificialmente em virtude da ausência de substrato científico para comprovar seus benefícios e malefícios”, defendendo a aplicação de responsabilidade civil e de afastamento de responsabilidade penal nesses casos (DIVINO, 2020, p. 169).

Neste contexto, o autor passa a apresentar dois conceitos para o que seria uma mente, para desenvolver o debate sobre a possibilidade de responsabilização penal de IA. Essa oposição se dá

entre o funcionalismo computacional – que define estados mentais como estados funcionais, ou seja, “aquilo que produz ou causa o comportamento de um organismo em condições específicas” (DIVINO, 2020, p. 172) – e o naturalismo biológico – que defende que a mente humana possui características únicas, como subjetividade e estados de humor, de modo que IA não possui estados mentais, podendo ter conhecimento, mas não inteligência, sendo incapaz de expressar *mens rea*, o que impede o estabelecimento de sua responsabilidade penal (DIVINO, 2020, p. 178).

Neste contexto, o autor adota o modelo *the perpetration-by-another*, de modo que “a responsabilidade deverá recair sob aqueles que o programaram, desenvolveram ou utilizaram para o cometimento de atos ilícitos”, pois “a falsa inteligência acompanhada por uma falsa mente não pode servir como instrumento a obstaculizar a efetiva aplicação da tutela penal àquele que realmente externou o *mens rea* ou que faltou com *culpa stricto sensu* no dever de vigilância desses entes” (DIVINO, 2020, p. 178)”.

No terceiro texto incluído na revisão de literatura, Varella (2021) afirma que, na era digital, há cada vez mais frequente emprego de tecnologia, gerando alterações tanto nas formas de IA quanto nas pessoas. Desta forma, em vez de um cenário em que somente pessoas têm autonomia e máquinas são heterônomas, passa-se a conviver com a figura de máquinas autônomas e de pessoas autômatos, cada vez menos autônomas em face da delegação de tarefas às máquinas. Neste contexto, surgem dificuldades de se compreender como se desenvolve a responsabilidade penal (VARELA, 2021, *passim*).

Em decorrência desse fenômeno, os riscos sociais são incrementados de forma imperceptível. No entanto, ante a ausência de um cenário regulatório claro em relação a esses riscos provenientes de máquinas autônomas, resta recorrer às estruturas clássicas do direito penal, como o risco socialmente tolerado para riscos de lesividade insignificante e a modalidade de imprudência para casos de aumento do risco automatizado⁶⁰ (VARELA, 2021, p. 28).

Neste sentido, a autora apresenta fundamentos de afastamento da responsabilidade penal individual, apontando a redução de liberdade das pessoas em face da tecnologia como forma de afastar o controle e a responsabilidade humana. Além disso, apresenta-se um cenário dúplice, em que a ausência de regulação pode ser entendida como permissão para as aplicações algorítmicas, mas também como ausência de permissão, demandando prudência de operadores e usuário. Por outro lado, aponta-se a necessidade de criação de deveres de cuidado por parte das pessoas e a perda de controle

⁶⁰ Tradução nossa. No original: “como el riesgo socialmente tolerado para riesgos de lesividad insignificante y la modalidad de la imprudencia para casos de incremento del riesgo automatizado”.

dos riscos sociais como fundamento para defender a responsabilidade penal individual (VARELA, 2021, p. 22-27).

O texto, mais do que defender uma determinada postura em relação à responsabilização penal dos seres humanos em face de atos “praticados” por IA, aponta as dificuldades do tratamento deste tema, seja ao se decidir pela incidência do direito penal, seja por sua ausência.

No quarto texto, Chaves Junior e Berzagui (2021) se propõem, a partir da teoria da ação significativa, a analisar se formas de IA devem possuir responsabilidade penal própria ou se, por outro lado, “quem cria, maneja e/ou coloca em atividade esses sistemas autônomos de inteligência artificial” (CHAVES JUNIOR; BERZAGUI, 2021, p. 1132) deve ser responsabilizado. Neste sentido, a aplicação da Teoria da Ação Significativa limita o conceito de ação, de modo que só há caracterização da responsabilidade penal “quando caracterizada uma ação cujo sentido é socialmente indesejado, de forma que meros acontecimentos não merecem atenção do Direito Penal” (CHAVES JUNIOR; BERZAGUI, 2021, p. 1144).

Dessa forma, os autores apontam que somente quando há intenção pode haver responsabilidade penal, afastando-se a responsabilidade penal de IA, mesmo das formas autônomas e capazes de autoaprendizado, pois estas ferramentas tecnológicas não seriam capazes de agir com intenção. Neste contexto,

[...] com base na Teoria da Ação Significativa, o que poderia ser considerado para fins de responsabilidade penal é a utilização da inteligência artificial como instrumento para a realização de um resultado típico. Nesse caso, o sistema não seria diferente de uma arma utilizada para o cometimento de um delito, já que consistiria no meio para a realização de uma intenção objetiva do ser humano que a maneja. A inteligência artificial, então, seria o meio escolhido pelo indivíduo para a realização de um comportamento dotado de significado, o que caracterizaria uma ação. (CHAVES JUNIOR; BERZAGUI, 2021, p. 1158)

Em decorrência disso, caso haja fato potencialmente criminoso praticado por meio de IA, “seria preciso indagar se esse fato pode ser interpretado, no âmbito da causalidade, como manifestação objetiva da vontade do ser humano que a controla” (CHAVES JUNIOR; BERZAGUI, 2021, p. 1159). Sobre a responsabilidade penal das formas de IA, tidas enquanto sujeitos ativos de crimes, concluem que “a pena para robôs será somente pensável quando o direito Penal perder o interesse humano” (CHAVES JUNIOR; BERZAGUI, 2021, p. 1160).

No último dos textos analisados, Farnosier (2022, P. 156) debate os conceitos de personalidade e responsabilidade no direito penal em face da ampliação da IA, com um enfoque no debate dogmático sobre o tema, não tratando das especificidades do sistema penal brasileiro. Segundo o autor, a despeito das evoluções tecnológicas relativas ao desenvolvimento da Inteligência Artificial, “os criadores desses sistemas ainda podem ser considerados responsáveis pelo uso criminoso [de IA]

na maior parte dos sistemas jurídicos pelo mundo⁶¹” (FARNOSIER, 2022, p. 158). Ainda, diante das possibilidades de desenvolvimento tecnológico, o autor aponta que em alguns sistemas a biotecnologia de clonagem de seres vivos foi proibida ou limitada, justamente para que se evitasse os riscos de seu desenvolvimento (FARNOSIER, 2022, p. 158-159).

Neste sentido, o autor afirma que alguns conceitos do direito penal deveriam ser revisados para que o uso de IA seja regulado neste campo do direito: (i) o conceito de culpabilidade vigente; (ii) o controle, pelos governos, de determinadas aplicações de IA, especialmente quando colocam em risco a segurança das pessoas; (iii) o não reconhecimento da responsabilidade penal das formas de IA, ante as dificuldades do reconhecimento de sua personalidade jurídica (FARNOSIER, 2022, p. 159). No entanto, segue o autor: “em situações intencionais ou culpáveis nas quais humanos e IA estão envolvidos conjuntamente na prática de crimes, a punição do humano, e não da entidade em si, deveria ser considerada”⁶² (FARNOSIER, 2022, p. 160).

Se, por um lado, o autor afirma que debater um modelo em que IA possui personalidade é algo difícil, apontando debater a responsabilizações das organizações, em um modelo de responsabilidade penal da pessoa jurídica, pode ser uma solução, dada a dificuldade de identificação da responsabilidade individual de cada agente envolvido no desenvolvimento e na operação dos algoritmos (FARNOSIER, 2022, p. 160). Dessa forma, ainda que seja possível conceber determinadas formas de IA que possuam crenças, desejos e intenções, uma solução é que a legislação limite esse nível de autonomia dos sistemas (FARNOSIER, 2022, p. 165).

O autor aponta três caminhos possíveis para essa relação entre robótica e direito penal: (i) insistir nos modelos tradicionais de responsabilidade penal, o que coloca no centro da responsabilidade penal os desenvolvedores e operadores de IA; (ii) aplicar um conceito funcional de responsabilidade penal, focado na adequação social às normas, de modo que “a responsabilidade penal não é baseada no reconhecimento do livre arbítrio ou liberdade de escolha humanos, mas em uma construção de que”, se robôs possuem cada vez mais funções humanas, “eles podem ser experienciados como ‘iguais’” (FARNOSIER, 2022, p. 166); (iii) superar o modelo vigente baseado em livre arbítrio e liberdade de escolha para uma lógica de “estabilização das normas por meio da punição” (FARNOSIER, 2022, p. 166).

Dessa forma, ante as dificuldades de aplicação das regras vigentes em relação a IA, deve-se fazer uma escolha entre qual o sistema de responsabilização aplicável – e em quais bases de direito

⁶¹ Tradução nossa. No original: “the creators of systems can still be held responsible for criminal use in the majority of law systems around the world”.

⁶² Tradução nossa. No original: “in intentional or culpable situations in which humans and AI are jointly involved in the occurrence of crimes, the punishment of the human, and not the entity itself, should be considered.”

penal – para que seja feita uma adequação às necessidades decorrentes do estado atual de desenvolvimento tecnológico experimentado.

Com base nos textos analisados, é possível perceber a dificuldade no estabelecimento de responsabilidade penal decorrente de IA, sendo a posição predominante dos autores no sentido de que máquinas não possuem intenção (*mens rea*) ou agir significativo, de modo que a responsabilização penal das máquinas é um caminho difícil de se tomar. Por outro lado, a responsabilização das pessoas envolvidas na criação ou operação de IA é um caminho possível, mas difícil, ante a ausência, muitas das vezes, de atuação humana direta na atuação e na “tomada de decisão”. Outra possibilidade é a criação de um modelo penal alternativo, que se dissocie dos conceitos de ação, culpa, responsabilidade e passe a aplicar punições tão somente para a estabilização das normas.

2.2.2. Estado da arte no mundo

No primeiro texto do recorte feito por meio de busca em bases de dados internacionais, Brodsky (2016) analisa o cenário regulatório em relação aos carros autônomos. Em linhas gerais, o texto aponta que a legislação e regulação vigentes não são suficientes para esta nova tecnologia, pois, seja pela insuficiência da restrição vigente, seja pelo risco de que a regulamentação inadequada freie a inovação no setor de tecnologia. Sobre o tema, o autor aponta que regulação deve estar associada ao contexto de inteligência artificial e de desenvolvimento de novas tecnologias, para além da aplicação de regulamentação das normas de trânsito já existentes, pensadas para um contexto humano (BRODSKY, 2016, *passim.*).

No segundo texto, Crootof (2016) busca entender as possibilidades de aplicação do direito penal internacional em face de crimes de guerra cometidos por sistemas de armamento autônomos utilizados em contexto de guerra – que a autora chama de robôs assassinos. Tais formas de armamento são definidas como sistemas capazes de selecionar e abordar os alvos sem a necessidade de tomada de decisão humana (CROOTOF, 2016, p. 1367). No entanto, como não há pessoas responsáveis pela definição dos alvos e da tomada de decisão, nem como estabelecer responsabilidade penal individual por ação intencional ou negligente (CROOTOF, 2016, p. 1347).

Em face dessa dificuldade, a autora aponta a existência de um vazio de responsabilidade (*accountability gap*) em relação a crimes de guerra cometidos pelos robôs assassinos (CROOTOF, 2016, p. 1353). Essa dificuldade de responsabilizar indivíduos é ampliada pela imprevisibilidade existente em sistemas com tomada de decisão autônoma e autoaprendizagem (CROOTOF, 2016, p. 1374).

Em decorrência disso, a autora defende que – nesses casos – não há responsabilidade penal direta ou indireta de humanos, uma vez que a tomada de decisão e a ação ocorrem independentemente de ação ou ordem intencionais de um ser humano (CROOTOF, 2016, p. 1376-1381). Em relação à possibilidade de punição por negligência, afirma que esta forma não possui previsão expressa e que, caso passe a ser ampliada, todos os atos de guerra poderiam ser criminalizados, acabando com o direito penal internacional referente aos conflitos entre estados (CROOTOF, 2016, p. 1381-1385).

Assim, em face das dificuldades de responsabilização dos atos praticados por robôs assassinos, a autora defende a responsabilização dos Estados. Dessa forma, a proposta é de afastamento do direito penal e criação de um sistema de responsabilidade civil (*tort liability regime*) como solução para superar o vazio de responsabilidade em relação às armas autônomas no contexto do direito internacional (CROOTOF, 2016, p. 1402).

No terceiro texto, Gless, Silverman e Weigend (2016) buscam compreender a responsabilidade decorrente de danos a bens jurídicos penalmente relevantes causados por carros autônomos, apontando, inicialmente, a dificuldade de se estabelecer responsabilidade, uma vez que a pilotagem é feita por IA, e não exatamente por pessoas. Neste sentido, questiona-se como pode haver a punição e quem são os agentes que podem ser punidos. (GLESS; SILVERMAN; WEIGEND, 2016, p. 412-414)⁶³.

Para desenvolver seus argumentos, o texto faz referência a dois modelos penais distintos. De um lado estão países como os Estados Unidos, caracterizados por maior pragmatismo no sistema penal, em que atos de não humanos podem ser integrados ao sistema penal. Neste sentido, aplica-se doutrinas como a da *respondeat superior*, em que é possível “a imposição de responsabilidade penal de empresas por atos praticados por seus agentes em benefício das próprias empresas, dentro do escopo de atribuição do agente⁶⁴” (GLESS; SILVERMAN; WEIGEND, 2016, p. 417).

De outro lado, aponta-se o caso de países como a Alemanha, que – pelas bases idealistas de seu sistema penal – resiste à ideia de responsabilidade penal de não humanos, como é o caso das pessoas jurídicas. No sistema alemão, a responsabilidade penal exige vontade autônoma de um agente, com ações orientadas ao contexto criminoso. Além disso, no sistema alemão, existe a necessidade de que o agente tenha culpabilidade, o que, segundo os autores, pressupõe que haja discernimento entre certo e errado, sendo o agente capaz de evitar o cometimento do ato ilícito – o

⁶³ Neste texto, e em alguns textos que debatem especificamente a responsabilidade penal por veículos autônomos, há, além dos debates apresentados, questionamentos em relação à responsabilidade penal pelo produto. No entanto, como a seleção de empregos por meio de ferramentas tecnológicas é um serviço, e não um produto, este debate não aproveita à pesquisa proposta, motivo pelo qual não será objeto deste trabalho.

⁶⁴ Tradução nossa. No original: “the imposition of criminal liability on corporations for acts committed by their agents on behalf of the corporation, for the benefit of the corporation, and within the scope of the agent’s authority”.

que, segundo os autores, não existe no caso de robôs⁶⁵ (GLESS; SILVERMAN; WEIGEND, 2016, p. 415-422).

Outra possibilidade debatida no texto é a de atribuição de responsabilidade penal diretamente aos robôs. No entanto, os autores apontam ser um caminho difícil, considerando que robôs não podem ser punidos dentro do direito penal vigente, além de não ter intenção e culpabilidade. Dessa forma, entenderam pela inadequação da reponsabilidade penal de robôs (GLESS; SILVERMAN; WEIGEND, 2016, p. 422-424).

Superada essa possibilidade, o texto passa a tratar da responsabilidade dos humanos por trás da máquina. Nos casos em que os danos causados são intencionalmente programados ou conhecidos, os autores defendem ser viável a aplicação das regras de imputações tradicionais do direito penal. No entanto, nos casos de negligência, os conceitos tradicionais de responsabilidade penal não são adequados à responsabilidade por danos causados por carros autônomos (GLESS; SILVERMAN; WEIGEND, 2016, p. 424-425).

Para o modelo alemão, caso os danos sejam antevistos, o agente deve reprimir a ação que gerará a violação à norma penal por omissão. No entanto, como apontam os autores, há situações delicadas no contexto da responsabilidade penal omissiva de pessoas que trabalham com IA, em que “um ato lícito – colocar em mercado um carro autônomo de acordos com o estado da arte das normas e do conhecimento tecnológico – pode gerar responsabilidade penal por omissão”⁶⁶ (GLESS; SILVERMAN; WEIGEND, 2016, p. 428).

Os autores apontam as dificuldades de responsabilização dos programadores, uma vez que a própria ação dos robôs pode interromper onexo causal entre a ação do programador e o resultado. Dessa forma, considerando-se a dificuldade de estabelecer responsabilidade da pessoa por trás da máquina e a impossibilidade de responsabilização penal de robôs, haveria um vazio de responsabilidade, o que “aconselha contra um afastamento da responsabilidade de operadores de robôs por danos causados pelos robôs”⁶⁷ (GLESS; SILVERMAN; WEIGEND, 2016, p. 432).

Segundo os autores, deve ser feita a escolha de quais riscos entendemos como aceitáveis ou inaceitáveis, sendo que tecnologias sempre trazem riscos, como é o caso de carros pilotados por

⁶⁵ Os textos usam indistintamente robôs como forma de se referir a formas de IA. Neste trabalho, também haverá este tratamento indistinto, ainda que possa existir esta distinção, uma vez que esta diferenciação não é necessária para o desenvolvimento deste trabalho.

⁶⁶ Tradução nossa. No original: “a perfectly legal act-the marketing of a self-driving car in accordance with the current state of knowledge and technology-may trigger criminal liability for omission”.

⁶⁷ Tradução nossa. No original: “counsel against generally absolving the robot's operators of responsibility for harm caused by the robot”.

humanos, que acarretam acidentes e mortes. A questão é, portanto, determinar se os riscos produzidos pelos robôs são normais ou se extrapolam o aceitável. Dessa forma:

Mesmo quando é previsível que os robôs possam cometer erros fatais resultantes de uma autoaprendizagem equivocada, o dever dos operadores de (pelo menos alguns) robôs pode ser reduzido a empregar o melhor conhecimento e tecnologia disponível na fabricação, programação, teste e monitoramento. Os produtores de robôs devem testá-los extensivamente e devem observar de perto, em particular, seus processos de aprendizagem autônomos. Além disso, eles devem monitorar o *feedback* dos clientes e devem reagir imediatamente às denúncias de conduta prejudicial.⁶⁸ (GLESS; SILVERMAN; WEIGEND, 2016, p. 434)

Assim, considerando-se os riscos de funcionamento dos carros autônomos como aceitáveis, um robô causar danos por desvios de seus procedimentos de aprendizagem deve ser considerado uma responsabilidade da sociedade de forma geral, e não de seus operadores.

Sobre as regras de funcionamento dos veículos autônomos, no direito alemão, sistemas autônomos de condução de veículos só podem ser empregados caso haja um ser humano capaz de assumir a direção, caso necessário. Em relação aos Estados Unidos, aponta-se a necessidade de que um condutor habilitado assuma a condução do veículo caso haja problemas técnicos. (GLESS; SILVERMAN; WEIGEND, 2016, p. 433-436)

Neste contexto, a conclusão a que o texto chega é que formas de IA como carros autônomos forçam os limites da responsabilidade penal, uma vez que robôs não podem ser punidos criminalmente por ausência de culpabilidade em suas ações e por não poderem ser sujeitos às punições penais tradicionalmente estabelecidas. Acerca da punição dos operadores, aponta-se que quanto maiores as possibilidades de sua responsabilização, menores os incentivos ao desenvolvimento tecnológico, de modo que

[...] cada sociedade deve responder, por si mesma, à questão de saber se investimento em chances de uma vida melhor [por meio da tecnologia] deve ser recompensado com uma isenção de responsabilidade criminal por alguns dos riscos envolvidos - e quais são esses riscos⁶⁹. (GLESS; SILVERMAN; WEIGEND, 2016, p. 436)

Assim, o texto apresenta as principais dificuldades e riscos na imputação de responsabilidade penal a pessoas envolvidas no desenvolvimento e operação de algoritmos.

No quarto texto, Yadav (2016) aponta a dificuldade de responsabilização por atos de inteligência artificial no mercado de ações, com foco central em responsabilidade civil. As conclusões

⁶⁸ Tradução nossa. No original: “Even where it is foreseeable that robots may commit fatal mistakes resulting from misguided self-learning, the duty of operators of (at least some) robots could be reduced to employing the best knowledge and technology available in manufacturing, programming, testing, and monitoring them. Producers of robots must test them extensively and must closely observe, in particular, their autonomous learning processes. They must, moreover, monitor feedback from customers and must react immediately to reports of harmful conduct.”

⁶⁹ Tradução nossa. No original: “each society must answer for itself the question of whether investment in chances for a better life should be rewarded with an exemption from criminal responsibility for some of the risks involved-and what these risks are.”

são que os panoramas de responsabilidade vigentes são insuficientes para o contexto da IA, de modo que deve surgir um novo paradigma regulatório para os algoritmos de investimentos e, além disso, deve haver uma abordagem que complete as lacunas deixadas por legislações ineficientes para cobrir falhas nessa aplicação algorítmica (YADAV, 2016, *passim*).

O quinto texto objeto da revisão de literatura é o de Tutt (2017), que aponta a crescente complexidade das aplicações de algoritmos e o uso de métodos de aprendizagem de máquina como uma tendência, havendo, nas formas de IA, cada vez mais elementos não explicitamente programados. Desse modo, muitas das vezes, nem mesmo o desenvolvedor é responsável pelo caminho de operação dos algoritmos. Em decorrência disso, as formas de regulação existentes, como responsabilidade penal ou civil, são insuficientes diante da existência de conhecimentos sobre as novas formas de tecnologia. Por isso, o autor apresenta o caminho regulatório, por meio da criação de uma agência regulatória federal nos Estados Unidos como solução para essa dificuldade de responsabilização no país (TUTT, 2017, *passim*).

Ainda que alguns detalhes do artigo não interessem diretamente à análise realizada, o argumento central do artigo nos interessa. Neste contexto, a proposta é que a agência faça um controle prévio do lançamento de determinados algoritmos no mercado, até que estes sejam testados e tenham sua segurança comprovada. Por isso, a proposta é a criação de uma agência nos parâmetros da FDA⁷⁰, que realiza controle de segurança de medicamentos antes de seu lançamento, garantindo não somente a eficácia das drogas, mas também sua segurança. Dessa forma, o autor enfrenta as limitações à responsabilização decorrente da aplicação de IA, sugerindo o modelo regulatório como adequado, incluindo-se a previsão de controle prévio de algoritmos. (TUTT, 2017, p. 118-123)

No último texto selecionado, Warren e Hillas (2018) tratam de armamentos de guerra, apontando a tendência de exclusão dos humanos nas armas autônomas. Neste sentido, aponta-se a necessidade de supervisão humana, incluindo-se, entre os motivos para isso, o possível cometimento de crime com o emprego dessa forma de armamento e eventual cenário de irresponsabilidade decorrente da ausência de decisão ou supervisão humana (WARREN; HILLAS, 2018, *passim*).

Com base nos textos analisados, é possível compreender a insuficiência das formas de responsabilidade vigentes, criando verdadeiros vazios de responsabilidade penal e civil. No entanto, existe uma tendência doutrinária da defesa da criação de um modelo regulatório em face das novas tecnologias, e não da ampliação do direito penal. Contudo, nos casos em que houver responsabilidade penal, a posição majoritária é no sentido de afastar a responsabilização das máquinas e buscar

⁷⁰ Trata-se da agência reguladora estadunidense que faz trabalho semelhante ao da ANVISA no cenário brasileiro, controlando o lançamento de remédios no mercado, por exemplo.

responsabilizar as pessoas envolvidas em sua operação, como desenvolvedores e operadores. Especificamente em relação à aplicação do direito penal, no entanto, as regras vigentes não parecem suficientes para um cenário tecnológico cada vez mais complexo e com cada vez mais influência sobre bens jurídicos penalmente relevantes.

Assim, além do que apresentam os textos, deve-se ter em mente, como também tratado por Gless, Silverman e Weigend, qual escolha política deve ser feita para endereçar, limitar ou consentir com os riscos decorrentes da terceirização, para máquinas, de decisões eminentemente humanas. Neste sentido, há autores que não tomam a aceitação da aplicação de IA em suas máximas potencialidades em contraste com a redução do papel humano na tomada de decisões.

Neste contexto, por ser útil rememorar os princípios do constitucionalismo digital, especialmente os relacionados à necessidade de supervisão humana, no sentido de impedir que máquinas efetivamente tomem as ‘decisões’, afastando-se o controle e a responsabilidade dos humanos envolvidos nos processos. Além disso, os princípios relacionados a democracia e explicabilidade também são centrais para este trabalho.

Consoante Rodrigues, (2020, p. 45), deve-se pensar qual é o projeto humano para a era digital. Assim, uma questão que se coloca é se “ao entregar a decisão à máquina, o homem não lhe transfere, desta forma, a responsabilidade da decisão que lhe deve caber a ele, enquanto agente moral, responsável pelas suas decisões em sociedade” (RODRIGUES, 2020, p. 47). A autora afirma – sobre essa entrega de decisões morais a máquinas – que deve haver meios para evitar essa transferência de responsabilidade para fora da esfera de controle humano, sendo que “a nossa escolha mais ética, ainda aberta à discussão, é escolher que não o queremos fazer” (RODRIGUES, 2020, p. 48).

Não pretendemos fazer a defesa da impossibilidade completa de determinadas aplicações de IA, em face dos riscos que podem ser causados às pessoas. No entanto, deve haver clareza de que esta é uma possibilidade que deve ser levada ao debate e, qualquer que seja a solução dada, que o ser humano deve ser pensado como o centro deste debate, e não somente uma peça neste tabuleiro.

Além disso, nenhum dos textos analisados tratou especificamente sobre como as regras do direito penal brasileiro são afetadas pela IA, mas tão somente sobre regras genéricas ao longo do mundo ou regras de responsabilidade penal de outros sistemas jurídicos, como o alemão e o estadunidense. Assim, estabelecidas essas premissas, e com base na revisão de literatura realizada, busca-se, a seguir, compreender responsabilidade penal por IA segundo as normas e estruturas de imputação vigentes no Brasil.

2.3. MODELOS DE RESPONSABILIDADE PENAL NO DIREITO BRASILEIRO E IA

Como apresentado anteriormente (2.2), ainda que em muitos casos algoritmos pratiquem fatos definidos como crime, ou ainda, pessoas pratiquem crimes por meio de IA, existem dificuldades e fatores limitantes para que haja a aplicação do direito penal. Neste contexto, buscaremos apresentar, de acordo com as regras vigentes no direito brasileiro, as possibilidades de responsabilidade penal, esclarecendo quem pode ser sujeito ativo e quais modelos de ação e autoria permitem essa compreensão.

Dessa forma, inicialmente será apresentado o panorama da responsabilidade penal da pessoa jurídica no direito brasileiro (2.3.1), e como isso cria uma limitação concreta à responsabilização penal direta de algoritmos, tidos enquanto não entes ou entes não naturais. Relativamente às pessoas naturais envolvidas no desenvolvimento ou operação de IA, é necessário a compreender o conceito de ação aplicável (2.3.2), as regras de autoria (2.3.3), as regras de culpabilidade (2.3.4) e dos limites impostos pelo princípio da legalidade (2.3.5) em nosso direito penal.

2.3.1. Responsabilidade penal da pessoa jurídica no direito penal brasileiro

No direito brasileiro, a regra é a responsabilidade penal de pessoas físicas. No entanto, a Constituição Federal determinou como possível a responsabilidade penal de pessoas jurídicas no caso dos crimes ambientais⁷¹ e contra a ordem econômica⁷². Em hierarquia legal, a Lei 9.605/1998 regulamentou a questão, passando a prever punição para pessoas jurídicas nos crimes ambientais⁷³.

Em relação ao tema, existe amplo debate doutrinário, com autores defendendo a inadequação dogmática da criminalização de condutas das pessoas jurídicas, pelas dificuldades de se estabelecer tipicidade, ilicitude e, especialmente, culpabilidade própria das pessoas jurídicas – sentido para o qual aponta Lobato (2020, *passim.*)⁷⁴. No entanto, a maior parte da doutrina reconhece a legitimidade penal ativa das pessoas jurídicas no caso dos crimes ambientais. Neste sentido, Bello Filho (2015, p. 254) afirma que “não há ofensa a quaisquer princípios constitucionais penais que regem a imposição de

⁷¹ Art. 225, § 3º: “As condutas e atividades consideradas lesivas ao meio ambiente sujeitarão os infratores, pessoas físicas ou jurídicas, a sanções penais e administrativas, independentemente da obrigação de reparar os danos causados.”

⁷² Segundo o §3º do art. 173 da Constituição Federal: “A lei, sem prejuízo da responsabilidade individual dos dirigentes da pessoa jurídica, estabelecerá a responsabilidade desta, sujeitando-a às punições compatíveis com sua natureza, nos atos praticados contra a ordem econômica e financeira e contra a economia popular”.

⁷³ O referido diploma normativo apresenta a seguinte previsão: “Art. 3º As pessoas jurídicas serão responsabilizadas administrativa, civil e penalmente conforme o disposto nesta Lei, nos casos em que a infração seja cometida por decisão de seu representante legal ou contratual, ou de seu órgão colegiado, no interesse ou benefício da sua entidade. Parágrafo único. A responsabilidade das pessoas jurídicas não exclui a das pessoas físicas, autoras, co-autoras ou partícipes do mesmo fato.”

⁷⁴ Em textos mais antigos, a irresponsabilidade penal da pessoa jurídica é defendida por Dotti (1995, *passim.*), que defende serem imputáveis somente os seres humanos, só havendo responsabilidade penal de pessoas físicas. No mesmo sentido, *vide*: (BITENCOURT, 2010).

sanções desta natureza”. Em sentido semelhante, Krebs (2020, *passim*.) defende que a possibilidade de cometimento de crimes por pessoas jurídicas não representa violação aos limites constitucionais de um direito penal mínimo, mas tão somente compatibiliza esses limites com as previsões de defesa do meio ambiente e da ordem econômica e social.

Na jurisprudência, contudo, este debate já foi resolvido, de modo que tanto o Supremo Tribunal Federal quanto o Superior Tribunal de Justiça já reconheceram a possibilidade de criminalização de pessoas jurídicas por crimes ambientais (BELLO FILHO, 2015, p. 236/238)⁷⁵. Assim, a despeito do debate doutrinário, os tribunais têm aplicado a responsabilidade penal ambiental da pessoa jurídica.

Contudo, existe outra questão que emerge da responsabilidade penal das pessoas jurídicas: a da dupla imputação. Ou seja, passou a haver debate doutrinário e jurisprudencial acerca da necessidade de responsabilização penal das pessoas físicas como requisito de atribuição de responsabilidade penal às pessoas jurídicas. O Superior Tribunal de Justiça passou a exigir a necessidade de responsabilização penal das pessoas físicas envolvidas enquanto requisito necessário para a caracterização da responsabilidade penal das pessoas jurídicas⁷⁶ (COSTA, 2015, p. 233/234). O Supremo Tribunal Federal, por sua vez, afastou esse requisito, permitindo a criminalização de pessoas jurídicas sem que houvesse o reconhecimento, concomitante ou anterior, da responsabilidade de pessoas físicas na mesma situação fática⁷⁷, “uma vez que o art. 225, §3.º, não condiciona a responsabilização penal da pessoa jurídica à identificação da pessoa física” (COSTA, 2015, p. 242).

Na doutrina, Ney Bello Filho (2015, p. 253/254) reconhece a desnecessidade de dupla imputação para caracterização de crime ambiental por pessoa jurídica. No mesmo sentido aponta Rafael de Oliveira Costa (2015, p. 244), segundo o qual não se pode interpretar pela existência de vedação não criada pela Constituição Federal ou por lei, como é o caso da exigência da dupla imputação para atribuição de responsabilidade penal a pessoa jurídica. Neste sentido, o entendimento do STF sobre a desnecessidade de dupla imputação passou a ganhar força na doutrina. Por outro lado, autores como Martins (2016, p. 175) defendem existir controvérsias sobre a dupla imputação, apontando dificuldades de atribuição de responsabilidade penal a pessoas jurídicas com as bases vigentes do sistema penal pátrio. Tais dificuldades se verificam até mesmo pela questão probatória, pois é improvável o cometimento de crime por pessoa jurídica sem o envolvimento de qualquer

⁷⁵ No STF, a possibilidade de responsabilização penal das pessoas jurídicas foi reconhecida em decisões como no HC 88747 e RE 593.729. No STJ, tal possibilidade foi referendada em julgados como o REsp 800817 e HC 89297.

⁷⁶ A impossibilidade de reponsabilidade penal da pessoa jurídica sem que haja dupla imputação foi estabelecida nos seguintes julgados do STJ, dentre outros: RHC 24239, REsp 610114.

⁷⁷ Esse entendimento foi instituído no julgamento do RE 548.181/PR.

pessoa física, e é difícil que se comprove um crime contra a pessoa jurídica sem que haja provas implicando pessoas físicas.

Dessa forma, a despeito do amplo debate sobre o tema, o recorte atual do direito brasileiro não permite a responsabilidade penal de pessoas jurídicas por IA, à exceção da responsabilidade nos crimes ambientais, já existente, regulamentada e reconhecida judicialmente. Não se desconhece, portanto, a possibilidade teórica de que empresa de tecnologia venha a cometer eventual crime ambiental, de acordo com as regras vigentes no Brasil. No entanto, debates em relação à responsabilidade penal da pessoa jurídica por algoritmos não serão abordados no presente trabalho, que observa especificamente o crime de racismo, cuja normatização vigente não prevê essa modalidade de responsabilidade.

2.3.2. O direito penal brasileiro, as teorias da ação e o dolo no tipo penal

A seguir, passa-se a apresentar as principais teorias desenvolvidas e relação à ação no direito penal, buscando compreender qual o sistema vigente no Brasil. As teorias abordadas serão da ação causal, da ação final, e da imputação objetiva. Consoante Greco (2014, p. 27), a teoria da ação causal, mais antiga das três, limitava o tipo à mera causalidade, de modo que “o tipo nada mais era que a causação de uma lesão a um bem jurídico”, havendo consideração do desvalor do resultado sem qualquer juízo quanto ao desvalor da ação. Neste contexto, Welzel (2015, p. 39-40) descreve a ação causal como “o mero processo causal que desencadeia a vontade (o “impulso voluntário”, ou a “enervação”) no mundo exterior (= efeito da vontade), independentemente de que o autor o tenha querido ou possa sequer prevê-lo (= conteúdo da vontade)”.

Assim, a ação causal busca estabelecer a relação de causa e efeito entre a ação e a violação do bem jurídico-penal. Ainda que haja grande importância no desvalor do resultado, o desvalor da ação é colocado, portanto, em segundo plano, fora, inclusive, da tipicidade penal, residindo tão somente na culpabilidade ou na antijuridicidade (ROXIN, 2008, p. 101-102).

Alguns problemas apontados nesta teoria da ação são que: (i) a tentativa de crime não é coberta por esta teoria, uma vez que, nos crimes tentados, não há desvalor no resultado, mas somente na ação (WELZEL, 2015, p. 41); (ii) a relação de causalidade desta teoria não limita a responsabilidade penal a uma finalidade, mas a um processo causal subsequente, de modo que pode haver responsabilização penal mesmo por consequências indesejáveis e imprevisíveis de uma ação voluntária do autor (WELZEL, 2015, p. 42); (iii) em relação aos crimes culposos, também o desvalor da ação se coloca como superior ao do resultado (WELZEL, 2015, p. 44).

Para esclarecer esta última questão, Hans Welzel (2015, p. 44) apresenta um exemplo em que dois carros, vindo em direções opostas, se chocam após fazerem uma curva em que não havia visibilidade e ambos sofrem lesões corporais. No entanto, o carro A permanece em sua faixa, enquanto B estava fazendo uma ultrapassagem proibida na outra faixa, em que estava A. Neste caso, o resultado lesivo pode ser atribuído aos dois motoristas, de modo que o desvalor do resultado estaria presente na conduta de ambos, o que leva o autor a concluir que “o elemento decisivo do injusto dos delitos culposos é também o desvalor da ação e não o mero desvalor do resultado” (WELZEL, 2015, p. 44).

Em resposta a esta teoria, surge, a partir da década de 30 do Século XX, a teoria finalista – atribuída principalmente a Hans Welzel – segundo a qual a ação humana é composta, em sua essência, “não no puro fenômeno natural da causação, e sim no direcionamento, guiado pela vontade humana, de um curso causal no sentido de um determinado fim antes tomado em vista” (ROXIN, 2008, p. 102). O finalismo passou a exigir o “desvalor da ação, ao lado do desvalor do resultado”, sendo que “esse desvalor da ação esgotava-se, por sua vez, em algo subjetivo, na finalidade” (GRECO, 2014, p. 27).

Assim, de acordo com Welzel (2015, p. 37):

A estrutura final da conduta humana deve necessariamente ser levada em conta pelas normas de Direito Penal. As normas jurídicas, isto é, as proibições e os mandatos do Direito, não podem se destinar a processos causais cegos, mas tão somente a ações, que podem configurar finalisticamente o futuro. As normas só podem mandar ou proibir uma conduta final.

Segundo o finalismo, portanto, o elemento subjetivo passa a ser necessário para a tipificação de um delito, de modo que o dolo deixa de ser elemento constitutivo tão somente da culpabilidade, passando a integrar a tipicidade das condutas (ROXIN, 2008, p. 102).

Formalmente, o Código Penal brasileiro acolhe a teoria do finalismo, estabelecendo como regra geral a necessidade de dolo – ou culpa, excepcionalmente – para que haja crime⁷⁸. No entanto, o finalismo não é uma teoria completa e longe de críticas, havendo, também, suas próprias limitações.

Segundo Roxin (2008, p. 103), em face de ser a causalidade o critério de tipicidade objetiva, “o tipo continua demasiado extenso”. Assim, “se a doutrina anterior deixava o tipo esgotar-se na causalidade, o finalismo passa a considerar a causalidade apenas a metade da teoria do tipo, o

⁷⁸ Neste sentido, o Código Penal estabelece como regra geral os crimes dolosos, vinculando a criminalização à intenção do agente: “Art. 18 - Diz-se o crime: Crime doloso: I - doloso, quando o agente quis o resultado ou assumiu o risco de produzi-lo; Crime culposo: II - culposo, quando o agente deu causa ao resultado por imprudência, negligência ou imperícia. Parágrafo único - Salvo os casos expressos em lei, ninguém pode ser punido por fato previsto como crime, senão quando o pratica dolosamente. (Incluído pela Lei nº 7.209, de 11.7.1984)” (BRASIL, 1940). Dessa forma, o citado dispositivo de lei estabelece a necessidade do dolo como regra para a criminalização e, ainda, a excepcional responsabilização por culpa, de modo que o elemento subjetivo passa a constituir, necessariamente, o delito.

chamado tipo ‘objetivo’” (GRECO, 2014, p. 25). Neste contexto, Luís Greco (2014, p., 25) afirma que situações como a da mãe que dá à luz ao criminoso seriam objetivamente típicas para o finalismo, de modo que somente na tipicidade subjetiva (finalidade/ dolo) é que se afastaria qualquer conduta criminosa dessa mãe⁷⁹. Ou seja, o simples dar à luz a um criminoso geraria uma relação de causalidade para o cometimento posterior do crime desse filho.

Para superar esse problema, surge a teoria da imputação objetiva, que “enuncia o conjunto de pressupostos que fazem de uma causação objetivamente típica” (GRECO, 2014, p. 24) se compõe de duas ideias: “a criação de um risco juridicamente desaprovado e a realização deste risco no resultado” (GRECO, 2014, p. 26). Esclarece ainda Roxin (2008, p. 104) a necessidade deste resultado se encontrar “dentro do alcance do tipo”. Consoante Luís Greco, “a imputação objetiva acrescenta ao injusto um desvalor objetivo da ação (a criação de um risco juridicamente desaprovado) e dá ao desvalor do resultado uma nova dimensão (realização do risco juridicamente desaprovado)” (GRECO, 2014, p. 28).

Assim, a imputação objetiva busca compreender, na tipicidade objetiva, não somente o desvalor do resultado, mas passa a trabalhar com um desvalor objetivo da ação, de modo que esta teoria se aplicaria em casos limítrofes para as teorias anteriores. Por exemplo, no apresentado caso de acidente de carro, a responsabilidade penal de A seria afastada, uma vez que não há risco desaprovado em dirigir de acordo com as regras de trânsito, sendo um risco permitido pelo direito, diferentemente de B, que teria criado um risco não permitido pelo direito ao desrespeitar as normas de trânsito vigentes (ROXIN, 2008, p. 110-11). A situação da mãe de um futuro criminoso também se resolve objetivamente por esta teoria, uma vez que parir não é criação de risco, nem contrário ao direito, de modo que não há tipicidade objetiva nesta ação.

Desta forma, a regra vigente no direito penal brasileiro é a da responsabilidade penal subjetiva (por dolo ou culpa) de base finalista, ainda que a imputação objetiva possa ser empregada como mecanismo para auxiliar a compreender a existência de tipicidade objetiva nos casos analisados.

Em relação ao dolo, Juarez Tavares (1971, p. 107) o define como “consciência e vontade em relação aos elementos objetivos pertencentes ao tipo”. Tamanha a importância da consciência para a constituição do elemento subjetivo do tipo penal, que sua ausência exclui o dolo, no direito brasileiro⁸⁰.

⁷⁹ Para a teoria da ação causal, tal situação é ainda mais complicada, pois também haveria tipicidade, que não seria afastada pela ausência do elemento subjetivo, de modo que deveria haver recurso ao rompimento do nexa causal.

⁸⁰ Segundo determina o Art. 20 do Código Penal brasileiro, “o erro sobre elemento constitutivo do tipo legal de crime exclui o dolo, mas permite a punição por crime culposo, se previsto em lei. (Redação dada pela Lei nº 7.209, de 11.7.1984)” (BRASIL, 1941).

Sobre as modalidades do dolo, este pode ser direto ou eventual. Segundo o autor, “há dolo direto, quando a vontade de realização associa-se necessariamente à representação das circunstâncias típicas como objetivo final de sua ação, meio para a consecução de outros objetivos ou como conseqüências acompanhantes ao fato” (TAVARES, 1971, p. 110). Assim, o dolo é direto quando o autor almeja o resultado diretamente ou como meio para chegar a um outro resultado efetivamente desejado, de modo que “no dolo direto o agente quer o resultado típico ou toma-o como necessário para a consecução de outros propósitos ou como conseqüência necessária de sua atividade” (TAVARES, 1971, p. 110).

No dolo eventual, por sua vez, o agente não deseja o resultado, mas entende que este é provável e consente com o prosseguimento da ação que pode levar a sua materialização. Assim, nas palavras de Juarez Tavares (1971, p. 112), “o autor neste caso representa as conseqüências como possíveis de serem derivadas do emprego de meios ou do alcance do objetivo e se conforma com a sua verificação”. Desse modo, a diferenciação entre dolo direto e eventual pode ser apresentada da seguinte forma: “naquele [dolo direto], a associação representativa se fixa na certeza de que as conseqüências ligar-se-ão necessariamente aos meios ou ao fim, enquanto que neste [dolo eventual] esta ligação se representa apenas como possível na mente do autor” (TAVARES, 1971, p. 112).

Ou seja, ainda que haja plena consciência dos elementos do tipo penal, no dolo direto o resultado delituoso é tomado como certo, e o autor age orientado a esse resultado. Já no dolo eventual, o resultado não é certo, mas provável, e o autor consente que essa probabilidade possa se materializar no resultado delituoso. Neste mesmo sentido, Luzón Peña (2020, p. 45) afirma que o dolo eventual se diferencia do dolo direto “porque o sujeito sabe que não é certo, senão somente possível – uma eventualidade, portanto –, que com sua conduta se realize o fato típico e, se necessário, o resultado”.

Uma outra questão relevante é a diferenciação entre o dolo eventual e a culpa consciente, modalidades limítrofe para o direito penal. Neste sentido, a diferenciação da culpa consciente para o dolo eventual não está no conhecimento, que existe de forma semelhante em ambas as situações, no entanto, para a culpa consciente acredita-se que o resultado possivelmente delituoso não ocorrerá. Neste sentido, Tavares (1971, p. 112-113) afirma que:

Porém ao representar como possível a verificação das conseqüências, pode o autor penetrar em duas situações distintas: a) - contar ou conformar-se com o acontecimento, ou b) - confiar na sua não verificação.

Na primeira (contar ou conformar-se com o acontecimento), a vontade de realização se caracteriza como dolo eventual. Na segunda, não há propriamente vontade de realização, sendo somente admitida a hipótese de culpa consciente.

A despeito da existência de divergência doutrinária em relação a qual elemento adicional deve separar o dolo eventual da culpa consciente, Luzón Peña afirma que a principal teoria que busca fazer

essa diferenciação é a do consentimento (majoritária na doutrina e jurisprudência espanholas, por exemplo), segundo a qual:

o dolo eventual, para além do conhecimento da realização - eventual - dos elementos objetivos do tipo, também requer o elemento volitivo e esse é, justamente, o elemento que o diferencia da culpa consciente, na qual não existe tal vontade. No dolo eventual, a vontade é menos intensa porque o sujeito, a rigor, não deseja nem pretende realizar o tipo, tampouco está certo sobre se produzirá ou não o resultado, mas consente com isso, considerando a eventualidade de que produza o fato típico. (LUZON PEÑA, 2020, p. 47)

Segundo o autor, a versão subjetivista desta teoria deve ser rechaçada, devendo haver “uma formulação estrita em relação à dupla conceitual aceitar/confiar na não produção” (LUZON PEÑA, 2020, p. 48). Assim, segundo o autor, amparado nesta concepção restritiva do consentimento:

[...] o dolo eventual requer o elemento volitivo como forma de aceitação ou consentimento – ou termos similares – da eventual produção do fato, mas que mediante uma *avaliação objetivo-normativa* é possível restringir o que se entende por aceitação/não aceitação. A restrição mais usual e importante consiste em considerar que a aceitação (ou consentimento ou similares) não se exclui com base em confiança irracional e infundada na produção do fato – julgando-se que essa confiança meramente subjetiva não é autêntica confiança, senão esperança ou desejo juridicamente irrelevante –, mas sim que *a aceitação ou o consentimento somente são excluídos por uma confiança minimamente fundada objetivamente*, conquanto errônea, de que o fato não se produza. (LUZON PEÑA, 2020, p. 55)

Em sentido semelhante, Claus Roxin (2020, p. 25), foge da delimitação unicamente subjetiva para separar o dolo eventual da culpa consciente, afirmando que:

a delimitação entre o *dolus eventualis* e a culpa consciente não pode prescindir de parâmetros normativos de avaliação. O substrato dessa avaliação não pode, entretanto, ser delimitado por determinada representação do perigo. O juízo sobre se o autor – seja em uma emergência ou naturalmente – se decidiu contra o bem jurídico proibido deve ser traçado considerando-se todos os elementos – objetivos e subjetivos – da situação criminosa relevantes para a atitude do autor.

Dessa forma, nos dois casos há previsibilidade do acontecer típico, com a diferença que na culpa consciente o autor acredita que o resultado não ocorrerá e, por outro lado, no dolo eventual, embora não deseje o resultado, o autor prevê seu acontecimento e não encerra seu curso causal. Para o presente trabalho, portanto, a diferenciação entre dolo eventual e culpa consciente parte do proposto por Luzón Peña na teoria do consentimento restrita, havendo dolo quando o autor aceita/consente com a realização de resultado, só ocorrendo seu afastamento se há elementos sérios para fazê-lo pensar que o resultado criminoso será evitado, posição defendida tanto por Luzón Peña quanto por Roxin, conforme apresentado acima.

No entanto, as questões de direito material necessitam de comprovação, por meio das normas processuais aplicáveis. Por essa razão, e como defendido por Wunderlich e Ruivo (2022, p. 25), “sem a demonstração probatória da anuência ou aceitação do resultado, não se pode falar em dolo eventual,

pois segue em aberto a real hipótese de culpa consciente”. Sendo assim, há necessidade de efetiva comprovação do elemento subjetivo, também em relação ao dolo eventual.

Uma última questão relevante, mas que não será tratado com centralidade no presente trabalho é o fato de que o dolo eventual é uma modalidade de dolo como outra qualquer, admitindo que haja crime tentado quando o resultado não se consuma. Neste sentido, Luzón Peña (2020, p. 55) afirma que, “tanto conceitualmente quanto por razões materiais, a tentativa (e sua punição), do mesmo modo que é punível no caso de dolo direto, também é compatível com o dolo eventual”.

Assim, em resumo, a regra para a responsabilização penal no direito brasileiro é o dolo – seja direto, seja eventual – enquanto elemento subjetivo, e somente a partir da existência deste elemento que se pode falar de eventual existência de racismo algorítmico ou não. No mais, a teoria da imputação objetiva – além da mera causalidade – será o critério de atribuição de tipicidade objetiva nos casos estudados.

2.3.3. Autoria no direito penal brasileiro

Outra questão fundamental para compreender crimes cometidos por meio de estruturas empresariais complexas – como nas empresas de desenvolvimento de tecnologias – é a forma de atribuição e individualização da autoria dentro dessas estruturas, motivo pelo qual se passa a tratar, panoramicamente, das formas de atribuição de autoria no direito brasileiro. O Código Penal determina que é autor quem concorre para o crime, não diferenciando as figuras de autor e partícipe⁸¹, atribuindo as mesmas penas aos envolvidos na prática criminosa, à exceção da participação de menor importância, que possui tratamento mais brando⁸².

Dessa forma, consoante Juarez Cirino dos Santos (2008, p. 361), ainda que formalmente a autoria no direito brasileiro adote a teoria unitária, na prática é um sistema diferenciador⁸³. Em decorrência disso, prossegue o autor, “autoria e participação devem ser estudadas segundo os postulados da teoria do domínio do fato, generalizados na literatura contemporânea como critérios de definição de autor e de partícipe”. (SANTOS, p. 361). Sobre o domínio do fato, GRECO *et al.* (2014,

⁸¹ Segundo o *caput* do art. 29 do Código Penal, “quem, de qualquer modo, concorre para o crime incide nas penas a este cominadas, na medida de sua culpabilidade”.

⁸² No direito penal brasileiro, a participação só possui tratamento diferenciado se for considerada de menor importância, de acordo com o art. 29, §1º, do Código Penal. A regra, contudo, é que tanto autor quanto partícipe são punidos com a mesma gravidade.

⁸³ De acordo com Greco *et al.* (2014, p. 13-14): “o sistema unitário é aquele que (a) iguala todas as contribuições causais para o delito e prevê marcos penais idênticos para todos (*sistema unitário formal*), ou (b) que distingue no plano conceitual entre várias formas de contribuição, mas continua a prever os mesmos marcos penais para todas (*sistema unitário funcional*). O sistema diferenciador distingue várias formas de intervenção no delito, isto é, distingue entre autores e partícipes e pode prever, ademais, diferentes marcos penais para as diferentes formas de intervenção”.

p. 22) apontam que o objetivo de teoria é justamente o de “*distinguir entre autor e partícipe*”. Neste contexto, buscaremos desenvolver, de acordo com a mencionada teoria, os conceitos de autor e partícipe.

Segundo Roxin⁸⁴, autor é quem possui o domínio do fato, que pode se distinguir em 3 categorias: i) domínio da ação (ou autoria imediata), que consiste no “domínio de quem realiza, em sua própria pessoa, todos os elementos de um tipo” (GRECO, L. *et al.*, 2014, p. 25); (ii) domínio funcional do fato (ou coautoria), que ocorre quando “duas ou mais pessoas, partindo de uma decisão conjunta de praticar o fato, contribuem para a sua realização com um ato relevante de um delito” (GRECO, L. *et al.*, 2014, p. 30); (iii) domínio da vontade (ou autoria mediata), por sua vez, pode ocorrer de três diversas formas, como se passa a apresentar.

A primeira é a coação, que ocorre quando o autor mediato coage a pessoa que funcionará como seu instrumento para a prática criminosa. A segunda é o erro, que se dá quando o autor mediato se vale de um autor em erro – que não sabe estar cometendo crime, por exemplo – para a prática criminosa desejada. Além disso, é possível o domínio do fato por meio do domínio da organização (GRECO, L. *et al.*, 2014, p. 26-30). Esta última modalidade ocorre quando há, além do aparato organizado de poder – que deve ser, necessariamente, dissociado do estado – em que o superior dá ordem a executores fungíveis, de modo que quem determina a ordem não é partícipe (instigador), mas autor.

Existe debate na doutrina em relação à possibilidade de aplicação desta última variante do domínio do fato em relações a instituições em conformidade com o direito. Neste contexto, Greco (2014, p. 101) afirma que a teoria do domínio da organização foi inicialmente pensada por Roxin para o caso de estados totalitários, mas que o Tribunal Constitucional Alemão entendeu que a teoria seria “aplicável a organizações empresariais”.

A despeito do debate sobre os limites da aplicabilidade do *domínio da organização* enquanto domínio do fato, o presente trabalho se ampara nas ideias de Greco (2014, 104-105), no sentido de que a teoria só deve ser aplicável nos casos originalmente propostos por Roxin, seja pela inexistência de executores fungíveis, seja por não se tratar de estrutura apartada do direito.

Dessa forma, é pela aplicação das regras de autoria previstas no direito brasileiro e sua complementação pela teoria do domínio do fato que será feita a análise dos casos no capítulo seguinte.

⁸⁴ Consoante Greco *et al.* (2014, p. 21-24), a teoria surgiu no contexto alemão com o objetivo de diferenciar quem é autor e quem é partícipe de determinado crime, uma vez que o sistema alemão distingue essas figuras. Segundo os autores, ainda que haja trabalhos anteriores de autores como Hegler, Lobe e até mesmo Welzel, é somente com a tese de habilitação de Roxin sobre o tema, publicada originalmente em 1962, que a teoria toma contornos próximos dos atuais.

Além disso, ainda que haja relevantes questões de política criminal relacionadas à responsabilidade penal por IA, elas serão abordadas de forma mais precisa em momento posterior.

2.3.4. Culpabilidade no direito penal brasileiro

Consoante Roxin (2008, p. 134), a dogmática jurídico-penal se preocupa em “determinar sob que pressupostos e em que medida alguém pode ser responsabilizado por um comportamento socialmente lesivo, de maneira que se apliquem as sanções desse ramo do direito. Este é o cerne do conceito de culpabilidade. Sobre o princípio da culpabilidade, afirma Cirino dos Santos, que este é:

o segundo mais importante instrumento de proteção individual no moderno Estado Democrático de Direito, porque *proíbe punir* pessoas que não preenchem *os* requisitos do *juízo de reprovação*, segundo o estágio atual da teoria da culpabilidade, a saber: a) pessoas *incapazes de saber o que fazem* (inimputáveis); b) pessoas imputáveis que, realmente, *não sabem o que fazem*, porque estão em situação de *erro de proibição inevitável*; c) pessoas imputáveis, com conhecimento da proibição do fato, mas *sem o poder de não fazer o que fazem*, porque realizam o tipo de injusto em contextos de anormalidade definíveis como situações de exculparão. (SANTOS, 2003, p. 24-25)

Ainda buscando definir a culpabilidade, Roxin (2008, p. 138) a descreve como “a realização do injusto apesar da idoneidade para ser destinatários de normas e ter capacidade de autodeterminação que daí deve decorrer”. Dessa forma, e em linhas gerais, a culpabilidade do agente mora em: (i) ser capaz de saber o que faz; (ii) saber o que faz; (iii) ter podido agir de modo diverso.

Neste sentido, a própria legislação penal brasileira possui dispositivos no sentido de afastar e/ou mitigar a responsabilidade penal de agentes considerada sua inexistente ou baixa culpabilidade, como é o caso dos agentes considerados inimputáveis, segundo *caput* o art. 26 do Código Penal⁸⁵.

No entanto, e reconhecendo-se que a questão da culpabilidade é uma das mais profundas e complicadas no direito penal, o presente trabalho não busca se aprofundar nesta definição ou em suas consequências, mas, tão somente reconhecer que nos casos de responsabilidade penal decorrente de IA, a questão da culpabilidade ocupa papel central, de modo que não somente os agentes serão responsáveis pelos fatos em relação aos quais sejam objetiva e subjetivamente imputáveis, mas, ainda, em relação aos quais também sejam culpáveis.

Esta é, portanto, mais uma dificuldade para o estabelecimento de responsabilidade penal o IA, segundo o modelo vigente.

⁸⁵ Segundo a mencionada norma: “É isento de pena o agente que, por doença mental ou desenvolvimento mental incompleto ou retardado, era, ao tempo da ação ou da omissão, inteiramente incapaz de entender o caráter ilícito do fato ou de determinar-se de acordo com esse entendimento. (Redação dada pela Lei nº 7.209, de 11.7.1984)” (BRASIL, 1941, n.p.)

2.3.5. Legalidade no direito penal brasileiro

Outro princípio central a nortear o direito penal é o da *legalidade estrita*. Segundo Luigi Ferrajoli (2002, p. 30), a primeira condição (*reserva legal*) para a satisfação desse princípio é a *formal*, segundo a qual “o juiz não pode qualificar como delitos todos (ou somente) os fenômenos que considere imorais ou, em todo caso, merecedores de sanção, mas apenas (e todos) os que, independentemente de sua valoração, venham formalmente designados pela lei como pressupostos de uma pena”, correspondendo ao brocado latino “*nulla poena et nullum crimen sine lege*”. A segunda parte deste princípio corresponde à ideia de que “a definição legal do desvio deve ser produzida não com referência a figuras subjetivas de *status* ou de autor, mas somente a figuras empíricas e objetivas de comportamento, segundo a outra máxima clássica: *nulla poena sine crimine et sine culpa*” (FERRAJOLI, 2002, p. 30).

Ainda sobre o *princípio da legalidade*, Juarez Cirino dos Santos afirma que este:

[...] é o mais importante instrumento constitucional de proteção individual no moderno Estado Democrático de Direito, porque proíbe (a) a *retroatividade* como criminalização ou agravamento da pena de fato anterior, (b) o *costume* como fundamento ou agravamento de crimes e penas, (c) a *analogia* como método de criminalização ou de punição de condutas e (d) a *indeterminação* dos tipos legais e das sanções penais (art. 5º, XL, CR). (SANTOS, 2008, p. 20)

No Brasil, a Constituição Federal de 1988 estabelece a legalidade como princípio constitucional a reger o direito penal. No inciso XXXIX de seu art. 5º fica estabelecido que “não há crime sem lei anterior que o defina, nem pena sem prévia cominação legal”. No mesmo sentido, o art. 1º do Código Penal determina que “Não há crime sem lei anterior que o defina. Não há pena sem prévia cominação legal” (BRASIL, 1941, n.p.), dando aplicabilidade à legalidade estrita, também em nível infraconstitucional.

Em relação a este trabalho, é importante compreender a legalidade enquanto limitador de criminalizações não descritas na norma penal incriminatória e, ainda, de incriminações imprecisas. Este é, portanto, mais um obstáculo para a configuração do crime de racismo nos casos estudados: deve haver criminalização expressa e suficientemente precisa das condutas incriminadas.

2.4. CONSIDERAÇÕES FINAIS

A IA possui grandes potencialidades, ampliando as capacidades do ser humano em diversas aplicações. No entanto, essas potencialidades são geradas em um contexto de ampliação dos riscos em face dessas tecnologias. Esses riscos podem se materializar em diversas formas, como, por

exemplo, na violação à privacidade e aos dados pessoais, mas também na discriminação de grupos sociais. Em decorrência disso, é necessário que haja um cenário regulatório adequado à mitigação desses riscos. Neste contexto, os princípios do Constitucionalismo Digital são fundamentais para evitar a violação de direitos das pessoas envolvidas na operação de algoritmos pelas *Big Techs*.

Em relação aos algoritmos de seleção e recrutamento, é possível verificar diversos casos em que as empresas de tecnologia que têm intermediado esse processo, ainda que não intencionalmente, têm permitido ou praticado a discriminação e exclusão de grupos sociais, como ocorre com negros, estrangeiros e mulheres. Dessa forma, uma tecnologia que seria usada para reduzir discriminação e tornar o processo de seleção de emprego mais eficiente pode gerar distorções. Contudo, como as empresas não divulgam a forma de operação de seus algoritmos, essa discriminação pode ocorrer sem que se tenha conhecimento dos critérios empregados na seleção, tornando difícil o controle desses algoritmos. Por isso, e considerando-se as possibilidades de discriminação no sentido de limitar o acesso de determinados grupos ao mercado de trabalho, tal situação pode configurar, em tese, o crime de racismo.

Na literatura apresentada sobre a responsabilidade penal e IA, é possível perceber que existem dificuldades no estabelecimento de responsabilização em decorrência de algoritmos. Isso se dá, por um lado, pela rejeição à responsabilização penal da IA diretamente, e, por outro lado, pelas dificuldades de responsabilização das pessoas envolvidas na operação de algoritmos. Neste contexto, muitas das dificuldades decorrem da exigência de intenção (*mens rea*) nos sistemas penais pelo mundo, de modo que deve haver como demonstrar a intenção do desenvolvedor/operador da IA no sentido do cometimento de crimes.

Considerado esse cenário, surgem diversas possibilidades de soluções em relação à responsabilização decorrente de “atos” criminosos praticados por formas de IA. Entre essas soluções, apresenta-se como possibilidade a redução do direito penal e o surgimento da regulação enquanto ferramenta para evitar ou impor limites aos danos causados por IA. Outra solução possível é considerar determinados danos como decorrentes do risco existente na terceirização de decisões tipicamente humanas a ferramentas tecnológicas. Mais uma possibilidade apresentada seria a de proibir ou controlar previamente o lançamento de determinados algoritmos para mitigar ou impedir a realização de determinados riscos.

Tomando-se como cenário a legislação penal vigente no Brasil, a responsabilidade penal só pode ocorrer quando há criminalização expressa, em respeito aos critérios da legalidade estrita e à necessidade de caracterização do elemento subjetivo (dolo ou culpa) e em relação a pessoas físicas – excepcionando-se os crimes ambientais, que não são objeto do presente estudo. Assim, as dificuldades encontradas no estabelecimento de responsabilidade penal também existem no cenário

brasileiro, em que há necessidade de responsabilização subjetiva de pessoas físicas e de configuração das regras de autoria previstas no Código Penal.

Considerando-se o apresentado neste capítulo, após a compreensão do cenário discriminatório nos algoritmos de seleção e recrutamento e dos modelos penais vigentes, é necessário o prosseguimento das pesquisas para realizar o estudo de casos e compreender a adequação dos atuais modelos penais em relação ao crime de racismo nos algoritmos de seleção e recrutamento.

3. RACISMO ALGORÍTMICO: DEFINIÇÃO DO CONCEITO E ESTUDO DE CASOS

3.1. O RACISMO ENQUANTO FENÔMENO SOCIAL E JURÍDICO NO BRASIL

De acordo com Almeida (2019, p. 18), no contexto da expansão comercial burguesa e da cultura renascentista do Século XVI, “abriu-se as portas para a construção do moderno ideário filosófico que mais tarde transformaria o europeu no *homem universal*”, e “todos os povos e culturas não condizentes com os sistemas europeus em varrições menos evoluídas”. Ainda, segundo o autor, no iluminismo, o modelo filosófico colocou o homem em seu centro, e, com base em um conhecimento fundado na observação do homem e de suas diferenças, surge “a distinção filosófico-antropológica entre *civilizado* e *selvagem*, que no século seguinte daria lugar para o dístico *civilizado* e *primitivo*” (ALMEIDA, 2019, p. 19). Em decorrência disso, os *civilizados* foram levar essa *civilização* para os lugares que ainda não a conheciam, o que “redundou em um processo de destruição e morte, de espoliação e aviltamento, feito em nome da *razão* a que se denominou *colonialismo*” (ALMEIDA, 2019, p. 19)

Sobre o colonialismo, Duarte e Queiroz (2016, p. 10-13) apontam as tensões raciais existentes, com os crescentes direitos dos brancos conferidos na Revolução Francesa (e nas metrópoles) em contraste com a escravidão nas colônias. Neste contexto, os autores apontam que um grande lócus desta tensão foi a Revolução Haitiana, em que foi criado o “primeiro Estado independente constituído por ex-escravos e negros libertos em janeiro de 1805” (DUARTE; QUEIROZ, 2016, p. 29). Em relação ao debate racial, a Constituição Haitiana pós-independência subverteu os sentidos considerados naturais, decretando que “todos os habitantes haitianos devem ser tratados como ‘negros’, em uma verdadeira reapropriação da linguagem do colonizador pelo colonizado” (DUARTE; QUEIROZ, 2016, p. 32).

No entanto, segundo os autores, este fenômeno histórico fundamental para compreender a luta por direitos civis, especialmente dos negros, mas também dos colonizados de forma mais ampla, “se perdeu ou foi ignorado pela historiografia e filosofia política justamente por estar atrelado a um

silenciamento ainda mais profundo, que é aquele exercido sobre o que foi e o que simbolizou a Revolução Haitiana” (DUARTE; QUEIROZ, 2016, p. 11). Neste sentido, concluem os autores que o “não-pensar da Revolução Haitiana significa o não-pensar da humanidade de negros e negras” (DUARTE; QUEIROZ, 2016, p. 35).

Neste contexto, o racismo, segundo Sílvio de Almeida, é “uma forma sistemática de discriminação que tem a raça como fundamento, e que se manifesta por meio de práticas conscientes ou inconscientes que culminam em desvantagens ou privilégios para indivíduos, a depender do grupo racial ao qual pertençam” (ALMEIDA, 2019, p. 22).

Já a discriminação racial, componente do racismo, pode ser definida como:

a atribuição de tratamento diferenciado a membros de grupos racialmente identificados. Portanto, a discriminação tem como requisito fundamental o poder, ou seja, a possibilidade efetiva do uso da força, sem o qual não é possível atribuir vantagens ou desvantagens por conta da raça. Assim, a discriminação pode ser direta ou indireta. A discriminação direta é o repúdio ostensivo a indivíduos ou grupos, motivado pela condição racial, exemplo do que ocorre em países que proíbem a entrada de negros, judeus, muçulmanos, pessoas de origem árabe ou persa, ou ainda lojas que se recusem a atender clientes de determinada raça. [...] Já a discriminação indireta é um processo em que a situação específica de grupos minoritários é ignorada – discriminação de fato –, ou sobre a qual são impostas regras de “neutralidade racial” – colorblindness – sem que se leve em conta a existência de diferenças sociais significativas – discriminação pelo direito ou discriminação por impacto adverso. (ALMEIDA, 2019, p. 23)

O racismo possui três formas de se apresentar, a *individual*, a *institucional* e a *estrutural*⁸⁶ (ALMEIDA, 2019, p. 15). Em perspectiva *individual*, o racismo seria uma patologia, como um “fenômeno ético ou psicológico de caráter individual ou coletivo, atribuído a grupos isolados” (ALMEIDA, 2019, p. 25), sendo, portanto, um fenômeno individual dissociado de um contexto social. Já a concepção *institucional* defende que o “racismo não se resume a comportamentos individuais, mas é tratado como o resultado do funcionamento das instituições, que passam a atuar em uma dinâmica que confere, ainda que indiretamente, desvantagens e privilégios com base na raça” (ALMEIDA, 2019, p. 26). Dentro da concepção institucional de racismo,

as instituições atuam na formulação de regras e imposição de padrões sociais que atribuem privilégios a um determinado grupo racial, no caso, os brancos. E um exemplo disso é a exigência de ‘boa aparência’ para se candidatar a uma vaga de emprego, que simultaneamente é associada a características estéticas próprias de pessoas brancas. (ALMEIDA, 2019, p. 30-31)

⁸⁶ Segundo o autor, “a tese central é a de que o racismo é sempre estrutural, ou seja, de que ele é um elemento que integra a organização econômica e política da sociedade. Em suma, o que queremos explicitar é que o racismo é a manifestação normal de uma sociedade, e não um fenômeno patológico ou que expressa algum tipo de anormalidade. O racismo fornece o sentido, a lógica e a tecnologia para a reprodução das formas de desigualdade e violência que moldam a vida social contemporânea”. (ALMEIDA, 2019, p. 15)

Neste sentido, ainda que com uma aparência de legitimidade e de imparcialidade, as instituições passam a criar critérios subjetivos que permitem discriminar sem assumir diretamente que o fizeram.

A concepção *estrutural*, por sua vez, compreende que o racismo está presente nas instituições por algo ainda mais profundo, estando entranhado nas estruturas das sociedades. Neste sentido:

[...] se é possível falar de um racismo institucional, significa que a imposição de regras e padrões racistas por parte da instituição é de alguma maneira vinculada à ordem social que ela visa resguardar. Assim como a instituição tem sua atuação condicionada a uma estrutura social previamente existente – com todos os conflitos que lhe são inerentes –, o racismo que essa instituição venha a expressar é também parte dessa mesma estrutura. As instituições são apenas a materialização de uma estrutura social ou de um modo de socialização que tem o racismo como um de seus componentes orgânicos. Dito de modo mais direto: as instituições são racistas porque a sociedade é racista. (ALMEIDA, 2019, p. 31)

Fugindo da concepção individualista, Eduardo Bonilla-Silva (2017, p. 11) aponta a existência de um racismo sem ‘racistas’ ou racismo com neutralidade racial (*color-blind racism*). Segundo o autor:

Essa ideologia, que adquiriu coesão e se tornou predominante no final da década de 1960, explica a desigualdade racial contemporânea como resultado de dinâmicas não-raciais. Enquanto o racismo de Jim Crow explicava a posição social dos negros como resultado de sua inferioridade biológica e moral, o racismo racialmente neutro evita tais argumentos fáceis. Em vez disso, os brancos racionalizam o status contemporâneo das minorias como produto da dinâmica do mercado, fenômenos naturais e limitações culturais imputadas aos negros. Por exemplo, os brancos podem atribuir a alta taxa de pobreza dos latinos a uma ética de trabalho relaxada [...] ou segregação residencial como resultado de tendências naturais entre grupos. (BONILLA-SILVA, 2017, p. 11)⁸⁷.

Dessa forma, o racismo presente na atualidade é o que desconsidera o histórico racial e a existência de determinadas estruturas sociais que favorecem brancos em relação a negros, apontando que essa crença de que, na atualidade, as relações raciais não geram distinções entre as pessoas “se tornou uma formidável ferramenta política para a manutenção da ordem racial” (BONILLA-SILVA, 2017, p. 12)⁸⁸.

Para o presente trabalho, sem desconsiderar que o racismo pode ser demonstrado de diversas formas, adota-se sua concepção enquanto fenômeno estrutural. A seguir, passa-se a contextualizar e apresentar o racismo no cenário brasileiro, e, ainda, no que trata de sua variante tecnológica, o racismo algorítmico.

⁸⁷ Tradução nossa. No original: “This ideology, which acquired cohesiveness and dominance in the late 1960s, explains contemporary racial inequality as the outcome of nonracial dynamics. Whereas Jim Crow racism explained blacks’ social standing as the result of their biological and moral inferiority, color-blind racism avoids such facile arguments. Instead, whites rationalize minorities’ contemporary status as the product of market dynamics, naturally occurring phenomena, and blacks’ imputed cultural limitations. For instance, whites can attribute Latinos’ high poverty rate to a relaxed work ethic [...] or residential segregation as the result of natural tendencies among groups.”

⁸⁸ Tradução nossa. No original: “has become a formidable political tool for the maintenance of the racial order”.

3.1.1. O cenário antes da Lei 7.716/1989 e da Constituição Federal de 1988

No contexto brasileiro anterior à Constituição Federal de 1988, “a Lei n. 1390, de 3 de julho de 1951, foi a primeira lei penal brasileira a prever expressamente como contravenções penais⁸⁹ condutas baseadas na etnia ou cor de pele” (SILVA; SFOGGIA, 1997, p. 11). A lei é conhecida como Afonso Arinos em homenagem ao jurista e Deputado Federal que apresentou seu projeto e “teria surgido a partir de uma motivação de ordem pessoal de Afonso Arinos, após um episódio em que seu motorista particular teria sido barrado numa confeitaria do Rio de Janeiro, por ser negro” (SILVA; SFOGGIA, 1997, p. 11). Outra situação a que também se atribui a criação da lei é o episódio em que “a bailarina norte-americana Katherine Dunham fora barrada num hotel de luxo em São Paulo ou Rio de Janeiro por ser negra” (SILVA; SFOGGIA, 1997, p. 12).

Entretanto, por suas imperfeições técnicas, a mencionada lei teve pouca aplicabilidade prática (SILVA; SFOGGIA, 1997, p. 12). Quando falam de imperfeições técnicas os autores parecem fazer referência às formas de discriminação tão somente como vedação de acesso a determinados lugares e, ainda, à enumeração dos locais em que teria aplicabilidade⁹⁰. Neste sentido:

Se, de um lado, a lei contempla uma tautológica tipificação de atos de recusa, oposição ou negação de acesso (material ou formal - este na hipótese de matrícula, inscrição ou admissão em escola), com a sisífica tentativa de abarcar todos os lugares em que o delito poderia ocorrer, acaba, por outro lado, excluindo uma série de outros atos discriminatórios, já que o legislador jamais será capaz de enumerar, de forma exaustiva, a miríade de condutas racistas, multiplicadas pela rica imaginação e criatividade daquele que as pratica. A título de exemplo, lembramos a mais comum das manifestações racistas, que é difamar ou injuriar determinada pessoa, expondo-a a desprezo público, por causa de sua raça, cor da pele ou de sua origem étnica. (SILVA; SFOGGIA, 1997, p. 13)

Para resolver essa impropriedade técnica, Silva e Sfoggia defendem que “a tipificação do crime de racismo jamais poderá ser enumerativa, até por força das restrições ao emprego da analogia, em tema de Direito Penal” (SILVA; SFOGGIA, 1997, p. 14).

⁸⁹ Sobre as contravenções penais, afirma C ezar Bitencourt (2010, p. 253): “As contraven es, que por vezes s o chamadas de *crimes-an es*, s o condutas que apresentam menor gravidade em rela ao aos crimes, por isso sofrem san es mais brandas. O fundamento da distin ao   puramente *pol tico-criminal*. O crit rio   simplesmente quantitativo ou extr nseco, com base na san ao assumindo car ter formal. Com efeito; nosso ordenamento jur dico aplica a pena de pris o, para os crimes, sob as modalidades de reclus o e deten ao, e, para as contraven es, quando for o caso, a de pris o simples (Decreto-lei n. 3.914/41). Assim, o crit rio distintivo entre crime e contraven ao   dado pela natureza da pena privativa de liberdade cominada.”

⁹⁰ Neste sentido, algumas das contraven es penais previstas s o: “Art. 2 . Recusar algu m hospedagem em hotel, pens o, estalagem ou estabelecimento da mesma finalidade, por preconceito de ra a ou de cor. [...] Art. 3 . Recusar a venda de mercadorias em lojas de qualquer g nero, ou atender clientes em restaurantes, bares, confeitarias e locais semelhantes, abertos ao p blico, onde se sirvam alimentos, bebidas, refrigerantes e guloseimas, por preconceito de ra a ou de cor. Art. 4 . Recusar entrada em estabelecimento p blico, de divers es ou esporte, bem como em sal es de barbearias ou cabeleireiros, por preconceito de ra a ou de cor.” (BRASIL, 1951, n.p.)

Mais especificamente no que trata da discriminação em relação ao acesso ao mercado de trabalho, a lei aponta como contravenções as seguintes condutas:

Art. 6º Obstar o acesso de alguém a qualquer cargo do funcionalismo público ou ao serviço em qualquer ramo das forças armadas, por preconceito de raça ou de cor. [...]

Art. 7º Negar emprego ou trabalho a alguém em autarquia, sociedade de economia mista, empresa concessionária de serviço público ou empresa privada, por preconceito de raça ou de cor. (BRASIL, 1951, n.p.)

Sobres os citados arts. 6º e 7º, os autores defendem que “tínhamos aqui mais um dispositivo inócuo, eis que centrado na esperança de que o empregador, ostensivamente, recusasse o ingresso de determinado candidato, explicitado seu preconceito” (SILVA; SFOGGIA, 1997, p. 14).

Em 20 de dezembro de 1985 foi editada a Lei nº 7.437, que alterou a Afonso Arinos, “incluindo sob sua tutela os atos discriminatórios centrados em preconceitos de ‘sexo ou estado civil’⁹¹ (SILVA; SFOGGIA, 1997, p. 14). Além disso, foi criado o art. 6º prevendo como contravenção “recusar a entrada de alguém em qualquer tipo de estabelecimento comercial ou de prestação de serviço, por preconceito de raça, de cor, de sexo ou de estado civil” (BRASIL, 1985, n.p.). No entanto, Silva e Sfoggia (1997, p. 15) defendem que as alterações realizadas não cobrem outras formas de discriminação como recusar-se a atender pessoa em função de discriminação e, também, não vale para edifícios residenciais, mas, tão somente, comerciais.

Este é o cenário de combate ao racismo anteriormente à edição da Constituição Federal de 1988. A seguir, passa-se a apresentar o cenário mais recente, já no atual momento constitucional.

3.1.2. A lei de racismo, seu contexto e seus tipos penais

Antes de apresentar a lei vigente de criminalização das práticas de racismo, entende-se por necessário apresentar o panorama de direito internacional e de direito constitucional que ampara a legislação de combate ao racismo.

Consoante Flávia Piovesan (2022, p. 69), “a internacionalização dos direitos humanos constitui, assim, um movimento extremamente recente na história, que surgiu a partir do pós-guerra,

⁹¹ A título de exemplo, apresenta-se os artigos da Lei Afonso Arinos – após as alterações feitas pela Lei nº 7.437/1985 – destinados à discriminação no contexto do mercado de trabalho, que passaram a incluir, como os demais, discriminação em função de sexo ou estado civil: “Art. 8º. Obstar o acesso de alguém a qualquer cargo público civil ou militar, por preconceito de raça, de cor, de sexo ou de estado civil. Art. 9º. Negar emprego ou trabalho a alguém em autarquia, sociedade de economia mista, empresa concessionária de serviço público ou empresa privada, por preconceito de raça, de cor, de sexo ou de estado civil.” (BRASIL, 1985, n.p.)

como resposta às atrocidades e aos horrores cometidos durante o nazismo”. Neste contexto, e em um ambiente de consenso⁹², surge a Declaração Universal dos Direitos Humanos⁹³. Segundo a autora

A Declaração Universal de 1948 objetiva delinear uma ordem pública mundial fundada no respeito à dignidade humana, ao consagrar valores básicos universais. Desde seu preâmbulo, afirma que “todas as pessoas nascem livres e iguais em dignidade e direitos”. Para a Declaração Universal a condição de pessoa é o requisito único e exclusivo para a titularidade de direitos. A universalidade dos direitos humanos traduz a absoluta ruptura com o legado nazista, que condicionava a titularidade de direitos à pertinência à determinada raça (a raça pura ariana). (PIOVESAN, 2022, p. 75)

Neste sentido, a Declaração Universal dos Direitos Humanos estabelece a igualdade entre as pessoas, vedando distinções de raça:

Artigo 2.1. Todo ser humano tem capacidade para gozar os direitos e as liberdades estabelecidos nesta Declaração, sem distinção de qualquer espécie, seja de raça, cor, sexo, língua, religião, opinião política ou de outra natureza, origem nacional ou social, riqueza, nascimento, ou qualquer outra condição.

Artigo 7. Todos são iguais perante a lei e têm direito, sem qualquer distinção, a igual proteção da lei. Todos têm direito a igual proteção contra qualquer discriminação que viole a presente Declaração e contra qualquer incitamento a tal discriminação. (NAÇÕES UNIDAS, 1948, n.p.)

Buscando combater especificamente a discriminação racial, surge – em 1963 – a Convenção Internacional sobre a Eliminação de todas as formas de Discriminação Racial, que apresenta a definição da discriminação que busca combater:

Artigo 1º - 1. Para fins da presente Convenção, a expressão "discriminação racial" significará toda distinção, exclusão, restrição ou preferência baseada em raça, cor, descendência ou origem nacional ou étnica que tenha por objeto ou resultado anular ou restringir o reconhecimento, gozo ou exercício em um mesmo plano (em igualdade de condição) de direitos humanos e liberdades fundamentais nos campos político, econômico, social, cultural ou em qualquer outro campo da vida pública. (NAÇÕES UNIDAS, 1963, n.p.)

Como afirma Piovesan (2022, p. 89), o referido tratado internacional proíbe tanto a discriminação direta que se quanto a indireta, sendo que a forma direta ocorre quando “há a intenção de discriminar”, sendo que na indireta “uma suposta neutralidade vem de forma desproporcional a impactar grupos raciais, limitando o exercício de seus direitos” (PIOVESAN, 2022, p. 89). Neste sentido, e para combater a discriminação racial, o tratado determina que os estados-parte devem “tomar as medidas eficazes, a fim de rever as políticas governamentais nacionais e locais e para modificar, ab-rogar ou anular qualquer disposição regulamentar que tenha como objetivo criar a

⁹² Desta forma: “[...] a Declaração logrou um surpreendente consenso interestatal sobre a relevância dos direitos humanos, considerando a diversidade dos regimes políticos, dos sistemas filosóficos e religiosos e das tradições culturais dos Estados-membros da ONU que a proclamaram na Resolução n. 217-A (iii) da Assembleia Geral” (PIOVESAN, 2022, p. 75).

⁹³ Segundo Flávia Piovesan (2022, p. 77), a declaração não apresenta valor jurídico de um tratado internacional, mas de uma “resolução, que, por sua vez, não apresenta forma de lei” (PIOVESAN, 2022, p. 77). Contudo, houve seu processo de ‘juridização’ pela incorporação de seus direitos a outros documentos, como o Pacto Internacional dos Direitos Civis e Políticos e Pacto Internacional dos Direitos Econômicos, Sociais e Culturais (PIOVESAN, 2022, p. 80).

discriminação⁹⁴ ou perpetrá-la onde já existir” e, ainda, tomar “as medidas legislativas, proibir e por fim, a discriminação racial praticadas por pessoa, por grupo ou das organizações” (NAÇÕES UNIDAS, 1963, n.p.). Dessa forma, a mencionada convenção internacional passa a integrar o direito brasileiro, com hierarquia superior às leis ordinárias, seja como norma materialmente constitucional,⁹⁵ seja como norma infraconstitucional, mas supralegal⁹⁶.

Especificamente no ordenamento jurídico brasileiro, a Constituição Federal de 1988 adotou o combate ao racismo como princípio que rege o país em suas relações internacionais⁹⁷. Além disso, em seu artigo 5º *caput*, ficou estabelecido que “todos são iguais perante a lei, sem distinção de qualquer natureza, garantindo-se aos brasileiros e aos estrangeiros residentes no País a inviolabilidade do direito à vida, à liberdade, à igualdade, à segurança e à propriedade, nos termos seguintes” (BRASIL, 1988, n.p.). Ao enumerar e estabelecer estes direitos e garantias fundamentais, ficou determinado no inciso XLII de seu art. 5º que “a prática do racismo constitui crime inafiançável e imprescritível, sujeito à pena de reclusão, nos termos da lei” (BRASIL, 1988, n.p.).

Consoante Gustavo Rodrigues (2012, p. 71), esta determinação constitui o “mandado de criminalização mais explícito constante em nossa Lei fundamental”. Ao se buscar a origem desta norma constitucional, Silveira (2006, p. 67) aponta que decorreu de emenda aditiva apresentada pelo Deputado Carlos Alberto Caó, no sentido de combater a “discriminação racial, velada ou ostensiva, que afetam mais de metade da população brasileira constituída de negros ou descendentes de negros privados do exercício da cidadania em sua plenitude”.

No mesmo contexto, ainda antes da promulgação da Constituição Federal, o Deputado Caó apresentou “no dia 11 de maio de 1998, projeto de lei com o intuito de criminalizar a prática do racismo” (SILVEIRA, 2006, p. 67-68). O projeto foi aprovado como Lei nº 7.716/1989, também

⁹⁴ No entanto, o tratado prevê expressamente a possibilidade de discriminação positiva (ações afirmativas). Neste sentido, o art. 1, §4º, determina que “não serão consideradas discriminação racial as medidas especiais tomadas com o único objetivo de assegurar progresso adequado de certos grupos raciais ou étnicos ou de indivíduos que necessitem da proteção que possa ser necessária para proporcionar a tais grupos ou indivíduos igual gozo ou exercício de direitos humanos e liberdades fundamentais, contanto que, tais medidas não conduzam, em consequência, à manutenção de direitos separados para diferentes grupos raciais e não prossigam após terem sido alcançados os seus objetivos.” (NAÇÕES UNIDAS, 1963, n.p.)

⁹⁵ MAZZOULI (2009, *passim*.) defende que os tratados internacionais que tratam de direitos humanos possuem status (hierarquia) materialmente constitucional, enquanto os demais tratados internacionais teriam caráter infraconstitucional, mas supralegal. Ante a constitucionalidade material dos tratados internacionais que versem sobre direitos humanos, a previsão de constitucionalização dos tratados internacionais de direitos humanos previstos na Emenda Constitucional 45 seriam somente um meio de fazer com que esses tratados passassem a possuir status formalmente constitucional.

⁹⁶ O Plenário do Supremo Tribunal Federal, no julgamento da ADI 1.055, fez maioria em torno do voto do Min. Gilmar Mendes, que defendeu a supralegalidade, mas infraconstitucionalidade, dos tratados internacionais que versam sobre direitos humanos (BRASIL, 2017). De qualquer modo, ficou evidenciada a necessidade de que as normas ordinárias de direito interno se adequem aos tratados internacionais de direitos humanos.

⁹⁷ Segundo o mencionado art. 4º da Constituição Federal: “Art. 4º A República Federativa do Brasil rege-se nas suas relações internacionais pelos seguintes princípios: VIII - repúdio ao terrorismo e ao racismo” (BRASIL, 1988, n.p.)

conhecida como Lei Caó, cujo conteúdo se passa a apresentar. A referida lei inovou ao prever como crime, e não como meras contravenções penais, as infrações nela estabelecidas⁹⁸.

De acordo com Silva e Sfoggia (1997, p. 17), a Lei Caó, a despeito de inovar no combate à discriminação no cenário brasileiro, “incidiu no principal defeito da Lei 1390/1951: sua casuística exacerbada⁹⁹”. Segundo os autores:

É certo que algumas práticas discriminatórias foram incluídas na nova legislação. Entretanto, preferiu tipificar as condutas em basicamente três núcleos verbais (recusar, impedir e obstar), incapazes por si sós de abarcar uma grande variedade de formas de discriminação, instituindo ainda na explicitação de tais condutas discriminatórias, embora a experiência secular já tivesse sido suficiente para demonstrar que a forma velada, dissimulada de racismo é talvez a mais corriqueira e devesse ser também coibida pela lei penal. (SILVA; SFOGGIA, 1997, p. 17-18)

Desta forma, a Lei Caó parece não ser suficiente para o combate às formas de discriminação, protegendo o acesso a lugares, mas não combatendo outras formas de discriminação e, ainda, pecando pelo casuísmo excessivo em relação ao combate ao racismo e à discriminação. Neste sentido, inclusive, Rodrigues defende que a lei, a despeito de vir para dar aplicação à norma constitucional que surgiu no mesmo contexto, foi insuficiente e não atuou na mesma extensão do mandado de criminalização constitucional, tanto que a despeito da imprescritibilidade do crime de racismo segundo o texto constitucional, a Lei Caó previa a prescrição deste delito no prazo de 1 ano (G. RODRIGUES, 2012, p. 84).

A seguir, passa-se a apresentar as alterações posteriormente feitas à Lei 7.716/1989, que ainda está vigente, para que se possa compreender sua delimitação atual. As alterações posteriores criaram o art. 20 na referida lei, que veda “praticar, induzir ou incitar a discriminação ou preconceito de raça, cor, etnia, religião ou procedência nacional”, além de “fabricar, comercializar, distribuir ou veicular símbolos, emblemas, ornamentos, distintivos ou propaganda que utilizem a cruz suástica ou gamada, para fins de divulgação do nazismo” (BRASIL, 1989, n.p.). Mesmo essa norma com maior abrangência, possui dificuldades em sua aplicabilidade, de modo que “o passionalismo e

⁹⁸ De acordo com sua Art. 1º, “serão punidos, na forma desta Lei, os crimes resultantes de preconceitos de raça ou de cor” (BRASIL, 1989, n.p.).

⁹⁹ A Lei Caó, mesmo quando acertou prevendo determinados tipos penais mais amplos, manteve sua enumeração excessiva, cometendo falhas de redação legislativa e política criminal, dando tratamento diverso a situações praticamente idênticas. Neste sentido: “Assim, por exemplo, se já entendermos que o artigo 6º. da Lei 1390/51 (com a redação dada pela Lei 7437/85), ao utilizar a fórmula genérica ‘qualquer tipo de estabelecimento comercial ou de prestação de serviço’ já tornaria dispensáveis os artigos 3º. e 5º. da referida lei a manutenção daquela expressão, na Lei 7716/89 poderia englobar, simultaneamente (já que se persistiu na inconveniente enumeração exaustiva), pelo menos 6 artigos (artigos 5º., 6º., 7º., 8º.; 9º. e 10º.). Da mesma forma, olvidou-se que a prestação de serviço em qualquer ramo das forças armadas constitui o exercício de cargo da administração direta (funcionalismo militar), o que ocasionou a dispensável inserção de diferentes artigos (3º. e 13º.) que, embora tutelando o mesmo bem jurídico, cominam penas distintas (2 a 5 anos, no caso do artigo 3º. e 2 a 4 anos, na hipótese do artigo 4º), sinalizando o legislador uma menor reprovabilidade da prática discriminatória obstativa do ingresso nas Forças Armadas, o que evidentemente não faz sentido.” (SILVA; SFOGGIA, 1997, p. 18)

improvisação com que são editadas novas leis e emendadas as já existentes podem ter um efeito devastador, quase tão grave quanto a inexistência de uma lei a respeito: a impunidade” (SILVA; SFOGGIA, 1997, p. 22-23).

Para o presente trabalho, a despeito de a norma prevista no art. 20 da Lei Caó, este trabalho trata especificamente do art. 4º por ser este último mais específico acerca das relações de trabalho e suas respectivas condições de acesso. Neste cenário, cumpre apresentar as normas penais relativas à discriminação para acesso e promoção no ambiente de trabalho, que decorrem de artigos da Lei Caó em sua redação atual:

Art. 4º Negar ou obstar emprego em empresa privada.

Pena: reclusão de dois a cinco anos.

§ 1º Incorre na mesma pena quem, por motivo de discriminação de raça ou de cor ou práticas resultantes do preconceito de descendência ou origem nacional ou étnica: (Incluído pela Lei nº 12.288, de 2010) (Vigência)

I - deixar de conceder os equipamentos necessários ao empregado em igualdade de condições com os demais trabalhadores; (Incluído pela Lei nº 12.288, de 2010) (Vigência)

II - impedir a ascensão funcional do empregado ou obstar outra forma de benefício profissional; (Incluído pela Lei nº 12.288, de 2010) (Vigência)

III - proporcionar ao empregado tratamento diferenciado no ambiente de trabalho, especialmente quanto ao salário. (Incluído pela Lei nº 12.288, de 2010) (Vigência)

§ 2º Ficará sujeito às penas de multa e de prestação de serviços à comunidade, incluindo atividades de promoção da igualdade racial, quem, em anúncios ou qualquer outra forma de recrutamento de trabalhadores, exigir aspectos de aparência próprios de raça ou etnia para emprego cujas atividades não justifiquem essas exigências. (Incluído pela Lei nº 12.288, de 2010) (BRASIL, 1989, n.p.)

Sobre o mencionado tipo penal, não há necessidade de grandes comentários, uma vez que a determinação “negar ou obstar emprego em empresa privada” parece autoexplicativa. No entanto algumas questões relativas a este delito passam a ser apresentadas, conforme apontado pela doutrina. Segundo Guilherme Nucci (2010, p. 311), “o sujeito ativo [autor] é o proprietário da empresa ou a pessoa responsável pela contratação de pessoal” enquanto o “sujeito passivo [vítima] é a pessoa discriminada”. Neste sentido, o autor define que o delito é crime próprio, só podendo ser praticado “pelo encarregado da área de contratação de pessoa de empresa privada” (NUCCI, 2010, p. 312). Além do mais, segundo o autor, o tipo penal apontado é doloso. O posicionamento para o presente trabalho é idêntico, uma vez que a regra geral no direito penal brasileiro é a responsabilidade penal por dolo, devendo a modalidade culposa ser expressamente prevista¹⁰⁰ (*vide* 2.3.2.).

Cumpre, aqui, fazer um esclarecimento: a despeito de nosso posicionamento neste trabalho ser de que o crime previsto no art. 4º da Lei Caó é próprio (demandando uma característica especial

¹⁰⁰ A simplicidade na análise doutrinária apresentada em relação à Lei Caó se deu principalmente em face da dificuldade de encontrar doutrina que se aprofundasse acerca do tema. Além disso, não se desconhece a necessidade de aplicação das normas penais antirracistas, como defendido por Machado *et. al.* (2015), ou mesmo do tratamento desigual que grupos minoritários podem sofrer, sendo alvos preferenciais da tutela penal e sujeitos cujos bens jurídicos não são tão bem protegidos, o que foi descrito por Zaffaroni (2004) como culpabilidade por vulnerabilidade.

do autor para sua caracterização), este crime pode ser praticado, ao menos em tese, por programadores e outras pessoas envolvidas em empresas especializadas em promover ou contratar empregados. Nem a digitalização, nem a terceirização de determinadas atividades fazem com que não haja pessoas responsáveis, mas, tão somente que a responsabilidade pela contratação seja inteiramente terceirizada ou, ainda, dividida entre a empresa contratante e a empresa de tecnologia contratada para fins de contratação de mão se obra. Ou seja, a migração da eventual prática de racismo para um ambiente digital não descaracteriza sua prática ou impossibilita a responsabilização dos envolvidos. Sobre este tema, passa-se a apresentar o fenômeno conhecido como racismo algorítmico, que busca explicar como estes podem ser vetores de discriminação racial.

3.1.3. O Racismo algorítmico e a discriminação algorítmica: as dificuldades e dilemas de um mundo tecnológico

Tarcizio Silva (2020a, p. 122) aponta a existência de um fenômeno denominado “plataformização” da internet¹⁰¹, que consiste na percepção de que, “gradualmente, ambientes digitais como Facebook estenderam-se para toda a web em uma espécie de integração que priorizou a concentração de dados e valor – inclusive financeiro – em poucas empresas”. Segundo o autor, neste contexto de plataformização,

[...] o impulso do capital financeiro permitiu que grupos como Alphabet (detentora do Google), Facebook, Amazon, Microsoft e Apple investissem no desenvolvimento estratégico de suas plataformas, incluindo a construção de conglomerados de cientistas de variadas áreas para contínua análise e otimização dos fluxos de monetização da audiência, publicidade e comportamento de seus usuários [...]. (SILVA, 2020a, p. 122)

Como consequência dessa plataformização, o autor aponta a “concentração internacional de fluxos de dados e de capital” (SILVA, 2020a, p. 123). Entre esses agentes que concentram esses fluxos, podemos colocar empresas como a rede social com fins profissionais LinkedIn¹⁰² como plataforma voltada ao recrutamento e seleção de emprego. Ou seja, trata-se não apenas de um site ou aplicativo, mas uma verdadeira rede social com exposição pública dos currículos e interação dos candidatos com empresas para se colocar de forma competitiva no mercado de trabalho atual.

¹⁰¹ No mesmo sentido aponta SRNICEK (2017, *passim*).

¹⁰² O LinkedIn é a maior rede social para fins profissionais do mundo, tendo chegado a mais de 700 milhões de usuários pelo planeta, segundo o portal Finances Online. Disponível em: “<https://financesonline.com/number-of-linkedin-users/>”. Acesso em 31 de outubro de 2022.

No Brasil, a rede social possui mais de 50 milhões de usuários, segundo noticiado por Olhar Digital. Disponível em: “<https://olhardigital.com.br/2021/08/16/internet-e-redes-sociais/linkedin-bate-a-marca-50-milhoes-de-usuarios-no-brasil/>”. Acesso em 31 de outubro de 2022.

Para além da plataformização, amplia-se a digitalização e a algoritmização de condutas antes seriam tomadas/decididas por humanas, surgindo, em decorrência disso, o fenômeno da dupla opacidade, que se caracteriza pelo “caráter difuso tanto da tecnologia, vista erroneamente como neutra, quanto das relações étnico-raciais na sociedade e, por consequente, na tecnologia” (SILVA, 2020b, p. 444-445).

Neste cenário supostamente neutro em face de questões étnico-raciais, o racismo algorítmico pode ser definido como uma manifestação do racismo estrutural no ambiente digital e algorítmico. Sobre este fenômeno, Bruna Lima afirma que:

O termo algoritmo é fator adjetivo da forma atual de praticar racismo atrelado não apenas aspectos éticos relacionados ao uso das tecnologias de informação e muito menos como algo independente criado pelo desenvolvimento da inteligência das máquinas: mas constitui um fenômeno sociotécnico de práticas de violência racial. (LIMA, 2022, p. 37)

Sobre as manifestações do racismo algorítmico, Silva (2020b, p. 130-134) apresenta diversos exemplos deste fenômeno, que vão desde exibição de anúncios de crime a pessoas negras, hipersexualização de resultados de imagens de garotas negras na busca do Google Imagens, marcação de pessoas negras como gorilas no Google Photos. Esses são apenas alguns exemplos das formas como o racismo pode ocorrer com uso de IA, o que, inclusive, independe da intenção de pessoas de praticar atos racistas. Neste sentido, volta-se à ideia do racismo como fenômeno estrutural, de modo que o racismo pode prescindir de um racista individualizado, operando por meio das concepções entranhadas nas estruturas sociais vigentes.

Duarte e Negócio (2021, p. 220) também apontam que algoritmos não são neutros, uma vez que estes carregam em seus dados “crenças, vieses e equívocos humanos, pois as escolhas que levam à construção desse sistema são feitas por humanos falíveis”.

Assim, o racismo algorítmico, se insere em um fenômeno mais amplo de discriminação algorítmica, conceito que, nas palavras de Laura Schertel Mendes e Marcela Mattiuzzo (2019, p. 51) serve para designar “tanto cenários que envolvem afirmações estatisticamente inconsistentes quanto cenários em que as afirmações, embora estatisticamente lógicas, de alguma forma tomam os indivíduos que dela são objeto não de forma efetivamente individualizada, mas apenas como parte de um grupo”.

Segundo as autoras, este fenômeno pode se dividir em quatro categorias diversas. A primeira delas é a *discriminação por erro estatístico*, que ocorre quando há falha na contabilização correta dos dados, sendo “o tipo de discriminação que decorre de um erro cometido pelos engenheiros ou cientistas de dados responsáveis pelo desenho do algoritmo” (MENDES; MATTIUZZO, 2019, p. 51). O segundo grupo é o da *discriminação por generalização*, é um modelo que “embora

estatisticamente correto, “leva a uma situação na qual algumas pessoas são equivocadamente classificadas em certos grupos” (MENDES; MATTIUZZO, 2019, p. 51). Neste sentido, as autoras apresentam os critérios para que haja a discriminação por generalização: “embora o algoritmo esteja correto e as informações também, ainda assim o resultado será uma generalização incorreta, na medida em que mesmo um resultado estatisticamente relevante apresentará um percentual de pessoas que não se encaixam perfeitamente naquela média” (MENDES; MATTIUZZO, 2019, p. 52).

A terceira forma é a *discriminação pelo uso de informações sensíveis*, que se dá quando uma análise, ainda que possivelmente correta de uma perspectiva estatística, aplica *proxies*¹⁰³ e dados protegidos por lei, de modo que esta forma de discriminação possui dois requisitos constitutivos: “além de utilizar dados sensíveis, a classificação deve se basear em características endógenas¹⁰⁴ ou, então, deve destacar grupos historicamente discriminados” (MENDES; MATTIUZZO, 2019, p. 52). A quarta categoria, por sua vez é a discriminação que limita o exercício de direitos (MENDES; MATTIUZZO, 2019, p. 52-53)¹⁰⁵.

Neste sentido, o uso de determinadas características pode ampliar a discriminação de grupos socialmente desfavorecidos, como é o caso dos negros. Neste sentido:

Outro problema que levanta reflexões em matéria de igualdade surge do tratamento diferenciado baseado em características discriminatórias e estereotipadas clássicas, como nacionalidade, gênero, idade ou identidade sexual (terceira categoria). Por um lado, essas características estão intimamente relacionadas ao cerne da personalidade de cada um, mostrando-se, portanto, invariáveis. De outro, esses atributos representam diferenças históricas de tratamento e de estereotipização de grupos. Ademais, sua utilização como base para processos decisórios pode trazer efeitos colaterais, isto é, a discriminação de certos grupos na sociedade pode ser intensificada. (MENDES; MATTIUZZO, 2019, p. 52-53).

¹⁰³ Neste contexto algorítmico, *proxies* são dados que se relacionam diretamente com algum outro dado. Desta forma, ao usar uma *proxie* para raça ou gênero, por exemplo, eu posso não fazer referência explícita a uma modalidade discriminatória e, ainda assim, gerar esta discriminação.

¹⁰⁴ As variáveis endógenas na discriminação são aquelas que, por sua reprodução sociais contínua geram a distinção suposta. Ou seja, trata-se não de uma realidade existente, mas de uma discriminação que cria uma realidade e, conseqüentemente, confirma-se a hipótese inicial. Neste sentido: “Utilizando o exemplo anterior do mercado de trabalho para esclarecer essa distinção, é possível afirmar que as mulheres foram, historicamente, mais envolvidas na criação dos filhos e em tarefas domésticas do que os homens. Mas esse é um resultado do fato de que a elas foram dadas menos oportunidades profissionais, consideradas por muito tempo como incompatíveis com as tarefas domésticas, e não uma característica inerente do sexo feminino que torna as mulheres menos capacitadas ou menos interessadas em oportunidades profissionais. A consequência da discriminação, nesse caso, leva à confirmação da hipótese inicial, pois estamos analisando uma variável endógena.” (MENDES; MATTIUZZO, 2019, p. 51)

¹⁰⁵ Não se desconhece que a operação de algoritmos se dá por meio de discriminação de características encontradas nos dados disponíveis entre as pessoas, mas o uso com intenção ou conseqüências discriminatórias desse tratamento de dados em face de grupos socialmente desfavorecidos pode gerar essa forma indesejada de discriminação. Neste contexto, “na medida em que os algoritmos se baseiam, em grande parte, em discriminação estatística, isto é, na diferenciação de indivíduos baseada nas características de um grupo e na probabilidade de tal grupo agir de determinada maneira, torna-se indispensável compreender se os processos e critérios utilizados para classificar indivíduos são corretos, transparentes e, em última instância, justos.” (MENDES; MATTIUZZO, 2019, p. 41)

Assim, como muitas das vezes a discriminação existente não é intencional – ou não há prova de intencionalidade –, as ferramentas jurídicas atuais de combate ao racismo passam a ser insuficientes, uma vez que carregam “a ideia de intencionalidade na conduta discriminatória” e não compreendem “adequadamente as estruturas sociais e a semântica da discriminação que premia os constructos sociais e técnicos” (DUARTE; NEGÓCIO, 2021, p. 220).

No cenário estadunidense, Barocas e Selbst descrevem os impactos da discriminação algorítmica no mercado de trabalho. Segundo os autores, muito da desigualdade existente decorre da discriminação estrutural que afeta os tomadores de decisões, de modo que o processamento de dados pode reproduzir os padrões de discriminação existentes (BAROCAS; SELBST, 2016, p. 673-676). Dessa forma, os autores explicam a forma com que algoritmos discriminam pessoas:

Por definição, o processamento de dados [*data mining*] é *sempre* uma forma de discriminação estatística (e, portanto, aparentemente racional). De fato, o próprio objetivo do processamento de dados é fornecer uma base racional sobre a qual distinguir entre indivíduos e conferir ao indivíduo de forma confiável as qualidades possuídas por aqueles que parecem estatisticamente semelhantes. No entanto, o processamento de dados tem o potencial de descontar indevidamente membros de classes legalmente protegidas e colocá-los em desvantagem relativa sistemática. Ao contrário de formas mais subjetivas de tomada de decisão, os efeitos nocivos do processamento de dados geralmente não são rastreáveis ao viés humano, consciente ou inconsciente¹⁰⁶. (BAROCAS; SELBST, 2016, p. 677)

Em certa medida, portanto, a questão central é entender que algoritmos funcionam fazendo discriminação entre pessoas, colocando-as em determinados grupos cujas características serão deduzidas, mas, limitar que a forma com a qual os grupos e dados são selecionados não gerem/reproduzam as discriminações indesejadas e socialmente já existentes, como é o caso do racismo.

Neste sentido, decidir uma contratação ou determinar a confiabilidade do crédito de uma determinada pessoa não possui uma resposta tão simples quanto se um e-mail é ou não *spam*, em que é possível obter critérios e respostas objetivas (BAROCAS; SELBST, 2016, p. 679). Em decorrência disso, Barocas e Selbst (2016, p. 678) apontam que a etapa inicial do processamento de dados consiste em entender a demanda empresarial e convertê-la em um problema de processamento de dados [*data mining problem definition*], momento em que os programadores podem, ainda que sem qualquer

¹⁰⁶ Tradução nossa. No original: “By definition, data mining is *always* a form of statistical (and therefore seemingly rational) discrimination. Indeed, the very point of data mining is to provide a rational basis upon which to distinguish between individuals and to reliably confer to the individual the qualities possessed by those who seem statistically similar. Nevertheless, data mining holds the potential to unduly discount members of legally protected classes and to place them at systematic relative disadvantage. Unlike more subjective forms of decision making, data mining's ill effects are often not traceable to human bias, conscious or unconscious”

intenção, “descrever o problema de uma forma que possa sistematicamente prejudicar grupos sociais protegidos¹⁰⁷”.

Neste sentido, vários fatores podem gerar discriminação indesejada na operação de algoritmos. O primeiro desses exemplos é a seleção dos dados utilizados para a operação das formas de IA/algoritmos, uma vez que (i) a seleção dos dados históricos como parâmetro pode reproduzir preconceitos e discriminações aplicadas no passado¹⁰⁸, ou que (ii) a seleção de amostras enviesadas (com sub-representação de grupos, por exemplo), pode gerar desvantagens a grupos sub ou super-representados nos dados¹⁰⁹ (BAROCAS; SELBST, 2016, p. 681).

Outra situação potencialmente problemática sob a perspectiva da discriminação é o modo pelo qual serão selecionadas as ferramentas. Considerando que é muito caro estabelecer modelos mais precisos para obter as características desejadas na análise de dados (ou mesmo obter os dados necessários a esta análise), muitas vezes recorre-se a distinções de determinados grupos, e não a características pessoais. Um exemplo desse fenômeno é o uso das credenciais acadêmicas com grande peso à universidade que o candidato ao emprego frequentou, que não é um dado que determina as capacidades de trabalho de um indivíduo, mas tão somente caracteriza-o dentro de um grupo (BAROCAS; SELBST, 2016, p. 689). Existe, ainda, o uso de características de grupos protegidos – como é o caso da categoria raça – que configura o racismo racional (*rational racism*) (BAROCAS; SELBST, 2016, p. 690). Uma prática que também pode existir é o mascaramento (*masking*) que consiste em criar um algoritmo com um erro aparentemente indesejado com resultado discriminatório, quando, na verdade, o real objetivo era inserir aquela forma de discriminação (BAROCAS; SELBST, 2016, p. 692-693).

No sentido de combater a discriminação no mercado de trabalho, o Title VII do Civil Rights Act estadunidense apresenta duas teorias de responsabilização dos empregadores por discriminação de empregados: o de *tratamento desigual* (*disparate treatment*) e de *impacto desigual* (*disparate impact*) (BAROCAS; SELBST, 2016, p. 694). O *tratamento desigual*, ocorre quando há discriminação intencional mascarada ou não (p. 698-700). Um debate importante em relação a esta

¹⁰⁷ Tradução nossa. No original: “parse the problem in such a way that happens to systematically disadvantage protected classes”.

¹⁰⁸ Como apontam os autores, um exemplo de como os dados históricos podem reproduzir discriminações historicamente praticadas é o da ferramenta de recrutamento passivo Talent Match da LinkedIn, que podem reproduzir a discriminação dos empregadores em face de grupos vulneráveis. Segundo eles: “In particular, if LinkedIn’s algorithm observes that employers disfavor certain candidates who are members of a protected class, Talent Match may decrease the rate at which it recommends these candidates to employers. The recommendation engine would learn to cater to the prejudicial preferences of employers.” (BAROCAS; SELBST, 2016, p. 683)

¹⁰⁹ Neste sentido, Buolamwini e Gebru (2018, *passim*.) apontam dificuldades de reconhecimento facial de pessoas negras ante a existência de menor concentração de pessoas negras nas bases de dados desses algoritmos, sendo que esse grupo possui menor confiabilidade no reconhecimento realizado.

modalidade discriminatória é se dá quando o empregador que descobre o padrão discriminatório não-intencional e não faz nada para alterá-lo. Segundo Barocas e Selbst (2016, p. 700), haveria dificuldades de comprovar a intenção de discriminar pelo mero conhecimento de resultado discriminatório não intencional do mecanismo de seleção, sendo que o conhecimento poderia servir tão somente um elemento a indicar a discriminação intencional e não sua comprovação¹¹⁰.

Já em relação ao *impacto desigual*, este ocorre quando uma prática aparentemente neutra causa impacto diferenciado em um grupo protegido. O grande critério orientador de sua existência ou não na jurisprudência estadunidense é o da necessidade (*business necessity*), segundo o qual os testes/critérios de seleção empregados devem ser relacionados ao trabalho. A despeito de variação na abrangência desse critério na jurisprudência estadunidense, ficou estabelecido que práticas não estão relacionadas com a performance no trabalho devem ser excluídas. No entanto, segundo os autores, uma interpretação abrangente deste limitador, conjugado com uma lógica de que, na dúvida, presume-se legítima a diferenciação criada faz com que seja difícil questionar a ilegitimidade de determinados mecanismos de seleção de emprego (BAROCAS; SELBST, 2016, p. 701-712).

Entretanto, partindo-se da premissa que algoritmos e IA são ferramentas, e não uma forma tecnológica em si perversa, uma vez compreendida adequadamente a forma de operação da discriminação algorítmica, os algoritmos podem ser usados para “mitigar (para não dizer solucionar) esses mesmos problemas, por meio da disponibilização de ferramentas que outrora não existiam e que permitem uma melhor compreensão do fenômeno discriminatório” (DUARTE; NEGÓCIO, 2021, p. 239).

Barocas e Selbst, apontam alguns caminhos que podem ajudar a reduzir a discriminação algorítmica no mercado de trabalho. Inicialmente, defendem a necessidade de descrever o problema de programação de forma que realmente decisões de contratação sejam definidas por critérios pertinentes à realização do trabalho objeto da seleção, bem como que a seleção de dados não pode se amparar em discriminações passadas, evitando sua reprodução. Contudo, segundo os autores, a datificação das seleções de recrutamento com dados de quem pertence a cada grupo minoritário pode possibilitar a monitoração das evoluções ou involuções na busca de inclusão destes grupos, tornando mais fácil verificar a inclusão ou a exclusão de grupos por meio de seleções de emprego com aplicação de IA (BAROCAS; SELBST, 2016, p. 714-720).

Acerca das ferramentas de IA de seleção e recrutamento, aponta-se a necessidade de evitar mecanismos que, embora eficientes, sejam discriminatórios. Assim, uma escolha que pode ser feita é

¹¹⁰ Na análise de casos (3.2), debateremos melhor o que pensamos em relação a este conhecimento e a possível configuração do crime de racismo no direito brasileiro.

de “em vez de reduzir a taxa de erro para aqueles em classes protegidas, os programadores/cientistas de dados podem estruturar suas análises para minimizar a diferença nas taxas de erro entre os grupos¹¹¹” (BAROCAS; SELBST, 2016, p. 720). Ou seja, uma solução possível é que os algoritmos sejam igualmente disfuncionais entre grupos diversos de modo que, a despeito da redução de sua eficácia global, não deixam em situação mais vulnerável os grupos sociais já mais vulneráveis.

Especificamente sobre a regulação brasileira da discriminação algorítmica, Bruna Lima (2022, p. 115) afirma que “a lei brasileira atual de proteção de dados não é suficiente para combater o racismo algorítmico, pois optou por não tratar especificamente o uso de algoritmos” e seus possíveis vieses discriminatórios. Assim, para a autora, a solução para o racismo algorítmico passa pela existência de:

[...] previsão legislativa que atribua uma responsabilidade objetiva no sentido de garantir a efetivação da reparação ao indivíduo que sofreu o dano (reparação integral). Considerando a capacidade de aprendizagem e autonomia de sistemas de IA de realizar decisões quase impossíveis de serem explicadas por seus desenvolvedores – abarcados pela “imprevisibilidade” -, incorrer a quem sofreu o dano que na maioria das vezes sem conhecimento técnico explique a forma (tipo específico de falha e que o cometeu) e extensão do dano é desproporcional. (LIMA, 2022, p. 117)

Como afirmado anteriormente (2.1), o Marco Regulatório da IA no Brasil é uma proposta legislativa em tramitação que pode resolver esse problema. No entanto, a análise de caso proposta se dá em face da legislação vigente, de modo que terá que lidar com as dificuldades e insuficiências existentes no cenário regulatório atual, especialmente no âmbito do direito penal, razão pela qual parte-se para a análise de casos que, dentro do que as limitações que a realidade e o padrão atual de conhecimento sobre as estruturas internas de produção de IA nos permitem.

3.2. ESTUDO DE CASOS

A seguir, passa-se à análise dos casos, tentando compreender, em cada situação específica, se (e como) seria possível a configuração do crime previsto no art. 4º da Lei Caó. No entanto, é necessário levar para a análise dos casos que seleção e recrutamento são processos complexos, com várias etapas, passando pelo anúncio da vaga, a filtragem dos currículos, a entrevista e, somente após essas etapas, a seleção final (esta última geralmente feita por um humano). Em cada uma dessas etapas é possível que haja pessoas (e até mesmo empresas) distintas cuidando de cada uma das etapas. Além disso, muitas das vezes, essas pessoas ou empresas não irão compartilhar as informações que possuem

¹¹¹ Tradução nossa. No original: “rather than reducing the error rate for those in protected classes, data miners could structure their analyses to minimize the difference in error rates between groups”.

com os demais agentes, de modo que saber o que é de conhecimento de cada um desses agentes é também um desafio que deve ser enfrentado de forma individualizada.

Apresentados esses esclarecimentos, podemos iniciar a análise dos casos.

3.2.1. Caso 1: formas tradicionais de contratação e de racismo

O primeiro caso proposto é um exemplo de seleção e recrutamento clássicos, sem o emprego de tecnologias de inteligência artificial e por meio da tomada de decisões humanas. Sendo assim, o Caso 1 foi estruturado da seguinte forma: Lucas, chefe de RH da Empresa Alfa, deixa de contratar José por este último ser negro. Esse é o caso clássico de discriminação direta para o qual a legislação penal de combate ao racismo foi pensada.

Não parece haver qualquer dúvida de que a discriminação racial deliberada no ambiente de trabalho configure o apontado crime, sendo o exemplo que configura sua definição literal de “negar ou obstar emprego em empresa privada” (BRASIL, 1989, n.p.). Dessa forma, esta atitude adequa-se perfeitamente ao tipo penal, deixando evidente a prática direta de racismo cometido por Lucas contra José. No entanto, e não se pode desconsiderar essas complexidades da vida real, a configuração do delito é clara somente quando há comprovação de que os elementos do tipo foram comprovados, inclusive, o dolo (elemento subjetivo).

Assim, mesmo um caso simples como esse vai esbarrar no problema de que, como afirma Silva e Sfoggia (1997, p. 17-18), dificilmente um empregador racista vai afirmar categoricamente que deixou de contratar um empregado negro *em razão* de sua “raça”, sendo o padrão de discriminação racial vigente de preponderância de sua modalidade indireta. Ou seja, como um exemplo de livro pode ser suficiente dizer que seria possível a configuração desse crime, mas enquanto dado da realidade – que deve ser comprovado num ambiente com todas as formalidades inerentes ao Processo Penal – há grandes dificuldades para fazer prova da intencionalidade na discriminação.

3.2.2. Caso 2: a filtragem de currículos por meio de IA

No Caso 2, iniciam-se as modalidades algorítmicas de seleção e recrutamento, não sendo a seleção feita por uma pessoa individualizada (como Lucas, no Caso 1). O cenário hipotético apresentado é o seguinte: Lucas, chefe de RH da Empresa Alfa, contrata a Empresa Beta para filtrar somente os 3% melhores currículos, que serão chamados para entrevistas, sendo que nem José nem

qualquer outra pessoa negra são chamadas para as entrevistas¹¹². Neste e nos demais casos analisados e nas correspondentes variações há a figura do programador Marcos – que não toma as decisões estratégicas, mas elabora os códigos pelos quais os algoritmos dotados de aprendizagem de máquina funcionam – e do operador¹¹³ Erick – que analisa os dados dos algoritmos envolvidos e que pode suspender ou iniciar sua execução e requerer alterações ao programador.

Das fases descritas no processo de seleção de emprego, o mecanismo de IA aplicado neste caso específico é a filtragem de currículos (terceirizada à empresa de tecnologia Beta) que, por uma razão desconhecida, excluiu os candidatos negros em sua integralidade. É difícil esclarecer exatamente qual seria a razão para a existência desta discriminação, mas é provável que o erro tenha ocorrido pelo trabalho do programador Marcos, seja na seleção da base de dados empregada para a análise do desempenho dos funcionários, seja pelo mecanismo utilizado em si – que passou a discriminar os candidatos negros em sua integralidade.

Sob a perspectiva da tipicidade objetiva, o crime de racismo parece se configurar, uma vez que foi negado (ou ao menos obstado) o acesso à contratação de todas as pessoas negras envolvidas na seleção. Os sujeitos passivos (vítimas) desse crime seriam as pessoas negras não selecionadas para as fases seguintes do processo de recrutamento, grupo composto, entre outros, pelo candidato José. No entanto, é sob a perspectiva da tipicidade subjetiva e em relação a quais seriam os sujeitos ativos que respostas são mais difíceis de se obter, uma vez que a necessidade de dolo demanda que o agente responsável tenha conhecimento e vontade de realizar o resultado discriminatório¹¹⁴.

Nesse caso, o programador Marcos somente poderia ser responsabilizado caso houvesse a compreensão de que este intencionalmente selecionou a base de dados ou programou o algoritmo de forma a excluir os candidatos negros. Considerando-se o cenário de algoritmos com *machine learning*, ainda é necessário comprovar que a programação discriminatória decorreu dos comandos de programação e/ou aprendizado do programador Marcos, e não por um caminho indesejado e não programado tomado pela própria aprendizagem da máquina, pois aqui o controle do curso causal (*domínio do fato*) não seria mais do programador. As possibilidades descritas são de prova difícil – se possível, a depender das possibilidades de acesso aos dados e informações internas das empresas –, sendo improvável a configuração prática da responsabilidade penal.

¹¹² Esse caso conta com uma extrapolação. Via de regra, a discriminação algorítmica não impede o acesso de todos os negros ao prosseguimento para as etapas seguintes da seleção de emprego, no entanto, para que a situação fique mais evidente e torne mais fácil o debate, a hipótese inicial do Caso 2 parte de uma possibilidade de exclusão absoluta, e não de uma exclusão estatística em que, na média, pessoas negras são mais prejudicadas que as não negras. As variações desses casos poderão trazer a lógica do caso para mais perto de possibilidades realistas.

¹¹³ Para o presente trabalho, o operador é quem controla a atuação dos algoritmos, inserindo os dados dos candidatos para a seleção: é a pessoa que aperta o botão para iniciar a seleção e pode, também, apertar o botão para que ela pare.

¹¹⁴ Sobre a constituição do elemento subjetivo, o tema foi tratado anteriormente (2.3.2).

Observando-se a figura do operador Erick (também integrante da empresa de tecnologia Beta), a situação é bem semelhante à do programador Marcos: somente em havendo elementos que comprovem que havia conhecimento de que negros estariam sendo discriminados e preteridos nas seleções de emprego cuja seleção é feita pela IA de seleção da empresa Beta, seria pensável a configuração da tipicidade subjetiva (dolo), possibilitando eventual responsabilidade penal pelo crime previsto no art. 4º da Lei Caó. Neste caso específico, seria mais remota a possibilidade de responsabilização penal em relação a Lucas (chefe de RH da empresa que terceirizou a primeira etapa da contratação), uma vez que ele não teria participado de qualquer ato que gerou a exclusão de negros de processo seletivo.

Dessa forma, mesmo nesse caso em que todas as pessoas negras teriam sido excluídas da seleção de emprego, haveria grandes dificuldades teóricas e práticas para que se configure o crime de racismo. No entanto, após a apresentação das variações deste caso, será debatida uma possibilidade adicional de configuração de vínculo subjetivo que, possivelmente, permitirá uma ampliação dos casos em que será possível haver a configuração do crime de racismo em face da operação de algoritmos.

Na primeira variação desse caso, ao invés de nenhuma pessoa negra, são chamadas somente 3 pessoas negras dentro de 100 entrevistados, metade da quantidade habitual, de acordo com o histórico de seleções anteriores, sendo que o candidato José foi excluído nessa etapa. Ao final das entrevistas, nenhuma pessoa negra é contratada por Lucas.

Nesse caso, ainda que muito semelhante à hipótese inicial, é ainda mais difícil o estabelecimento de responsabilidade penal, pois a redução do percentual de pessoas negras não necessariamente tem relação com um critério discriminatório, menos ainda deliberadamente e intencionalmente discriminatório, de modo que a situação do programador Marcos e do operador Erick seriam juridicamente mais simples, havendo maiores obstáculos ainda maiores ao estabelecimento de responsabilidade penal desses agentes.

Uma última situação a ser analisada é a do Chefe de RH Lucas, que não selecionou nenhuma pessoa negra após as entrevistas. Mesmo aqui, haveria dificuldades para o estabelecimento de responsabilidade penal, uma vez que a proporção de pessoas negras na etapa de entrevistas foi bem menor, no entanto, aqui sua responsabilização já parece, em tese, possível, diferentemente dos contornos originais do Caso 2, dependendo da comprovação da intencionalidade de discriminar por razões de raça.

Nesta 1ª variação, as eventuais vítimas do crime de racismo seriam, também, as pessoas negras que participaram da seleção. No entanto, somente podem ser vítimas pessoas afastadas na etapa em que eventual autor a tenha excluído. Ou seja, José, que foi excluído na etapa de análise de currículos

por IA, pode ter sido vítima do crime de racismo praticado pelo programador Marcos ou pelo operador Érick, mas não pelo chefe de RH Lucas. Por outro lado, os candidatos entrevistados só podem ser vítimas do crime de racismo praticado por Lucas, em contexto semelhante ao Caso 1, de racismo “tradicional”, e não de sua variante algorítmica.

Na segunda variação do Caso 2, são chamadas 10 pessoas negras (incluindo-se José) dentro de 100 entrevistados, mais do que a quantidade habitual, de acordo com o histórico de seleções anteriores. Ao final das entrevistas, nenhuma pessoa negra é contratada. Nesse caso, a responsabilização do programador Marcos e do operador Erick parece mais remota do que no caso original e na primeira variação, uma vez que, por meio do emprego de IA, houve aumento de participação de pessoas negras no histórico de contratação da empresa. No entanto, mesmo assim, somente 10% das pessoas chamadas para a entrevista são negras, havendo a possibilidade de algum viés discriminatório ser determinante para esse baixo (ainda que maior) percentual de seleção de pessoas negras.

Por outro lado, pode haver responsabilidade penal em relação a Lucas, uma vez que este não selecionou nenhuma pessoa negra para as vagas, ainda que tenha havido maior percentual na seleção. Sua situação seria semelhante à do Caso 1, havendo necessidade de comprovação da intenção discriminatória, de modo que, se este último praticou racismo, é uma prática de racismo comum e não em sua modalidade algorítmica.

Mais uma vez, as dificuldades de configuração do crime de racismo passam por haver legislação que busca combater o racismo como discriminação direta e, por outro lado uma realidade de reprodução de um racismo estrutural, e não direto. E mais, mesmo nos casos de racismo direto, ninguém assumirá que o motivo para a exclusão do candidato para o emprego será de ordem racial – sendo comum, no contexto algorítmico a prática do mascaramento, que, pelo uso de *proxies*, os reais elementos usados para selecionar e classificar pessoas não são evidenciados.

Como vimos anteriormente, o caminho mais adequado para combater a discriminação algorítmica no mercado de trabalho é por meio de aplicação de critérios de impacto desigual, e não de tratamento desigual¹¹⁵. No entanto, a legislação penal de combate ao racismo não parece possuir mecanismos de combate ao impacto desigual – provavelmente mais afetado por um racismo estrutural que por formas mais explícitas de discriminação racial –, de modo que são diversas as dificuldades de se estabelecer responsabilidade penal relacionada a criação ou operação de IA de seleção e recrutamento. Assim, as dificuldades de configuração do crime de racismo – especialmente pela ausência do dolo enquanto elemento subjetivo do tipo penal – partem da premissa de que o racismo

¹¹⁵ Tema abordado anteriormente no tópico 3.1.3.

algoritmo não é intencional e, ainda, que os agentes envolvidos na operação e controle dessas formas de IA não conhecem seus efeitos discriminatórios.

Partindo-se da hipótese de que pessoas envolvidas nesses mecanismos de recrutamento e seleção conhecem os efeitos discriminatórios dessas ferramentas tecnológicas, há possibilidades razoáveis de configuração do dolo, ainda que em sua modalidade eventual. Sendo o dolo composto de consciência e vontade¹¹⁶, saber que há um padrão discriminatório sendo criado e/ou repetido pela IA de seleção e recrutamento faz com que uma parte do elemento subjetivo esteja satisfeito. Uma vez configurada a consciência do mecanismo de racismo algorítmico, a questão, aqui é compreender se existe a vontade. Para a configuração do dolo direto, seria necessário enfrentar dificuldades no sentido de superar a comprovação de intencionalidade do agente, o que esbarra nas dificuldades existentes na legislação vigente de combate ao racismo, e, ainda, nas dificuldades de acessar e compreender a operação de algoritmos.

No caso do dolo eventual, por sua vez, a situação é menos restritiva, uma vez que, conhecendo o funcionamento discriminatório das formas de IA, ainda que não deseje o resultado, é provável que o autor tenha consentido com sua realização, configurando-se essa modalidade de dolo. Um questionamento possível neste momento é – considerando-se a tênue diferenciação entre dolo eventual e culpa consciente – se o prosseguimento de IA de processo seleção e recrutamento sabidamente discriminatório não poderia ocorrer como forma de culpa consciente. No entanto, esta situação é amplamente improvável em decorrência de uma questão específica: como as formas de IA de seleção e recrutamento tendem a repetir as operações, manter um processo ou procedimento sabidamente discriminatório possui ampla probabilidade de novo resultada discriminatório, de modo que este eventual resultado pode não ser, necessariamente, desejado, mas a manutenção da mesma forma de IA após o conhecimento das práticas discriminatórias indica o consentimento do risco de materialização no resultado – sendo possível a caracterização do dolo eventual nesses casos.

Neste sentido, e retomando-se as ideias de Luzon Peña (2020, p. 55), para afastamento do consentimento ou aceitação em relação ao dolo eventual, o agente deve possuir “*uma confiança minimamente fundada objetivamente*” para acreditar que o resultado criminoso não ocorrerá. No entanto, e considerando-se que algoritmos tendem a repetir de forma idêntica ou semelhante suas operações, há dificuldades de configurar esses casos em que há crença razoável e justificada de que o resultado discriminatório não ocorrerá. Dessa forma, nos casos de prévio conhecimento do histórico discriminatório da ferramenta tecnológica empregada para seleção e recrutamento, defende-se a possibilidade de responsabilização penal subjetiva na modalidade de dolo eventual.

¹¹⁶ Vide 2.3.2.

Restaria compreender quem – dentre as pessoas envolvidas na operação da algoritmos – pode ser considerado responsável por este eventual delito. Neste sentido, e considerando-se que o autor é quem controla o curso causal (o acontecer típico), cada um poderá ser responsável por falha existente na etapa de sua participação. Assim, voltando-se à 1ª Variação do Caso 2, o gerente de RH Lucas é responsável por terceirizar a seleção de currículos da empresa Alfa para a empresa Beta. Por sua vez, o programador Marcos e o operador Erick, ambos da empresa Beta, respondem por sua participação na forma de IA. Em relação ao desenvolvedor Marcos, se este, após tomar conhecimento de falha de programação, não toma nenhuma atitude para corrigi-la, passa a ser possível pensar em sua responsabilidade penal. O operador Erick, por sua vez, ao tomar conhecimento de falha na IA e não cessar sua operação ou requerer ao programador que haja sua correção, pode cometer o crime do art. 4º da Lei Caó.

Em relação ao chefe de RH Lucas, este pode, também, ser responsabilizado em face da IA – e não somente de sua conduta, como já apresentado – na versão original do Caso 2. Considerando que este percebe que pessoas negras são excluídas sistematicamente pela IA de seleção e não altera o mecanismo de seleção, também aqui é o caso de pensar em sua responsabilidade penal.

No entanto, e a não ser que exista combinação intencional (coautoria) dos agentes, o conhecimento de um agente sobre a discriminação não implica os outros. Ou seja, se o programador Marcos não sabe da falha no algoritmo que ele mesmo criou, mas o controlador Erick conhece e não faz nada para que o erro seja corrigido, pode-se pensar na responsabilidade penal de Erick, mas não de Marcos. Por outro lado, se Erick – ciente da discriminação existente em decorrência de falha de programação ou de incorreta aprendizagem de máquina – informa ao programador Marcos a falha existente e requer que as correções sejam feitas, a partir desse momento, é o programador que passa a poder ser responsabilizado, afastando-se a responsabilidade penal do operador, que adotou as medidas necessárias, requerendo a correção ao responsável pela programação da IA. Cada agente é, portanto, responsável pelo curso causal que controla dentro de suas atribuições.

Por fim, há, ainda, um requisito para que seja possível a responsabilização penal no contexto do Caso 2: segundo a teoria da imputação objetiva¹¹⁷, uma ação é objetivamente típica quando cria ou amplia o risco para o bem jurídico, mas no caso de redução do risco para o bem jurídico não há tipicidade objetiva. Neste contexto, na 2ª Variação da Caso 2 – em que há efetivo aumento de pessoas negras selecionadas pelo algoritmo, em comparação com os dados das seleções anteriores, ainda que em padrão menor que o desejável – haveria uma redução de risco para o bem jurídico ou, somente

¹¹⁷ Vide 2.3.2.

uma criação de risco diferente de outra modalidade de risco existente, de acordo com a teoria da *imputação objetiva*?

Ainda que o tema seja complexo e demande considerações mais aprofundadas em trabalhos posteriores, entendemos que haveria, nesse caso, a criação de um risco, e não sua redução. Mesmo que outras formas de seleção possam representar riscos maiores, o padrão da comparação para a criação de riscos discriminatório é o cenário em que a discriminação não existe. Discriminar em grau menor que outras pessoas ou instituições ainda é discriminação e merece ser objeto da tutela penal correspondente.

Cumpra esclarecer que, em relação ao Caso 2, os resultados obtidos coincidiram parcialmente com a hipótese inicial de amplas dificuldades na responsabilização penal por IA de seleção e recrutamento. Especificamente quando os agentes envolvidos tiverem conhecimento do viés discriminatório dos algoritmos, será possível a responsabilização penal a título de dolo eventual, havendo uma possível abertura para a punição criminal de práticas de racismo algorítmico¹¹⁸. Os demais casos ficaram dentro do que previa a hipótese, havendo amplas restrições à aplicação da lei penal nesses casos.

3.2.3. Caso 3: os anúncios de emprego desiguais

O Caso 3 foi estruturado da seguinte forma: Lucas, chefe de RH da Empresa Alfa, contrata a Empresa Ômega para fazer anúncios de emprego para uma vaga aberta. Os anúncios são mostrados prioritariamente para pessoas brancas, chegando a 80% de pessoas brancas e 20% de negras, em um espectro amostral de 50% de brancos e negros. Em decorrência disso, José não é chamado para as entrevistas e são chamadas somente 3 pessoas negras dentro de 100 entrevistados, metade da quantidade habitual, de acordo com o histórico de seleções anteriores.

Em relação às possibilidades de responsabilização ou não, a situação é muito semelhante à Variação 1 do Caso 2. No entanto, em relação aos anúncios de emprego, em respeito à legalidade estrita que rege o direito penal (2.3.5), haveria o debate se mostrar anúncios de forma desigual para grupos distintos significaria *impedir* ou *obstar* o acesso de negros a empregos. Em relação a impedir, parece não haver grandes detalhes, pois o anúncio desigual não impede que pessoas negras se candidatem. Ainda em relação a obstar, não parece ser o caso, pois aqui, não haveria um ato de não

¹¹⁸ Em relação a esta possibilidade de responsabilização penal por dolo eventual em decorrência do racismo algorítmico, esta será possível em relação a outras ações discriminatórias e em relação a outros grupos sociais, desde que haja tipos penais criminalizando especificamente o ato discriminatório praticado e a descrição de como possíveis vítimas o grupo discriminado em cada caso, em respeito do princípio da *legalidade*.

divulgação, mas um não ato de divulgação, não havendo óbice pela ausência de ação em relação a determinadas pessoas.

Neste sentido, a posição aqui defendida é que esta discriminação específica não é compreendida pelo art. 4º da Lei Caó, por ausência de *reserva legal*, o que impede a responsabilização penal dos agentes envolvidos no anúncio de vaga no cenário apresentado neste Caso 3. Nesse caso, pode até haver vítimas, mas não há crime –segundo a legislação penal vigente no Brasil.

3.2.4. Caso 4: os algoritmos de recrutamento passivo

O Caso 4 descreve uma situação de recrutamento passivo, como é feito pela plataforma Talent Match do LinkedIn, com a seguinte situação: Lucas, chefe de RH da Empresa Alfa, contrata a Empresa Pi para fazer recrutamento passivo, abordando diretamente os candidatos que combinam com o perfil da vaga em sua plataforma digital de seleção e recrutamento. Em decorrência disso, nenhum candidato negro é contactado ou contratado, só havendo contato com o candidato (branco) que foi efetivamente contratado. Ninguém além do candidato selecionado sequer tomou conhecimento da vaga e o Chefe de RH Lucas entrevistou este único candidato e optou por contratá-lo.

Neste caso específico, comprovada a discriminação, haveria a possibilidade de pensar na prática do crime previsto no art. 4º da Lei Caó. No entanto, existem dificuldades práticas e probatórias que devem ser superadas para que seja possível essa análise. Algumas dessas questões são: (i) identificar potenciais vítimas, (ii) compreender com dados se a seleção é realmente discriminatória e quais critérios foram empregados, (iii) identificar se há intencionalidade na discriminação praticada.

Nos casos anteriores, a figura do candidato José funcionou como uma forma de individualizar a situação de candidatos discriminados. Neste caso, buscando-se responder ao primeiro questionamento, nem José nem qualquer outro candidato chegariam a tomar conhecimento da existência da seleção. Assim, como pediríamos identificar as vítimas de discriminação? Além dessa dificuldade prática de identificar as vítimas, há uma dificuldade probatória, como comprovar que alguém que sequer tomou conhecimento do processo de seleção teve seu direito acessar um emprego negado ou obstado?

Se em *International Brotherhood of Teamsters v. United States* a Suprema Corte estadunidense reconheceu que políticas discriminatórias sistemáticas que fazem os candidatos desistirem de concorrer ao emprego por saberem que não serão selecionados em virtude de discriminação recorrente faz com que eles possam demandar judicialmente para questionar a discriminação (UNITED STATES, 1977, n.p.), aqui a situação é diversa. No precedente mencionado,

os candidatos ficaram sabendo da existência do processo de recrutamento e não participaram pela certeza do tratamento discriminatório. Já no caso do recrutamento passivo, não vai haver nem mesmo o conhecimento da existência da seleção ou de eventual caráter discriminatório existente.

A segunda questão também demanda atenção. Como saber se um algoritmo de seleção que escolheu apenas um candidato, e não informou quem foram as outras pessoas avaliadas para a vaga discriminou ou não? A resposta a essa pergunta só será possível caso tenha havido registrado dos dados dos empregados não selecionados. Dessa forma, a datificação pode ajudar, desde que haja como acessar esses dados. De qualquer modo, os princípios do constitucionalismo digital relacionados a explicabilidade e democracia possuem papel fundamental para que seja possível avaliar qualquer elemento que indique existência ou não de discriminação.

A terceira questão – de identificar intencionalidade em prática eventualmente discriminatória – é difícil de ser respondida, uma vez que o direito penal brasileiro exige que haja verificação do elemento subjetivo – no caso dos crimes de racismo, o dolo. Em relação à caracterização do dolo, a solução neste caso ocorre de modo semelhante ao Caso 2. No caso de conhecimento das práticas discriminatórias, os agentes envolvidos podem ser responsabilizados por dolo eventual. Entretanto, seria remota a possibilidade de responsabilização do Chefe de RH Lucas, que tão somente gostou de um candidato pré-selecionado pela IA de recrutamento passivo.

Em relação a este último caso, ainda que a responsabilidade penal não seja um cenário impossível, sua aplicabilidade prática será reduzida, uma vez que há grandes dificuldades de conhecer os mecanismos de seleção, individualizar as vítimas e, ainda, comprovar a intenção discriminatória apta a constituir o dolo, ainda que em sua modalidade eventual.

3.3. CONSIDERAÇÕES FINAIS

Como se pôde observar, o racismo é um fenômeno discriminatório ainda muito presente na realidade brasileira, sendo, também, algo entranhado nas estruturas da nossa sociedade, o que a literatura especializada discute como racismo estrutural. Também no ambiente algorítmico, há ocorrência de discriminação racial, trazendo para o ambiente digital problemas do mundo físico. Neste contexto, existe o que se convencionou chamar de racismo algorítmico.

Ainda que haja a discriminação racial direta e explícita, como o caso do dono de bar e restaurante que impede o acesso ou expulsa o cliente negro afirmando expressamente a motivação de seus atos com falas explicitamente racistas, essa não é a regra. No contexto do mercado de trabalho, ainda que haja discriminação racial, os empregadores não fundamentam esta discriminação em seus reais motivos.

Essa situação gera problemas. Por um lado, a forma mais recorrente do racismo não é explícita, mas indireta, e (em alguns casos) até mesmo não intencional, funcionando como mera reprodução da lógica estrutural existente. Por outro lado, a legislação brasileira de combate ao racismo – ao menos as normas penais analisadas – possui previsões no sentido de combater a discriminação direta e exigem a intencionalidade dos agentes (dolo).

Entretanto, a discriminação algorítmica e o racismo algorítmico operam de formas distintas, muitas vezes sem o uso dos dados discriminatórios (mas pelo emprego de *proxies* que levarão ao mesmo resultado que o uso desses dados) e até mesmo sem a intencionalidade. O que muitas vezes ocorre é a reprodução de discriminações historicamente existente pela sua verificação nos dados que amparam as formas de IA ou, ainda, a aplicação de mecanismos de seleção de emprego que são mais eficientes nos grupos mais favorecidos, prejudicando determinados grupos, como é o caso dos negros.

Neste sentido, o combate ao racismo algorítmico passa pela seleção adequada dos dados que a IA usará para fazer a seleção – de modo que não sejam usados dados marcados por preconceitos preexistentes – e por mecanismos de operação de IA que, além de não se preocupar em discriminar, se preocupem diretamente em não discriminar.

Em relação aos casos analisados, no Caso 1 – de contratação tradicional feita diretamente por pessoas – é difícil caracterizar o crime previsto no art. 4º da Lei Caó, uma vez que há necessidade de comprovação do dolo em um contexto de discriminação racial que é, na maior parte das vezes, indireta e não necessariamente individualizada, tornando difíceis tanto sua percepção quanto sua comprovação.

No Caso 2, surge uma possibilidade que infirma parcialmente a hipótese inicial, que é a responsabilização penal nos agentes envolvidos no desenvolvimento ou operação de algoritmos a partir do momento em que estes tomam conhecimento/são avisados do caráter discriminatório nos algoritmos de seleção e recrutamento, hipótese em que a responsabilização penal pode ocorrer, considerando-se o dolo eventual dos agentes. Assim, há razoáveis possibilidades de criminalização das condutas dos agentes em relação ao art. 4º da Lei 7.716/1989, especificamente em relação ao racismo algorítmico.

Em relação aos Casos 3 e 4, as dificuldades de criminalização das condutas residem nos limites do princípio da *legalidade*, que tem que prever expressamente a modalidade discriminatória existente, havendo dúvidas se os contextos discriminatórios apresentados em ambos os casos estão no alcance da norma penal incriminatória. Além disso, estes casos também esbarram nas dificuldades probatórias em relação ao dolo, de modo semelhante ao que foi apresentado no Caso 1.

Assim, em relação ao estudo dos casos, a hipótese foi parcialmente comprovada, no sentido de que há grandes dificuldades da configuração do crime previsto no art. 4º da Lei Caó em relação ao

racismo algorítmico ocorrido em processos de seleção e recrutamento. Por outro lado, na medida em que a possibilidade do reconhecimento do dolo eventual de programadores e controladores que possuem consciência do caráter discriminatório dos algoritmos – ainda que não necessariamente desejem promover essa discriminação – há uma refutação parcial da hipótese que abre as portas para a criminalização prática do racismo algorítmico, o que demandará mais pesquisas.

4. PONTE PARA O FUTURO: QUESTÕES DE POLÍTICA CRIMINAL E REGULATÓRIA PARA OS ALGORÍTMOS DE SELEÇÃO E RECRUTAMENTO

De acordo com o já apresentado neste trabalho, é possível perceber que as ferramentas tecnológicas – incluindo-se aqui a IA e as técnicas de aprendizagem de máquina – podem ampliar as capacidades e habilidades humanas, mas podem, também, ampliar os riscos para as pessoas em faces de “ações” e “decisões” não necessariamente tomadas por pessoas. Neste sentido, Beck (1992, p. 10) afirma que os contornos da sociedade industrial deram forma a uma nova organização social, que ele define como sociedade do risco. Segundo o autor, “no centro estão os riscos e consequências da modernização, que se revelam como ameaças irreversíveis à vida de plantas, animais e seres humanos¹¹⁹” (BECK, 1992, p. 13). Neste contexto, o que se busca fazer é, a partir da compreensão do conceito de sociedade do risco, debater os modelos de direito penal e regulatórios adequados para possibilitar um efetivo combate ao racismo algoritmo sem que, por outro lado, haja violação de direitos e garantias constitucionais aplicáveis ao direito penal e ao processo penal no contexto brasileiro.

Assim, alguns temas tratados anteriormente serão retomados neste capítulo, mas com enfoque diferente. Se no capítulo anterior a apresentação das normas e sistemas penais aplicáveis para a responsabilização por danos decorrentes do uso de IA tomava como premissa a realidade atual (normas vigentes e aplicáveis), neste capítulo, o debate extrapolará essas normas, e debaterá modelos teóricos e possíveis alterações nesse sistema normativo vigente de responsabilização por IA, buscando adequar o modelo existente às especificidades das novas tecnologias e dos riscos destas decorrentes.

Para isso, este capítulo está dividido em duas partes: a primeira delas debate as dificuldades e possibilidades político-criminais existentes no contexto algorítmico e o combate ao racismo algorítmico no mercado de trabalho; a segunda apresenta algumas possibilidades em relação à

¹¹⁹ Tradução nossa. No original: “at the center lie the risks and consequences of modernization, which are revealed as irreversible threats to the life of plants, animals, and human beings”.

regulação de IA, com objetivo de analisar as possibilidades regulatórias no que trata dos casos e soluções mapeadas neste trabalho.

4.1. MODELOS PENAIIS NA SOCIEDADE CONTEMPORÂNEA E A IA DE SELEÇÃO E RECRUTAMENTO

No estudo de casos, ficaram evidentes as dificuldades de estabelecimento de responsabilidade penal decorrente de IA, justamente pelo afastamento de pessoas da tomada de decisões e, ainda, pelo fracionamento das etapas pela qual a decisão é tomada por humanos. Não bastasse isso, muitos casos reais serão mais complexos que os analisados: se nos casos apresentados cada etapa foi realizada por somente uma pessoa, há estruturas empresariais em que uma pluralidade de agentes responderá pela programação do algoritmo e outra pluralidade por sua operação, por exemplo. Desse modo, muitos dos modelos de política criminal podem ser insuficientes para esse novo contexto em que o cometimento de crimes pode ocorrer.

Retornando-se ao conceito de sociedade do risco, José Luis Díez Ripollés (2015, p. 59-61) distingue três realidades que a compõem¹²⁰: i) a generalização de novos riscos para a coletividade, usualmente decorrentes das novas tecnologias; ii) dificuldade de responsabilização e individualização da produção desses novos riscos, pela complexidade e da multiplicidade de fatores de sua geração; iii) a potencialização da percepção desses riscos, pela intensa cobertura dos meios de comunicação em massa. Segundo o autor, a política criminal que busca responder a essas demandas sociais se caracteriza pela: i) ampliação do objeto da intervenção do direito penal; ii) busca de responsabilização penal dos poderosos; iii) intervenção penal como instrumento preponderante de controle social, colocando em segundo plano a lógica da subsidiariedade da intervenção penal; iv) adequação do conteúdo penal (material e processual) às novas dificuldades decorrentes da persecução penal na sociedade do risco, com flexibilização das regras de imputação e do sistema de garantias do processo penal.

Silva Sánchez (2001, p. 109-110), por sua vez, afirma que o fenômeno da globalização gera uma tendência de universalização do direito, incluindo-se o direito penal e, mais especificamente sobre a criminalidade organizada, que a soberania estatal não mais pode ser aplicável como uma barreira à responsabilização penal dos poderosos. Esse fenômeno de universalização do direito,

¹²⁰ Como muitos desses riscos decorrem das ferramentas tecnológicas que interconectam as pessoas, o conceito de sociedade em rede também é interessante para mostrar um dos principais mecanismos de ampliação desses riscos.

segundo o autor, gera uma tendência “mais de expansão do que de restrição do Direito Penal¹²¹” (SILVA SÁNCHEZ, 2001, p. 111).

Neste contexto de expansão do direito penal, Hassemer (2008, p. 40) aponta um processo de simplificação de procedimentos penais como a busca de “economia através da abolição de formalidades”, e de forma crítica, afirma que esse “controle do desvio não formalizado, que se serve dos instrumentos drásticos do Direito Penal, é um dos fantasmas assustadores da modernização” (HASSEMER, 2008, p. 43). Dessa forma, o autor defende que a necessidade de adaptação penal às novas situações de um mundo cada vez mais complexo deve respeitar “as tendências de proteção dos direitos individuais que caracterizam a Filosofia iluminista” (HASSEMER, 2008, p. 45).

Em sentido semelhante e acerca de expansão do direito penal, Ripollés afirma que existe uma moda do ‘risco’ dentro da qual se tenta:

[...] equiparar os riscos derivados do uso das novas tecnologias com aqueles costumeiros da vida cotidiana, como consequência da crescente presença de auxílio desemprego e marginalização social, supõe aludir a duas fontes de risco radicalmente distintas em sua origem, agentes sociais causadores, natureza objetiva e subjetiva dos comportamentos e consequências nocivas produzidas. (RIPOLLÉS, 2015, p. 69)

No contexto do surgimento de novos riscos, talvez o grande dilema do direito penal contemporâneo seja, justamente, saber se é possível equacionar as garantias do direito penal clássico e, ainda assim, responder adequadamente a esses novos riscos. Em decorrência disso, passa-se a apresentar algumas teorias criadas para tentar responder a esse questionamento.

Para tentar compreender o funcionamento do direito penal ao longo da história, Silva Sánchez (2001, p. 163) divide-o em velocidades, sendo que a primeira velocidade do direito penal corresponde ao direito penal clássico, com incidência da pena de prisão, mas em que devem ser rigidamente mantidos “os princípios político-criminais clássicos, as regras de imputação e os princípios processuais¹²²”. A segunda velocidade, por sua vez, afasta a pena de prisão e, por limitar outros direitos não afetados pela prisão, pode ter por flexibilizado o rol de limitações e garantias atribuíveis ao direito penal tradicional (SILVA SÁNCHEZ, 2001, p. 163). Ou seja, na segunda velocidade, afasta-se a prisão, mas mitigam-se as garantias aplicáveis no contexto do direito penal de primeira velocidade.

O autor ainda discorre sobre a aplicabilidade de uma terceira velocidade, que derivaria da simplificação de um sistema de limitações e garantias sem que se abrisse mão da aplicação da pena de privação de liberdade (SILVA SÁNCHEZ, 2001, p. 163). Silva Sánchez (2001, p. 165-166) admite

¹²¹ Tradução nossa. No original: “más de expansión que de restricción del Derecho penal”.

¹²² Tradução nossa. No original: “los principios político-criminales clásicos, las reglas de imputación y los principios procesales”.

a possibilidade de aplicação da terceira velocidade, condicionando-a aos casos de emergência em uma espécie de “direito de guerra” em face de um conflito de gravidade extrema. Por outro lado, o autor reconhece que a tendência existente é de ampliação dessa forma de aplicação mais grave e menos criteriosa do direito penal, devendo essa terceira velocidade ser evitada (SILVA SÁNCHEZ, 2001, p. 167).

Assim, o autor defende a primeira velocidade para aplicação ao direito penal clássico em que os bens jurídicos são bem individualizados, como o caso dos crimes contra a vida e o patrimônio por exemplo. Por outro lado, em relação à criminalização de condutas contra bens jurídicos não universalizáveis, deve ser aplicada a segunda velocidade do direito penal (SILVA SÁNCHEZ, 2001, p. 159-162).

Em sentido semelhante, Winfried Hassemer defende a aplicação de um direito penal clássico, com as garantias da filosofia iluminista para os crimes comuns e, por outro lado, a aplicação de uma forma de punição apresentada como direito de intervenção especial para a criminalidade cujos bens jurídicos não são individualizáveis – a criminalidade da modernidade. Segundo o autor:

[...] quanto àqueles problemas da sociedade moderna que provocam a modernização do direito penal, de que fossem regulados em um direito de intervenção especial, o qual está situado entre o direito penal e o direito da contrariedade à ordem pública, entre o direito civil e o direito público, o que dispõe, na verdade, de garantias e regramentos processuais menos exigentes do que o direito penal, mas que, em contrapartida, está equiparado com sanções menos intensas em face do indivíduo. Um direito de tal natureza ‘moderna’ não seria somente menos grave normativamente, ele seria, também, de fato, mais adequado para recepcionar os problemas especiais da sociedade moderna. (HASSEMER, 2007, p. 207)

Neste contexto, o direito penal clássico se desenvolve após a derrocada do direito natural e não surge de princípios superiores e incontestáveis, mas de “um acordo de uma ordem jurídica possível por meio daqueles que eram por ela alcançados: o contrato social” (HASSEMER, 2007, p. 192). Sendo este contrato permitido pelo convívio harmonioso entre os indivíduos, “os limites da renúncia à liberdade devem ser marcados com garantias precisas” (HASSEMER, 2007, p. 192). Neste contexto, surge uma série de limitações ao poder de punir do estado, sendo que entre elas se encontra a teoria do bem jurídico, que – no contexto do direito penal clássico – funciona com o “um critério negativo de uma incriminação legítima”, de modo que “sem uma violação palpável a um bem jurídico, não há um fato típico” (HASSEMER, 2007, p. 193).

No entanto, na passagem para o direito penal moderno, a teoria do bem jurídico que era aplicada como um limitador de punições infundadas – ainda que legalmente previstas, numa espécie de limitação ao legislador – passa a ser um conceito positivo – fazendo um apelo a este mesmo legislador, “para que este imponha uma pena a determinados tipos de comportamento” (HASSEMER, 2007, p. 195). Neste contexto, tomam forma no direito penal princípios como o da vedação à proteção

deficiente, que surgem para permitir e incentivar a criminalização de condutas, “na medida em que se pretende legitimar uma atuação estatal protetora de direitos fundamentais” (MENTOR, 2016, p. 36)¹²³.

Ainda sobre os bens jurídicos, Hassemer define que no direito penal moderno estão algumas áreas decorrentes das alterações do mundo, como meio ambiente, economia, drogas e, também, a programação de dados (HASSEMER, 2007, p. 197). Assim, os bens jurídicos relacionados às novas tecnologias de informação estão dentro do escopo deste novo direito penal – em que as regras e criminalizações do direito penal clássico se mostram insuficientes.

É justamente neste contexto – e após o reconhecimento da insuficiência do direito penal clássico – que são legitimados:

discursos de endurecimento das normas penais, sob a perspectiva de que sua contundência seria mais eficaz na prevenção dos novos delitos do que a intervenção de outros ramos do direito, proporcionando, desta feita, o incremento da produção legislativa criminalizadora de condutas aviltantes dessas atividades. (MENTOR, 2016, p. 71)

Em relação à segunda velocidade do direito penal de Silva Sanchez e ao direito de intervenção especial de Hassemer, Mentor (2016, p. 80) afirma que ambas as teorias seriam formas de proteger os bens jurídicos do direito penal contemporâneo, “na medida em que permitiriam a flexibilização de algumas garantias processuais diante da impossibilidade de adoção de penas privativas de liberdade”, aplicando à criminalidade econômica “as sanções restritivas de direitos”. Acerca da diferenciação entre as teorias, o autor afirma que:

O autor espanhol [Silva Sanchez], em vista do histórico de arbitrariedades perpetradas pela Administração Pública durante o período ditatorial de Franco, aposta na criação de uma nova teoria geral do crime, com uma segunda velocidade do direito penal, cuja imposição de sanções permaneceria no contexto de um processo penal, com a imparcialidade de um juiz togado e com o aproveitamento de todo o arsenal teórico da parte geral do direito penal. Por sua vez, WINFRIED HASSEMER prefere transferir a atribuição de tutela e sanção desses novos bens jurídicos para um novo ramo do direito (que ele denomina de direito de intervenção), o qual seria intermediário entre o direito penal e o direito puramente administrativo, mas sempre com a possibilidade de controle judicial. (MENTOR, 2016, p. 100)

Assim, em decorrência de particularidades históricas dos países de origem dos autores (Espanha e Alemanha), alguns detalhes dos modelos foram diferenciados, com a proposta de um modelo de intervenção – fora do direito penal – por Hassemer, e um modelo de direito penal sem pena de prisão, no caso de Silva Sanchez. Se, por um lado, em face das preocupações com o fascismo espanhol, Silva Sanchez não afastou a figura do juiz, para Hassemer, o surgimento das infrações

¹²³ Na jurisprudência constitucional brasileira, o referido princípio surge no voto do Min. Gilmar Mendes no julgamento do HC 82.424 pelo Plenário do Supremo Tribunal Federal, tendo outros usos posteriores. Para uma análise mais detalhada, vide L. RODRIGUES (2012, *passim*).

administrativas em face de questões tributárias e econômicas¹²⁴ no pós-Segunda Guerra Mundial foi o fenômeno de observação que o permitiu pensar que determinadas infrações deveriam estar fora do direito penal.

No contexto brasileiro, há diversos mecanismos de segunda velocidade do direito penal. Neste contexto, a Lei 9.099/1994 previu acordos sem cumprimento de pena e com restrição de direitos para crimes com penas menores: a transação penal para os crimes de menor potencial ofensivo (pena máxima inferior a 2 anos) e a suspensão condicional do processo para crimes cujas penas mínimas não superem 1 ano. Mais recentemente, a Lei n.º 13.964/2019 alterou o Código de Processo Penal criando, no Brasil, o acordo de não persecução penal, que amplia as possibilidades de afastamento de prisão em casos não abrangidos pelos acordos anteriores, aplicando-se a delitos com pena mínima inferior a quatro anos¹²⁵.

Há, ainda, acordos como a colaboração premiada prevista na Lei 12.850/2013, na qual, embora seja possível o afastamento da pena, essa não é a regra¹²⁶. Nesta mesma esteira, Suxberger e Filho (2016, p. 392) apontam que:

[...] com a tutela de novos bens jurídicos, os sistemas penais dos Estados Democráticos de Direito, caminham para modelos de consenso, nos quais o Estado abre mão de parcela do *ius puniendi* e incorpora institutos negociais no ordenamento, sob a justificativa de redução de custos com o aparato judiciário estatal e de maior eficiência no combate à criminalidade organizada.

No entanto, a despeito desses mecanismos que afastam ou mitigam a aplicação da pena no processo penal – instrumentos de segunda velocidade –, sua aplicação não ocorre nos termos das velocidades do direito penal apontadas por Silva Sanchez. Isso porque são questões outras, e não os bens jurídicos tutelados, que possibilitam a aplicação desses mecanismos, a exemplo dos patamares de pena aplicáveis ou, ainda, da colaboração de um membro de organização criminosa. Assim, se o mecanismo é semelhante à definição de segunda velocidade do direito penal, sua forma de implementação parece não se relacionar aos bens jurídicos coletivos cuja proteção se busca ampliar no contexto da sociedade do risco.

¹²⁴ Em relação ao tema, MENTOR (2016, p. 98) afirma o seguinte: “Após 1945 surge na Alemanha uma nova categoria de ilícitos: as chamadas infrações administrativas, cujo objetivo era a regulação de infrações tributárias e econômicas, sendo regidas, porém, pelas regras do Estado de Direito¹⁶⁷. Desta feita, é perceptível que a transposição de certas infrações da seara penal para a administrativa, como nos ilícitos contra a ordem econômica, se deu com bases garantistas, escorada nos pilares das garantias individuais e do Estado de Direito, o que certamente contribuiu para que a teoria do direito de intervenção de WINFRIED HASSEMER se construísse a partir dessa pedra fundamental.”

¹²⁵ Para uma visão mais precisa sobre os requisitos e limitações do acordo de não persecução penal, *vide*: VASCONCELOS (2022).

¹²⁶ Segundo a mencionada legislação, entre uma ampla lista de benefícios possíveis, dois deles podem afastar o processo ou a penal. Um deles é a previsão de não denunciar o colaborador em relação a infração penal cuja respectiva ação penal não se tenha iniciado (art. 4º, §4º). A outra é a possibilidade de perdão judicial do colaborador (art. 4º, *caput*).

Já acerca da diferenciação entre as infrações administrativas e as penais, Mentor (2016, p. 105-108) aponta existirem duas correntes: uma conhecida como diferenciadora – defendida por James Goldschmidt – segundo a qual os ilícitos penais carregam necessariamente um desvalor ético não existente nas infrações administrativas e outra conhecida como unitária – defendida por autores como Heinz Mattes – que defende haverem infrações penais sem que haja um desvalor ético associado, e, por outro lado, que infrações administrativas podem possuir um desvalor ético a elas relacionado, não residindo aí a diferenciação entre infrações penais e administrativas.

Para o presente trabalho, a teoria adotada é a unitária, mas o que se faz com ressalvas que devem agora ser explicadas. Em respeito à teoria do bem jurídico – em sua vertente limitadora de incriminações infundadas pensada no contexto do direito penal clássico –, não se pode pensar em crimes (infrações penais) despidas de desvalor ético, no entanto, é possível que as infrações administrativas possuam ou não desvalor ético. Não se compreende, portanto, que ambas as modalidades podem ou não ser carregadas de desvalor ético, mas que infrações penais sempre possuem desvalor ético associado, sendo que as infrações administrativas podem possuir ou não esse desvalor, de modo que não é nesse ponto que reside a diferenciação entre as modalidades de infração, mas nas consequências, sendo que somente as penais podem resultar em privação de liberdade.

Neste contexto, ante a desnecessidade de que haja desvalor ético nas infrações administrativas, elas podem se aplicar até mesmo de forma objetiva, de modo que – comprovado o nexo causal – é possível a responsabilização administrativa ainda que inexistente dolo ou culpa em determinados casos (MENTOR, 2016, p. 115-116). Não se descarta a imputação subjetiva de infrações administrativas, mas apenas apresenta-se a sua desnecessidade para a realização de uma infração administrativa, especialmente em relação as pessoas jurídicas.

Cabe aqui, ainda, um esclarecimento. A despeito da mitigação de garantias das infrações administrativas – em comparação com as penais –, tem surgido defesa cada vez mais forte, na doutrina, da ampliação das garantias no contexto do direito administrativo sancionador¹²⁷. No entanto, a posição defendida por Mentor é que:

[...] essa transposição de garantias do Direito Penal para o direito sancionatório deve ser feita *cum grano salis*, posto que a adoção integral da parte geral do direito penal para o direito administrativo, criando-se uma espécie de parte geral sancionatória, reduziria ao nada a efetividade desse novo ramo do direito, o qual estaria fadado a ser uma espécie de direito penal especial, destinado unicamente à tutela da ordem econômica, mas sem as vantagens de eficácia e celeridade do direito sancionatório decorrentes das reduções de garantias implementadas nesse ramo do Direito em troca da impossibilidade de aplicação de penas privativas de liberdade. (p. 133)

¹²⁷ Neste sentido, autores como Guardia (2014, p. 789-790) defendem a aplicação das garantias do processo penal no direito administrativo sancionador, ante a existência de um único *ius puniendi* estatal.

Partindo dessa diferenciação entre as velocidades do direito penal feita por Silva Sanchez e de direito penal de direito de intervenção feita por Hassemer, Diogo Mentor (2016, p. 21) passa a debater a responsabilidade penal da pessoa jurídica, apontando que “na prática, as sanções penais impostas às pessoas jurídicas que cometem algum tipo de ilícito são muito semelhantes às punições administrativas, resumindo-se a penas restritivas de direitos e multas”. Sobre o tema, o autor aponta que a responsabilidade penal individual surge do ideário iluminista pós-Revolução Francesa, no contexto do direito penal clássico, sendo que, nos países de *Common Law*, a atribuição de responsabilidade penal a pessoas jurídicas passou a ser mais aceita a partir do século XIX (MENTOR, 2016, p. 28).

De qualquer modo, a despeito da ampla complexidade do debate acerca da responsabilidade penal de pessoas jurídicas, esta não é uma responsabilidade de direito penal clássico, em que incide a pena de prisão (até mesmo pela impossibilidade de aplicar uma pena de pessoas físicas às jurídicas), funcionando, portanto, como um instrumento de direito penal de segunda velocidade.

Retornando-se aos casos, foi possível perceber as dificuldades no estabelecimento de responsabilidade penal de pessoas relacionadas a algoritmos de seleção e recrutamento, seja pelas dificuldades probatórias existentes, seja pelas limitações existentes no direito penal brasileiro em relação à necessidade de dolo como elemento subjetivo. Assim, a despeito da possibilidade de responsabilização de agentes envolvidos em IA de seleção e recrutamento por dolo eventual, seria necessária a comprovação do conhecimento prévio dos agentes em relação ao contexto discriminatório. Além disso, o debate sobre responsabilidade penal da pessoa jurídica também surge como uma possibilidade de responsabilização por danos decorrentes de IA.

Neste sentido, a posição adotada no presente trabalho é de que somente por meio do afastamento de determinadas garantias inerentes ao direito penal será possível o combate adequado às violações da modernidade, segundo posições defendidas por Hassemer, Silva Sanchez e, no contexto brasileiro, por autores como Mentor. A questão, é, portanto, como resolver esse problema da tecnologia e da responsabilidade pelos danos dela decorrentes de IA de das novas tecnologias sem violar o núcleo duro do direito penal clássico.

Acerca do racismo algorítmico – diferentemente de outros riscos decorrentes das tecnologias de informação de comunicação – não há um bem jurídico transindividual, mas tão somente a vedação à discriminação racial, problema existente (possivelmente de forma mais marcante) nos termos do direito penal clássico. O racismo não foi criado pela tecnologia, sendo uma ideologia e uma realidade (enquanto prática social) já consagrada, ainda que absolutamente indesejada e criminosa.

No entanto, se nas modalidades tradicionais de cometimento do crime de racismo há autor e vítima individualizados – como no Caso 1 apresentado neste trabalho –, no caso do racismo

algorítmico a violação de bens jurídicos tende a ser transindividual ante a pluralidade de agentes envolvidos, havendo dificuldades de individualizar qual participação cada um desses agentes teria na prática da infração e na causação do resultado danoso. A questão – no racismo algorítmico – não é o bem jurídico em si, que poderia ser abarcado pelas estruturas do direito penal clássico, mas os mecanismos tecnológicos de violações a estes bens jurídicos: a questão é compreender, portanto, como o ambiente tecnológico e a inteligência artificial demandam tratamento especial para a responsabilização das violações existentes, ainda que a bens jurídicos de direito penal clássico.

Por outro lado, as limitações à responsabilização penal de entes coletivos que decorre das amplas dificuldades de individualização da prática do racismo algoritmo, seja das pessoas físicas autoras, seja das vítimas, como foi possível perceber nos casos estudados. Ao se afastar a responsabilidade penal, passa a ser desnecessário individualizar atitudes de pessoas física, havendo caminhos para a responsabilização objetiva das pessoas jurídicas que operam e criam algoritmos de seleção e recrutamento.

Talvez a solução para a responsabilidade por danos causados por IA – e mais especificamente IA de seleção e recrutamento – seja a manutenção da responsabilidade individual nos casos em que esta pode ser atribuída de acordo com as regras do direito penal clássico, como nos casos estudados no capítulo anterior. A responsabilização e eventual reparação dos danos não precisa passar, necessariamente pelo direito penal. Considerando-se isso, e sendo a subsidiariedade um dos princípios basilares do direito penal, se é possível evitar a pena privativa de liberdade e o direito penal, este deve ser o caminho seguido.

Assim, o que se propõe no presente trabalho é que a responsabilidade penal decorrente de IA – mesmo no caso de discriminação algorítmica – seja mantida nos casos abrangidos pela legislação atual, que exige a responsabilização do crime por dolo, ainda que em sua modalidade eventual. No entanto, nos casos de impossibilidade de individualização dos agentes e sua colaboração para o evento danoso, ou, ainda, de não caracterização do dolo, a resposta adequada deve residir – segundo nosso entendimento – fora do direito penal, com possibilidade de aplicação de sanções diretamente às pessoas jurídicas que operam as formas de IA de seleção e recrutamento e, até mesmo, de forma objetiva.

Há, porém, um problema que deve ser pensado na cumulação da responsabilidade penal e da administrativa: consoante Lobo da Costa (2013, p. 235-236), dada a expansão do direito penal econômico e do direito administrativo sancionador e a inexistência de finalidades distintas entre essas modalidades sancionatórias, há crescentes riscos de *bis in idem*, por meio da duplicidade de punições criminais e administrativas em relação a determinados ilícitos vinculados ao direito penal econômico. Neste sentido, a autora defende uma ampliação do *ne bis in idem* para abarcar situações em que a

cumulação de sanções administrativas e penais representem uma dupla punição (COSTA, 2013, p. 237). Esse risco é mais um motivo para evitar cumulação de responsabilidade penal e administrativa, de modo que a resposta administrativa ou por meio de um direito de intervenção especial pode servir, também, para dar aplicabilidade ao *ne bis in idem* na prática sancionatória contemporânea, não havendo dois espaços de punição das infrações econômicas, eliminando a concorrência e cumulação entre estes mecanismos de aplicação sancionatória¹²⁸.

A seguir, passa-se a discorrer sobre as possibilidades regulatórias no sentido de limitar o racismo algoritmo relacionado às atividades de seleção e recrutamento. Se o direito penal deve ter sua aplicação de forma subsidiária, o debate que se estabelece neste tópico seguinte é qual forma de solução primária pode ser aplicada para regular esta aplicação de IA.

4.2. FUGINDO DA PRISÃO? AS POSSIBILIDADES REGULATÓRIAS E O COMBATE AO RACISMO ALGORÍTMICO

Ao definir as funções do direito regulatório, Aranha (2015, p. 8) afirma que para além da proteção dos mercados, este ramo do direito busca resguardar “o direito à igualdade de condições concorrenciais”. No entanto, não é somente a igualdade que deve ser resguardada pela regulação, que “tem por finalidade preeminente a proteção dos direitos fundamentais” (ARANHA, 2015, p. 9). Assim, ante as dificuldades de normatização e legislação apta a compreender e regular as complexidades de uma realidade multifacetada e dinâmica:

[...] o conteúdo normativo encontra-se claramente remetido a decisões de normatização secundária, significando que o adensamento do conteúdo dos direitos fundamentais depende, hoje, em grande medida, de decisões estatais influentes sobre os setores tidos como essenciais ao desenvolvimento socioeconômico do país e o fenômeno da regulação ocupa posição privilegiada em tal espaço decisório. (ARANHA, 2015, p. 11)

¹²⁸ Na doutrina, há posições em defesa de uma aplicabilidade desse princípio entre áreas distintas do direito, como o administrativo sancionador e o penal, como a apresentada por Costa (2013, Cap. 4). No entanto, a aplicabilidade desses entendimentos em relação ao direito tributário deve ser mais amplamente debatida, motivo pelo qual o tema será tratado de modo apenas panorâmico neste trabalho.

Uma forma de se afastar o *ne bis in idem* entre distintos ramos do direito é o da independência das instâncias, especialmente entre a penal e a administrativa, como apontado por Osório (2020, n.p., Parte II, Cap. 3, 3.8.5 Causas de justificação). Segue o autor, apontando a radicalização da independência das instâncias, ao invés de promover a separação dos poderes, pode isolá-los, criando desarmonia e insegurança jurídica: “Por aí tem transitado tanto a ideia de proporcionalidade quanto, em maior medida, uma hermenêutica restrita e limitada em torno ao princípio do *non bis in idem* na jurisprudência e na doutrina brasileiras, ainda predominando o ideário ínsito à independência das instâncias, como corolário da separação de Poderes, em detrimento de outras alternativas igualmente legítimas. E vale destacar que a radicalização dessa independência das instâncias tem conduzido, não raro, ao isolacionismo das instâncias, à falta de harmonia entre os Poderes e à insegurança jurídica, em muitos níveis relevantes.” (OSÓRIO, 2020, n.p., Parte II, Cap. 4, 4.2.1.1 Descrição do *non bis in idem* no cenário brasileiro)

Sobre o tema, Fonseca e Fonseca (2020, *passim*) defendem que haja um diálogo entre as ordens jurídicas supranacional europeia e a brasileira no sentido de se debater a aplicação, no direito brasileiro, de decisões que vedam o *ne bis in idem* transversal, que ocorre entre diferentes ramos do direito (como penal e administrativo sancionador).

O Estado Regulador, portanto, busca trazer para o ambiente estatal “as noções de profissionalismo e expertise tradicionalmente aplicadas aos negócios privados”, adaptando “ao conceito de expertise na atividade de governar com a conotação de permanência, treinamento e especialização de funções” (ARANHA, 2015, p.12). Segue o autor apontando que a missão do Estado Regulador é “a defesa dos direitos fundamentais via conhecimento técnico, geral e setorial” (p. 26).

Neste sentido, a modalidade regulatória abordada no presente trabalho é a setorial, que “diferencia-se das demais por operar em determinados segmentos de atividades definidas convencionalmente como afins” (ARANHA, 2015, p.112). Consoante Aranha:

A política pública setorial depende, portanto, do *conhecimento setorial* para produzir regramentos viáveis. Ela depende de acompanhamento do desenvolvimento tecnológico para orientar eventuais desígnios utilitaristas de mercado (ou dos *atores* do mercado) na direção do interesse público. (ARANHA, 2015, p.112)

O que se busca compreender é como mecanismos de regulação setorial relativas à IA de seleção e recrutamento podem ser aplicadas de modo a manter a isonomia na aplicação dessas tecnologias e, especialmente, a realização dos direitos fundamentais no ambiente digital. Como tratando anteriormente (2.1), os princípios estabelecidos no constitucionalismo digital impõem limites éticos, a necessidade de tomada de decisão humana e a necessidade explicabilidade e democratização do conhecimento acerca dos processos algorítmicos – mas o grande sentido desses princípios parece ser o da garantia de direitos fundamentais no contexto digital e, além disso, a responsabilização de quem viola esses direitos.

No entanto, o caminho para a viabilização dessa responsabilidade não necessariamente é penal, como foi possível ver nos estados da arte apresentados no capítulo 2, que apresentaram alguns modelos de responsabilidade no sentido de fugir do vazio de responsabilidade decorrente da aplicação das regras restritivas e individualizadas do direito penal ao contexto algorítmico. As principais propostas em relação à responsabilidade penal giravam em torno de responsabilizar os desenvolvedores e pessoas envolvidas em sua operação ou, ainda, de responsabilizar penalmente as empresas (pessoas jurídicas) responsáveis pela criação das formas de IA.

Mas é justamente esse cenário de busca da responsabilização pelo direito penal – em decorrência das limitações que esse campo do direito necessariamente impõe – que cria os vazios de responsabilidade, em face das limitações criadas por esse ramo do direito. Por essa razão, as propostas de responsabilização fora do direito penal serão avaliadas com maior atenção.

De acordo com o estado da arte acerca da responsabilidade penal e IA, algumas soluções despontam como mais adequadas à realidade algorítmica, quais sejam: (i) responsabilização objetiva cumulada ou não com um modelo de seguro em relação a violações de direitos praticadas por meio

de algoritmos; (ii) uma solução regulatória prévia que somente libera determinadas aplicações de IA após garantir sua segurança para os usuários e consumidores; (iii) proibição do uso de determinadas formas de IA; (iv) criação de mecanismos de registros e de explicabilidade em relação às formas de IA.

Sobre a aplicação da responsabilidade objetiva (i), o cenário vigente no Brasil é o seguinte. Apesar da existência de alguns projetos de lei referentes à regulação de IA, o que está com tramitação mais avançada é o PL 21/2020 da Câmara dos Deputados, que foi aprovado naquela casa com a previsão de responsabilidade civil subjetiva em relação aos danos causados por essas tecnologias. No entanto, após essa aprovação, alguns relevantes juristas brasileiros apresentaram carta aberta no sentido de que a responsabilidade civil seja objetiva¹²⁹. Em seguida, o Senado Federal instituiu comissão de juristas para propor anteprojeto de Marco Legal de Inteligência Artificial com diversos juristas, dentre os quais alguns signatários da mencionada carta aberta.

Neste contexto, o anteprojeto do Marco Legal da IA deve apresentar parâmetros mais adequados para a regulação e para mitigar os riscos decorrentes da aplicação de algoritmos. Além disso, pela convergência de signatários da carta aberta de juristas defendendo um modelo de responsabilidade objetiva por danos decorrentes de IA com membros da comissão de juristas instituída pelo Senado, o caminho legislativo natural que parece surgir é justamente o da responsabilidade objetiva, o que, contudo, certamente demandará debate mais aprofundado do parlamento.

Algumas atividades de risco elevado possuem formas de responsabilização objetiva ilimitada ou com mecanismos de seguro para garantir que o pagamento seja realizado. Neste sentido, o sistema constitucional instituído em 1988 no Brasil possui uma proposta de responsabilidade civil independentemente de culpa para exploração de atividade nuclear¹³⁰. Evidentemente, guardadas as devidas proporções, o modelo de responsabilidade civil por danos causados por IA não precisa ser igual ao previsto para atividade nucleares, mas os limites para que possa haver a reparação dos danos tem que ser tão menor quanto maiores os riscos.

O debate entre responsabilidade objetiva ou subjetiva, ainda que não coincida exatamente, possui algumas semelhanças com o caso estadunidense dos conceitos de tratamento desigual e de impacto desigual, como apresentado por Barocas e Selbst (2016, p. 701-712). Ou seja, reconhecer que o mero resultado desigual – e não somente o tratamento distinto – gera danos e deve ser objeto de reparação abre possibilidades para que o impacto desigual seja reconhecido como dano o que

¹²⁹ Documento disponível online em: “<https://www.conjur.com.br/2021-out-27/especialistas-questionam-artigo-pl-marco-legal-ia>”. Acesso em 26 de novembro de 2022.

¹³⁰ Art. 21, inciso XXIII, alínea “d”.

permite estabelecer a responsabilidade dos envolvidos. No entanto, no debate em relação ao caso estadunidense da discriminação no ambiente de trabalho, há requisitos adicionais para estabelecimento de responsabilidade subjetiva. A responsabilidade objetiva, por sua vez, atribui a responsabilidade pelo mero resultado, independentemente de deveres de cuidado terem ou não sido violados.

Cumulado com esse regime de responsabilidade civil objetiva, há autores que defendem a criação de regimes de seguro obrigatório para danos patrimoniais decorrentes de carros autônomos. Schroll (2015, p. 822-833), por exemplo, defende como uma possibilidade para viabilizar o regime de responsabilidade de veículos autônomos a criação de um fundo nacional de responsabilidade civil. Segundo o autor, esse fundo proibiria que as pessoas cobrassem os danos patrimoniais pela via da responsabilidade civil, que serviria tão somente para danos não patrimoniais. Em sentido semelhante, Yoshikawa (2019, p. 1178-1186) defende a criação, no contexto estadunidense, de um fundo federal para acidentes de trânsito, semelhante ao que já existe na Nova Zelândia e cobre, também, acidentes de trânsito com envolvimento de carros autônomos.

A despeito de esta ser uma possível solução interessante para o caso dos carros autônomos, em que os danos causados são individualizáveis e economicamente quantificáveis, um sistema semelhante para o combate à discriminação algorítmica no mercado de trabalho seria de difícil definição e de mais difícil implementação. O que pode ser a grande barreira é estabelecer quem seriam as vítimas e quais as formas de repará-las. Outras ações, como por exemplo a criação de instituições com o objetivo de reparar essas distorções e o pagamento de multas ou taxas a essas instituições pode ser interessante, mas, ainda assim, não haveria como saber se as pessoas realmente afetadas pela discriminação seriam as beneficiárias efetivas das reparações.

Outra solução possível é a de regulação prévia das ferramentas algorítmicas de seleção e recrutamento (ii) e de outras formas de IA cujas aplicações apresentem riscos sociais representativos. Como apresentado anteriormente (2.1) a proposta legislativa de AIA da Comissão Europeia entende que algoritmos de seleção e recrutamento são de alto risco e que, em consequência disso, exige-se a testagem e documentação prévia desses sistemas. Ou seja, somente ao se comprovar a segurança e a confiabilidade dessa modalidade de IA é que haveria sua liberação.

Em sentido semelhante, há a proposta encontrada de revisão de literatura sobre responsabilidade penal e IA, apresentada por Tutt (2019, p. 118-123), que propõe a criação de uma agência reguladora aos modos da FDA estadunidense (ou da ANVISA, em um paralelo com o caso brasileiro). Segundo o autor, a criação de determinadas formas de IA pode funcionar como o desenvolvimento e liberação de um novo remédio, em que há testes controlados e, tão somente após

a comprovação de sua segurança e eficácia pela agência reguladora que o medicamento (e nesse caso o algoritmo) pode ser lançada no mercado.

Assim, essas propostas apresentam como perspectiva comum o reconhecimento de que, pelos possíveis elevados riscos que podem gerar e pelas dificuldades de responsabilização por eventuais violações, somente após controle prévio que se pode assegurar a liberação de determinadas formas de IA e, para a Comissão Europeia, os algoritmos de seleção e recrutamento se encontram nesse grupo. Neste sentido, a regulação prévia pode ser compreendida como corolário do princípio da prudência, consagrado no constitucionalismo digital.

Outra solução possível em relação a determinadas formas de IA é sua completa proibição ou banimento (iii). A própria Comissão Europeia, na proposta do AIA apresenta essa solução em determinados casos de aplicação algorítmicas. Essa é a resposta, por exemplo, aos aplicativos de *social scoring*, que categorizam as pessoas. No entanto, acreditamos que somente em casos extremos é que o banimento da aplicação algorítmica pode ser a solução, o que não ocorre no uso para seleção e recrutamento de mão de obra.

Outro caminho possível para o combate ao racismo algorítmico é o da explicabilidade (iv). Como apontado anteriormente (2.1), a explicabilidade e a democracia são princípios que regem o constitucionalismo digital e a lógica desses princípios é que as pessoas afetadas por algoritmos devem ter informações que as tornem capazes de compreender a atuação e as consequências dos algoritmos em suas vidas. Assim, de modo geral, a explicabilidade serve para que seja possível compreender o funcionamento do algoritmo.

Compreender a forma como o algoritmo opera pode ser uma ferramenta para evitar ou punir eventuais violações de direitos. Dessa forma, ainda que a explicabilidade não resolva ou evite qualquer forma de discriminação algorítmica, pode ser um mecanismo para que se possa perceber os mecanismos de racismo algorítmico, possibilitando fazer alterações a fim de evitar esse fenômeno e, até mesmo, ajudar a atribuir responsabilidades aos agentes envolvidos em sua prática.

As medidas apresentadas para combater a discriminação algorítmica certamente não esgotam as possibilidades de prevenção e responsabilização dessa prática, sendo somente uma tentativa de apresentar as principais ideias encontradas para responder a esse problema. A enumeração feita, além disso, não descarta as possibilidades de combinação entre as soluções, de modo que, por exemplo, a explicabilidade, pode ser feita de forma conjunta com outras medidas de prevenção e combate aos danos possivelmente provocados por formas de IA.

Há, portanto, diversas medidas distintas do direito penal para combater o racismo algorítmico, e, assim sendo, a pena (enquanto ferramenta do direito penal) deve ser aplicada somente em casos em que as demais ferramentas não forem adequadas ou forem insuficientes. Defende-se, portanto, o

emprego de ferramentas do direito de intervenção especial (ou uma modalidade de direito administrativo sancionador) em face das formas de IA em detrimento até mesmo do direito penal de segunda velocidade, que, a despeito de abrir mão da pena, é mantido no âmbito da justiça criminal, possuindo um tempo mais lento de julgamento e corpo técnico não especializado para responder a questões envolvendo temas de tecnologia. Inclusive, um dos motivos de que haja uma opção pela aplicações de soluções fora do direito penal é o medo que que, no contexto da expansão penal, sanções de um direito penal de segunda velocidade passem para a terceira velocidade, cujo abuso é uma tendência mundial, como reconhece o próprio Silva Sanchez.

Existem, ainda, outros obstáculos que devem ser pensados e que colocam a regulação da discriminação algorítmica por meio de IA de seleção e recrutamento. Diferentemente do caso do veículos autônomos, em que há um produto físico circulando no país (ainda que produção e tecnologia possam ser estrangeiras), no caso dos algoritmos de seleção e recrutamento, um serviço produzido e regulado em outro país pode gerar discriminação no contexto brasileiro, uma vez que são plataformas internacionais como a LinkedIn que estão tomando a dianteira nesses projetos. Uma questão que resta, independentemente da solução que se adote, é como responsabilizar agentes que não estão em território brasileiro.

Uma outra questão que se coloca é a do suposto debate existente entre regulação dessas novas tecnologias e o seu desenvolvimento. Esse debate muitas das vezes trata limitar e regular algumas aplicações dessas tecnologias como um impedimento para o desenvolvimento de novas tecnologias. Neste sentido, pode-se apontar a existência de um conflito entre interesses econômicos e regulatórios.

Sobre a influência da economia no direito regulatório e vice-versa, Kellner (2022, *passim*.) afirma que há divergência na economia sobre quais critérios e métodos aplicados e que a economia não necessariamente se preocupa com questões que devem importar ao regulador. Assim, é o regulador que deve definir quais argumentos econômicos importam à regulação e qual seu peso para decidir o que regular¹³¹.

Ademais, e retornando-se às ideias de Rodrigues (2020, p. 45-47), a responsabilidade moral de fazer as decisões do que é aceitável ou não, lícito ou não, cabe ao ser humano, e não à máquina.

¹³¹ Segundo o autor: “[...] a aplicação pela autoridade do direito regulatório deve funcionar como um delimitador normativo do que pode ser considerado como um argumento aceitável vis-à-vis inaceitável. O argumento parece se sustentar na constatação dos seguintes problemas: (i) a busca por eficiência econômica revela que argumentos econômicos, em regra, ignoram justiça distributiva e outros fins sociais, com os quais o sistema jurídico costuma se preocupar; (ii) os pressupostos básicos dos modelos econômicos mostram certas idealizações que requerem um escrutínio crítico quando de uma análise jurídica. No modelo de comportamento econômico, por exemplo, pressupõe-se fundamentalmente que os atores envolvidos estão perfeitamente informados, fato que desconsidera uma série de limitações aventadas acima; e (iii) precisão/imprecisão das afirmações que podem ser feitas usando os padrões emprestados da economia.” (KELLNER, 2022, p. 30)

Ainda, segundo a autora, ao transferirmos para máquinas a responsabilidade por decisões morais, devemos assumir os riscos decorrentes de violações éticas praticadas por essas máquinas. A dimensão ética deve possuir papel central na regulação das tecnologias de IA, especialmente em se tratando de algoritmos de seleção e recrutamento.

Assim, ainda que, neste trabalho, não haja defesa da necessidade de tomada de uma posição ou outra, entende-se que o papel atribuído ao direito penal na regulação de novas tecnologias deve ser restritivo e secundário. O papel central deve ser atribuído a ferramentas regulatórias, o que evita o enfrentamento da responsabilidade penal das pessoas jurídicas e a limitação à responsabilização objetiva no âmbito deste ramo do direito.

Em relação às ferramentas regulatórias empregadas, há necessidade de alguma forma de reparação dos danos, sendo interessantes as propostas de responsabilização civil objetiva, cumulada ou não com regimes de seguro obrigatório. A despeito deste tipo de medida poder ser útil para casos em que as vítimas são individualizadas – como nos carros autônomos –, para as formas de IA de seleção e recrutamento, ante a dificuldade de individualização das vítimas e de reparação de danos causados, mecanismos de controle prévio e de registro como os previstos no AIA podem funcionar de forma mais adequada. Além disso, possibilidades como de criação de uma agência reguladora ou órgão regulador para o controle das aplicações de IA ou ampliação das competências da Autoridade Nacional de Proteção de Dados, que possui vocação semelhante, a de regular a proteção de dados no país, podem ser soluções interessantes.

4.3. CONSIDERAÇÕES FINAIS

Como apresentado, o direito penal possui grandes dificuldades de proteger os bens jurídicos e de encontrar técnicas de tutela adequadas para combater as violações a direitos decorrentes dos novos riscos e – ao mesmo tempo – manter o respeito aos princípios que o regem, como a *fragmentariedade* e a *legalidade*. Neste sentido, o combate a violações praticadas por IA de seleção e recrutamento por meio do direito penal demandaria a criação de mecanismos de responsabilidade penal objetiva, bem como a responsabilização penal de pessoas jurídicas. Tal situação coloca o direito penal em seu limite e faz com que, inclusive, ele tenha que renunciar a sua grande característica diferenciadora para outros ramos do direito: a pena privativa de liberdade.

No entanto, a solução proposta neste trabalho, amparada no conceito de direito de intervenção mínima da Hassemer e na teoria do direito administrativo sancionador é que a regra para a responsabilização por danos decorrentes de IA de seleção e recrutamento seja a da responsabilização não penal. Assim, surgem possibilidades de responsabilização civil objetiva com ou sem um

mecanismo de seguro obrigatório para a reparação e danos. Em relação ao racismo algorítmico em relação a IA de seleção e recrutamento, é difícil reparar o dano, pois é difícil delimitar claramente o dano e individualizar as pessoas afetadas pela medida discriminatória.

Uma possibilidade que tem que ser pensada seriamente, portanto, é a solução regulatória, com limitação e testagem prévia de determinadas formas de IA, e documentação da operação de aplicações mais delicadas de algoritmos (como é o caso da seleção e do recrutamento), o que pode ser feito e regulado por uma nova agência regulatória ou órgão criado especificamente para esse fim, ou aproveitando-se de uma estrutura já existente, como a Autoridade Nacional de Proteção de Dados.

5. CONCLUSÕES

Após análise do estado da arte em relação a algoritmos de seleção e recrutamento no Brasil e no mundo, foi possível perceber que existem vieses discriminatórios nesta aplicação de IA, impactando negativamente grupos historicamente discriminados, como negros e mulheres. Neste sentido, a produção doutrinária majoritária aponta para maior necessidade de transparência, de maiores controles e de supervisão humana em relação a essa modalidade algorítmica – justamente para que sejam minimizados, evitados ou reparados esses riscos discriminatórios e suas consequências.

As regras do direito penal clássico – que demandam reponsabilidade subjetiva – criam diversas barreiras para o estabelecimento de responsabilidade penal para as pessoas envolvidas na operação e no desenvolvimento de algoritmos. Isso porque, por um lado, a individualização de conduta intencional de agentes é de difícil caracterização, e, por outro, pela dificuldade de concretizar propostas como a responsabilização penal das próprias tecnologias de IA, como se fossem entes autônomos. Surgem, nesse contexto, cenários de defesa de soluções de responsabilidade penal das pessoas jurídicas ou modelos de exclusão do direito penal e aplicação de direito regulatório em face dessas tecnologias.

Contudo, no ordenamento jurídico brasileiro não existe a previsão de responsabilidade penal de pessoas jurídicas – à exceção de crimes ambientais e contra o sistema financeiro nacional – e a responsabilidade penal de pessoas físicas só ocorre de forma subjetiva, o que cria amplas limitações para a caracterização do racismo algorítmico enquanto crime de racismo previsto no art. 4º da Lei n.º 7.716/1989. Ainda sobre a legislação penal de combate ao racismo, as criminalizações se referem a formas de discriminação direta, embora o racismo ocorra majoritariamente de forma velada – indireta – em razão de sua condição de fenômeno estrutural.

Em decorrência disso, na etapa de análise de casos de racismo algorítmico, esperava-se encontrar amplas dificuldades de caracterização deste crime no que se relaciona às atividades de seleção e recrutamento. Esta hipótese foi, em grande medida, confirmada, uma vez que a necessidade de existência de dolo (elemento subjetivo) e sua correspondente comprovação é ainda mais complexa em relação ao contexto de tecnologia – que conta com uma pluralidade de pessoas físicas e jurídicas envolvidas.

No entanto, no caso de conhecimento do caráter discriminatório da IA de seleção e recrutamento por parte dos agentes envolvidos, houve refutação parcial dessa hipótese – chegando-se à conclusão da possibilidade de caracterização do crime previsto no art. 4 da Lei n.º 7.716/1989, na modalidade de dolo eventual. Essa possibilidade em conjunto com a datificação – que pode advir do emprego de IA nas atividades de seleção e recrutamento, se empregadas corretamente – podem funcionar como formas de limitar ou, ainda, de tornar mais transparente o combate ao racismo algoritmo e mais viável a punição criminal, ainda que de forma excepcional.

A partir disso, foram debatidos modelos para dar aplicabilidade ao combate ao racismo, especialmente em sua modalidade algorítmica. Apesar da existência de outras possibilidades, o modelo defendido no presente trabalho se baseia no conceito de direito de intervenção especial de Hassemer, que defende que os riscos antigos – tutelados pelo direito penal clássico – devem ser mantidos dentro deste, e os riscos da modernidade devem ser colocados em um direito que não emprega a privação de liberdade e busca – fora do direito penal –, sem as amplas limitações e garantias deste campo do direito, responsabilizar agentes (pessoas físicas e jurídicas) que cometam essas violações.

No mesmo sentido, o estado da arte sobre responsabilidade penal e IA demonstrou que a solução regulatória – e não penal – é amplamente defendida, havendo diversos modelos propostos, dentre os quais o de responsabilidade civil objetiva – cumulada ou não com mecanismos de seguro – e outro de regulação prévia com limitação e controle de determinadas aplicações críticas de algoritmos, como os de seleção e recrutamento – este último modelo é o previsto – por exemplo – no Artificial Intelligence Act da Comissão Europeia.

Sobre essas propostas, e apesar de um modelo de responsabilidade objetiva ser interessante para a regulação de algoritmos de uma forma geral, no caso de algoritmos de seleção e recrutamento essa solução pode ser insuficiente ante as amplas dificuldades de individualização dos envolvidos – não somente das pessoas físicas e jurídicas responsáveis pelo dano, o que poderia ser resolvido por este proposto regime de responsabilização objetiva –, mas por uma dificuldade de compreender quais seriam os danos, sua extensão e quem seriam as vítimas desses danos, que receberiam, portanto, eventuais reparações destes decorrentes.

Em um cenário em que é difícil determinar a extensão dos danos, identificar as pessoas afetadas e reparar esses mesmos danos, a melhor solução é evitar que eles sejam realizados e manter controles e registros para sua identificação. Neste contexto, as soluções regulatórias que parecem mais adequadas aos algoritmos de seleção e recrutamento são as de regulação prévia e de controle de segurança dessa modalidade de algoritmos, seja por meio da criação de uma nova agência reguladora ou órgão executivo para este cenário de inteligência artificial, seja pelo aproveitamento da estrutura de um órgão já existente, como a Autoridade Nacional de Proteção de Dados, que possui vocação tecnológica desde seu nascimento.

Acerca dos motivos que nos levaram à defesa da solução regulatória, este trabalho não defende o completo afastamento do direito penal em relação a algoritmos de seleção e recrutamento. O que se entende por inadequado é uma aplicação demasiadamente ampla e sem critérios de formas de reponsabilidade penal objetiva, o que vai contra os limites filosóficos que o Iluminismo impõe a essa grave intervenção do estado na vida do jurisdicionado, e que não pode ser feito pelo medo de que ferramentas penais de segunda velocidade se convertam em instrumentos de terceira velocidade, cujo uso (e abuso) são tendências no mundo contemporâneo.

Neste sentido, nos casos em que a intencionalidade dos agentes estiver caracterizada, podendo-se atribuir reponsabilidade penal subjetiva, deve ser aplicado o direito penal, respeitadas suas regras de aplicação e as garantias inerentes a este ramo do direito. No entanto – e é nesse contexto em que a defesa da solução regulatória é feita – quando os modelos de responsabilidade penal vigentes se mostrarem insuficientes, é mais fácil e mais adequado sair do direito penal do que desfigurá-lo, o que pode permitir indesejada escalada autoritária do estado por meio de sua indevida ampliação.

Dessa forma, este trabalho, nas conclusões precárias que apresenta, não buscou encerrar qualquer debate sobre a responsabilidade penal por inteligência artificial, mas – de forma exatamente oposta – ampliá-lo. Assim, a partir da constatação de que a produção acadêmica existente trata, em sua ampla maioria, de formas de IA com correspondência no mundo físico (como os carros autônomos), há necessidade de maior produção acadêmica sobre a responsabilidade penal decorrente de algoritmos que não possuem materialização física e que podem ter consequências maiores e menos perceptíveis – como os empregados em atividades de seleção e recrutamento.

No entanto, uma solução enfaticamente defendida neste trabalho é a necessidade de que a regulação supra o lapso de responsabilidade, e não o direito penal, nos moldes de propostas

legislativas como o AI Act ou o Anteprojeto de Marco Regulatório da IA¹³² apresentados pela Comissão de Juristas instituída pelo Senado Federal.

¹³² Sobre o qual não há maiores comentários por ter sido apresentado após a finalização deste trabalho, mas cujas soluções propostas a doutrina certamente terá que se debruçar.

REFERÊNCIAS

- ALI, M. et al. Discrimination through optimization: How Facebook's ad delivery can lead to biased outcomes. **Proceedings of the ACM on Human-Computer Interaction**, v. 3, n. CSCW, 1 nov. 2019.
- ALMEIDA, Silvio Luiz de. **Racismo estrutural**. São Paulo: Pólen, 2019.
- ARANHA, M. I. **Manual de Direito Regulatório**. 3ª ed., London: Laccademia Publishing, 2015.
- BARBOSA, P. A. R. Corporate Criminal Law, Artificial Intelligence and Big Data: the Huawei case and its implications for global society. **Revista Paradigma**, v. 29, n. 1, p. 19–35, 2020.
- BAROCAS, S.; SELBST, A. D. Big Data's Disparate Impact. **California Law Review**, v. 104(3), 2016, p. 671-732.
- BECK, U. Risk Society: **Towards a New Modernity**. Translated by Mark Ritter. London: SAGE Publications, 1992.
- BELLO FILHO, N. DE B. Responsabilidade criminal da pessoa jurídica por crimes contra o ambiente: um balanço após 27 anos de Constituição. **Revista de Direito Ambiental**, v. 80, n. 20, p. 63–90, 2015.
- BITENCOURT, C. R. **Tratado de Direito Penal - Parte Geral 1**. 15a. ed. São Paulo: Saraiva, 2010.
- BONILLA-SILVA, Eduardo. **Racism Without Racists: Color-Blind Racism and the Persistence of Racial Inequality in America**. 5th ed. Lanham: ROWMAN & LITTLEFIELD, 2017.
- BORGES, G. S.; FILÓ, M. D. C. S. Inteligência artificial, gênero e direitos humanos: o caso Amazon. **Justiça do Direito**, v. 35, n. 3, p. 218–245, 2021.
- BRADFORD, A. The Brussels Effect. **Northwestern University Law Review**, v. 107, n. 1, p. 1–68, 2012.
- BRASIL, 1951. **Lei nº 1.390**, de 3 de julho de 1951. Inclui entre as contravenções penais a prática de atos resultantes de preconceitos de raça ou de côr. Rio de Janeiro, DOU de 10.7.1951.
- BRASIL, 1985. **Lei nº 7.437**, de 20 de dezembro de 1985. Inclui, entre as contravenções penais a prática de atos resultantes de preconceito de raça, de cor, de sexo ou de estado civil, dando nova redação à Lei nº 1.390, de 3 de julho de 1951 - Lei Afonso Arinos. Brasília, DOU de 23.12.1985.
- BRASIL. **Constituição Federal**. 1988.
- BRASIL, 1989. **Lei nº 7.716**, de 5 de janeiro de 1989. Define os crimes resultantes de preconceito de raça ou de cor. Brasília, DOU de 6.1.1989 e retificada em 9.1.1989.
- BRASIL, 2017. Supremo Tribunal Federal. Tribunal Pleno. **ADI 1.055**. DJe-168, DIVULG 31-07-2017, PUBLIC 01-08-2017.

BRODSKY, J. S. Autonomous Vehicle Regulation How an Uncertain Legal Landscape May Hit the Brakes On Self-Driving Cars. **Berkeley Tech Law Journal**, v. 31, n. 1, p. 851–878, 2016.

BRUNETTA, C.; LEITÃO, A. S.; DIAS, E. R. Da pessoa e para a pessoa: a regulação jurídica dos algoritmos. **Revista de Informação Legislativa**, v. 59, n. 233, p. 163–178, 2022.

BUOLAMWINI, J.; GEBRU, T. GENDER SHADES: INTERSECTIONAL ACCURACY DISPARITIES IN COMMERCIAL. **Proceedings of Machine Learning Research**, v. 81, p. 1–15, 2018.

CELESTE, E. Digital constitutionalism: a new systematic theorization. **International Review of Law, Computers & Technology**, v. 33, n. 1, p. 76–99, 2019.

CHAVES JUNIOR, A.; BERZAGUI, B. Máquinas autônomas e responsabilidade no âmbito do direito penal: uma análise a partir da teoria da ação significativa. **Revista Eletrônica Direito e Política**, Programa de Pós-Graduação Stricto Sensu em Ciência Jurídica da UNIVALI, v. 16, n. 3, p. 1131–1165, 2021.

COSTA, Helena Regina Lobo da. **Direito penal econômico e direito administrativo sancionador: ne bis in idem** como medida de política sancionadora integrada. 2013. Tese (Livre Docência) – Universidade de São Paulo, São Paulo, 2013.

COSTA, R. DE O. Responsabilidade penal ambiental da pessoa jurídica: uma análise crítica do RE 548.181/PR. **Revista de Direito Ambiental**, v. 79, n. 20, p. 231–246, 2015.

CROOTOF, R. War torts: accountability for autonomous weapons. **University of Pennsylvania Law Review**, v. 164, n. 6, p. 1347–1402, 2016.

DATTA, A. et al. Discrimination in Online Advertising a Multidisciplinary Inquiry. **Proceedings of Machine Learning Research**, v. 81, n. Section 3, p. 1–15, 2018.

DE MAURO, A.; GRECO, M.; GRIMALDI, M. A formal definition of Big Data based on its essential features. **Library Review**, v. 65, n. 3, p. 122–135, 2016.
Direito, 2020.

DIVINO, S. Responsabilidade penal de Inteligência Artificial: uma análise sob a ótica do naturalismo biológico de John Searle. **Revista Brasileira de Ciências Criminais**, v. 171, p. 153–183, 2020.

DOTTI, R. A. A incapacidade criminal da pessoa jurídica. **Revista Brasileira de Ciências Criminais**, v. 11, n. 201, p. 184–207, 1995.

DUARTE, A.; NEGÓCIO, R.V. Todos são iguais perante o algoritmo? Uma resposta cultural do Direito à discriminação algorítmica. **RDP**, Brasília, v. 18, n. 100, p. 218-244, out./dez. 2021.

DUARTE, E. C. P.; QUEIROZ, M. V. L. A Revolução Haitiana e o Atlântico Negro: o constitucionalismo em face do lado oculto da modernidade. **Direito, Estado e Sociedade**, v. 49, p. 10-42, 2016.

ESTELLITA, H.; LEITE, A. **Veículos autônomos e Direito Penal: uma introdução**. In: ESTELLITA; LEITE (org.). **Veículos Autônomos e Direito Penal**. São Paulo: Marcial Pons, 2019.

EUROPEAN COMMISSION. **Artificial Intelligence Act**. 2021.

FARNOSIER, M. DE O. Criminal Liability and artificial intelligence. **Revista Brasileira de Ciências Criminais**, v. 187, n. 30, p. 153–170, 2022.

FERRAJOLI, L. **Direito e razão: teoria do garantismo penal**. São Paulo: Revista dos Tribunais, 2002.

FLORIDI, L. et al. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. **Minds and Machines**, v. 28, n. 4, p. 689–707, 2018.

FLORIDI, L. The End of an Era: from Self-Regulation to Hard Law for the Digital Industry. **Philosophy and Technology**, v. 34, n. 4, p. 619–622, 2021.

FONSECA, R. S.; FONSECA, R. C. **Concretização do Princípio *ne bis in idem* no Processo Antitruste Sancionador**. In: Uma vida dedicada ao direito: estudo em homenagem a Roberto Rosas. Coordenação: Arruda Alvim, Eduardo Arruda Alvim, Flávio Galdino. Rio de Janeiro: GZ, 2020.

GAY, D. S.; KAGAN, A. M. Big Data and Employment Law: What Employers and Their Legal Counsel Need to Know. **ABA Journal of Labor & Employment Law**, v. 33, n. 2, p. 191–210, 2018.

GILL, L.; REDEKER, D.; GASSER, U. Towards digital constitutionalism? Mapping attempts to craft an Internet Bill of Rights. **International Communication Gazette**, v. 80, n. 4, p. 302–319, 2018.

GLESS, S.; SILVERMAN, E.; WEIGEND, T. If robots cause harm, who is to blame? Self-driving cars and criminal liability. **New Criminal Law Review**, v. 19, n. 3, p. 412–436, 2016.

GRECO, L. *et al.* **Autoria como domínio do fato: estudos introdutórios sobre o concurso de pessoas no direito penal brasileiro**. São Paulo: Marcial Pons, 2014.

GRECO, L. **Um panorama da teoria da imputação objetiva**. 4a. ed. São Paulo: Revista dos Tribunais, 2014.

GUARDIA, G. E. R. S. Princípios processuais no direito administrativo sancionador: um estudo à luz das garantias constitucionais. **Revista da Faculdade de Direito**, Universidade de São Paulo, v. 109, p. 773-793, 2014.

GUIMARÃES, R. R. C. A Inteligência Artificial e a disputa por diferentes caminhos em sua utilização preditiva no processo penal. **Revista Brasileira de Direito Processual Penal**, v. 5, n. 3, p. 1555–1588, 2019.

HASSEMER, W. **Direito Penal libertário**. [Trad. Regina Greve] Belo Horizonte: Del Rey, 2007.

HASSEMER, W. **Direito Penal: fundamentos, estrutura, política**. [Trad. Pablo Rodrigo Alflen da Silva] Porto Alegre: Sérgio Antônio Fabris Editor, 2008.

HOFFMANN-RIEM, W. **Teoria geral do direito digital: transformação digital: desafios para o direito**. 2. ed. Rio de Janeiro: Forense, 2022.

KELLNER, A. L. M. **Direito regulatório, risco e incerteza econômica**. In: LEAL, F.; MENDONÇA, J. V. S. (org.). Transformações do direito administrativo: debates e estudos empíricos em direito administrativo e regulatório. Rio de Janeiro: FGV Direito Rio, 2022, p. 10-32.

KIM, P. T. Manipulating Opportunity. **Virginia Law Review**, v. 106, n. 4, p. 867–935, 2020.

KREBS, P. A responsabilização penal da pessoa jurídica e a suposta violação do direito penal mínimo. **Revista Estudos Jurídicos**, v. 33, n. 88, p. 117–126, 2000.

KROLL, J. A. The fallacy of inscrutability. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 376, n. 2133, 2018.

LAMBRECHT, A.; TUCKER, C. Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of stem career ads. **Management Science**, v. 65, n. 7, p. 2966–2981, 2019.

LIMA, Bruna Dias Fernandes. **Racismo algorítmico: o enviesamento tecnológico e o impacto aos direitos fundamentais no Brasil**. Dissertação (Mestrado em Direito), Programa de Pós-Graduação em Direito, Universidade Federal de Sergipe. São Cristóvão, 127 f., 2022.

LOBATO, J. D. T. Responsabilidade penal da pessoa jurídica - uma inconsistência dogmática e de princípios. **Revista da EMERJ**, v. 13, n. 50, p. 268–286, 2020.

LUZÓN PEÑA, Diego-Manuel. Reflexões sobre o dolo e o dolo eventual. Trad. Eduardo Viana. **Revista de Estudos Criminais**, Porto Alegre, v. 19, n. 77, p. 35-60, 2020.

MACHADO, M. R. A.; SANTOS, N. N. S.; FERREIRA, C. C. Legislação antirracista punitiva no Brasil: uma aproximação à aplicação do direito pelos Tribunais de Justiça brasileiros. **Revista de Estudos Empíricos em Direito**. V. 2, n. 1, p. 60-92, 2015.

MARQUES, F.; MARTINEZ NETO, A. A. Vieses algorítmicos, direitos fundamentais e os sindicatos. **Revista Direito do Trabalho e Sefuridade Social**, v. 222, p. 201–219, 2022.

MARTINS, S. P. R. A Responsabilidade Penal da Pessoa Jurídica nos Crimes Ambientais e a Teoria da Dupla Imputação. **Rev. Ciênc. Juríd. Empres.**, v. 17, n. 2, p. 166–176, 2016.

MAZZUOLI, V. D. O. Teoria geral do controle de convencionalidade no direito brasileiro. **Revista de Informação Legislativa**, v. 181, p. 113–139, 2009.

MENDES, G. F.; FERNANDES, V. O. **Constitucionalismo digital e jurisdição constitucional: uma agenda de pesquisa para o caso brasileiro**. Revista Brasileira de Direito, Passo Fundo, vol. 16, n. 1, p. 1-33, Janeiro-Abril, 2020.

MENDES, L.S.; MATTIUZZO, M. Discriminação Algorítmica: Conceito, Fundamento Legal e Tipologia. **RDU**, Porto Alegre, v. 16, n. 90, p. 39-64, nov.-dez. 2019.

MENTOR, D. **Teoria do direito de intervenção: a alternativa de Winfried Hassemer à inflação dos crimes econômicos**. Rio de Janeiro: Lumen Juris, 2016.

NAÇÕES UNIDAS. **Convenção Internacional sobre a eliminação de todas as formas de**

Discriminação Racial. 1963.

NAÇÕES UNIDAS. **Declaração universal dos Direitos Humanos.** 1948.

NUCCI, Guilherme de Souza. **Leis Penais e Processuais Penais Comentadas.** 5ª ed., São Paulo: RT, 2010.

O'NEIL, C. **Weapons of Math Destruction: how Big Data increases inequality and threatens democracy.** New York: Crown, 2016.

OLIVEIRA, M. A. B. DE; LOPES, I. M. P. **Os princípios norteadores da proteção de dados no Brasil.** In: TEPEDINO et al (coord.). Lei Geral de Proteção de Dados Pessoais e suas repercussões no Direito brasileiro. 2ª ed., São Paulo.: Thomson Reuters Brasil, 2020.

OSÓRIO, F. M. **Direito administrativo sancionador.** 3a. ed. São Paulo: Thomson Reuters Brasil, 2020.

PARREIRA, A. C. R. Revolução digital e a relevância da transparência algorítmica para as relações de trabalho. **Revista do Tribunal Regional Federal da 3a Região**, v. 66, n. 102, p. 315–329, 2020.

PASQUALE, F. Platform neutrality: Enhancing freedom of expression in spheres of private power. **Theoretical Inquiries in Law**, v. 17, n. 2, p. 487–513, 2016.

PASQUALE, F. **The Black Box Society.** Cambridge: Harvard University Press, 2015.

PIOVESAN, Flávia. **Direitos humanos e o direito constitucional internacional.** 20ª ed. São Paulo: Saraiva Jur, 2022.

PIRES FINCATO, D.; WÜNSCH, G. Subordinação Algorítmica: Caminho Para O Direito Do Trabalho Na Encruzilhada Tecnológica? Algorithmic Subordination: Path To Labor Law in Technological Crossroads? **Rev. TST**, v. 86, p. 3, 2020.

RIPOLLÉS, J. L. D. **A política criminal na encruzilhada.** Trad. André Luís Callegari. Porto Alegre: Livraria do Advogado Editora, 2015.

RODRIGUES, A. M. **Inteligência Artificial no Direito Penal – a Justiça Preditiva entre a Americanização e a Europeização.** In: RODRIGUES (coord.). **Inteligência Artificial no Direito Penal.** Coimbra: Almedina, 2020.

RODRIGUES, Gustavo Távora. O racismo escondido sob o manto da Lei. **Revista Direito e Práxis**, vol. 3, n. 5, 2012.

RODRIGUES, Luis Fernando Matricardi. **O STF às voltas com a “Navalha de Ockham”: uma “Proibição de Insuficiência” como Controle de Proporcionalidade das Omissões?.** In: VOJVODIC, Adriana; PINTO, Henrique Motta; GORZONI, Paula; SOUZA, Rodrigo Pagani de, (Org.). **Jurisdição Constitucional no Brasil.** São Paulo: Malheiros Editores, 2012, p. 317-334

ROXIN, C. **Estudos de Direito Penal.** Trad. Luís Greco. Rio de Janeiro: Renovar, 2008.

ROXIN, C. Sobre a normatização do dolo eventual e a doutrina do perigo doloso. Trad. Eduardo

Viana. **Revista de Estudos Criminais**, Porto Alegre, v. 19, n. 79, p. 7-27, 2020.

SANTOS, J. C. **Direito Penal**. 3a. ed. Curitiba: Lumen Juris, 2008.

SCHEAU, M.-C.; ARSENE, A.-L.; POPESCU, G. Artificial Intelligence / Machine Learning Challenges and Evolution. **International Journal of Information Security and Cybercrime**, v. 7, n. 1, p. 11–22, 2018.

SCHROLL, Carrie. Splitting the Bill: Creating a National Car Insurance Fund to Pay for Accidents in Autonomous Vehicles. **Northwestern University Law Review**, vol. 109, no. 3, Spring 2015, pp. 803-834.

SILVA SÁNCHEZ, J. M. **La expansión del Derecho Penal: Aspectos de la política criminal en las sociedades postindustriales**. 2a. ed. Madrid: Civitas, 2001.

SILVA SÁNCHEZ, J. M.; Urbina Gimeno, I. O. **Introducción al Derecho Penal Económico-Empresarial**. In: CRESPO, E. D. (director); CALATAYUD, M. M. (org.). *Crisis financiera y Derecho penal económico*. Madrid: EDISOFER, 2014.

SILVA, Eliezer Gomes da; SFOGGIA, Ivonei. O crime de racismo na legislação penal brasileira: passado, presente e futuro. **Revista Igualdade**, v. XIV, jan.-mar. 1997.

SILVA, Tarcízio da. Visão computacional e racismo algorítmico: branquitude e opacidade no aprendizado de máquina. **Revista da ABPN**, v. 12, n. 31, dez. 2019–fev. 2020b, p. 428-448.

SILVA, Tarcízio. **Racismo Algorítmico em Plataformas Digitais: microagressões e discriminação em código**. In: SILVA, Tarcízio (org.). *Comunidades, algoritmos e ativismos digitais: Olhares afrodiaspóricos*. São Paulo: LiteraRUA, 2020a, p. 120-137.

SILVEIRA, Fabiano Augusto Martins. **Da criminalização do racismo: aspectos jurídicos e sociocriminológicos**. Belo Horizonte: Del Rey, 2006.

SOUSA, S. A. DE. “**Não fui eu, foi a máquina**”: **Teoria do Crime, Responsabilidade e Inteligência Artificial**. In: RODRIGUES (coord.). *Inteligência Artificial no Direito Penal*. Coimbra: Almedina, 2020.

SRNICEK, Nick. **Platform capitalism**. New Jersey: John Wiley & Sons, 2017.

SUXBERGER, A.; FILHO, D. Funcionalização e expansão do Direito Penal: o Direito Penal negocial. **Revista de Direito Internacional**, v. 13, n. 1, p. 377–396, 5 ago. 2016.

SUZOR, N. Digital Constitutionalism: Using the Rule of Law to Evaluate the Legitimacy of Governance by Platforms. **Social Media and Society**, v. 4, n. 3, 2018.

TAVARES, Juarez. Espécies de dolo e outros elementos subjetivos do tipo. **Revista da Faculdade de Direito UFPR**, v. 14, 1971.

TUTT, A. An FDA for algorithms. **Administrative Law Review**, v. 69, n. 1, p. 83–123, 2017.

UNITED STATES. 9th Circuit. **Fair Hous. Council v. Roommate.com. LLC** - 666 F.3d 1216, 2012.

UNITED STATES. 9th Circuit. **HomeAway.com v. City of Santa Monica**. No. 18-55367, 2019.

UNITED STATES. US Supreme Court. **International Brotherhood of Teamsters v. United States**. 431 U.S. 324, 365, 1977.

VARELA, L. Nuevas habituaciones sociales al riesgo tecnológico y responsabilidad penal individual. **Revista Electrónica de Ciencia Penal y Criminología**, v. 23, n. 6, p. 1–29, 2021.

VASCONCELOS, V. G. **Acordo de não persecução penal**. São Paulo: RT, 2022.

WARREN, A.; HILLAS, A. Lethal autonomous robotics: rethinking the dehumanization of warfare. **UCLA Journal of International Law and Foreign Affairs**, v. 22, n. 2, p. 218–249, 2018.

WELZEL, H. O novo sistema jurídico-penal. [Trad. Luiz Regis Prado] 4a ed., São Paulo: RT, 2015.

WUNDERLICH, A.; RUIVO, M. A. **Parecer CASO “BOATE KISS” – SANTA MARIA /RS:** Culpa consciente e dolo eventual e a impossibilidade de reexame probatório em recursos aos tribunais superiores. In: WUNDERLICH; RUIVO; CARVALHO. **Dolo Eventual: imputação e determinação da pena – estudos sobre o caso Boate Kiss**. Porto Alegre: Livraria do Advogado, 2022.

YADAV, Y. The Failure of Liability in Modern Markets. **Virginia Law Review**, v. 102, n. 4, p. 1031–1100, 2016.

YOSHIKAWA, Jin. Sharing the Costs of Artificial Intelligence: Universal No-Fault Social Insurance for Personal Injuries. **Vanderbilt Journal of Entertainment & Technology Law**, vol. 21, no. 4, Summer 2019, pp. 1155-1188.

ZAFFARONI, Eugenio Raúl. Culpabilidad por la vulnerabilidad. **Revista anthropos: Huellas del conocimiento**. n. 204, p. 136-151, 2004.