

VI ENCONTRO VIRTUAL DO CONPEDI

DIREITO, GOVERNANÇA E NOVAS TECNOLOGIAS II

DANIELLE JACON AYRES PINTO

EDSON RICARDO SALEME

FERNANDO GALINDO AYUDA

Todos os direitos reservados e protegidos. Nenhuma parte destes anais poderá ser reproduzida ou transmitida sejam quais forem os meios empregados sem prévia autorização dos editores.

Diretoria - CONPEDI

Presidente - Prof. Dr. Orides Mezzaroba - UFSC - Santa Catarina

Diretora Executiva - Profa. Dra. Samyra Haydêe Dal Farra Naspolini - UNIVEM/FMU - São Paulo

Vice-presidente Norte - Prof. Dr. Jean Carlos Dias - Cesupa - Pará

Vice-presidente Centro-Oeste - Prof. Dr. José Querino Tavares Neto - UFG - Goiás

Vice-presidente Sul - Prof. Dr. Leonel Severo Rocha - Unisinos - Rio Grande do Sul

Vice-presidente Sudeste - Profa. Dra. Rosângela Lunardelli Cavallazzi - UFRJ/PUCRio - Rio de Janeiro

Vice-presidente Nordeste - Profa. Dra. Gina Vidal Marcilio Pompeu - UNIFOR - Ceará

Representante Discente: Prof. Dra. Sinara Lacerda Andrade - UNIMAR/FEPODI - São Paulo

Conselho Fiscal:

Prof. Dr. Caio Augusto Souza Lara - ESDHC - Minas Gerais

Prof. Dr. João Marcelo de Lima Assafim - UCAM - Rio de Janeiro

Prof. Dr. José Filomeno de Moraes Filho - Ceará

Prof. Dr. Lucas Gonçalves da Silva - UFS - Sergipe

Prof. Dr. Valter Moura do Carmo - UNIMAR - São Paulo

Secretarias

Relações Institucionais:

Prof. Dra. Daniela Marques De Moraes - UNB - Distrito Federal

Prof. Dr. Horácio Wanderlei Rodrigues - UNIVEM - São Paulo

Prof. Dr. Yuri Nathan da Costa Lannes - Mackenzie - São Paulo

Comunicação:

Prof. Dr. Liton Lanes Pilau Sobrinho - UPF/Univali - Rio Grande do Sul

Profa. Dra. Maria Creusa De Araújo Borges - UFPB - Paraíba

Prof. Dr. Matheus Felipe de Castro - UNOESC - Santa Catarina

Relações Internacionais para o Continente Americano:

Prof. Dr. Heron José de Santana Gordilho - UFBA - Bahia

Prof. Dr. Jerônimo Siqueira Tybusch - UFSM - Rio Grande do Sul

Prof. Dr. Paulo Roberto Barbosa Ramos - UFMA - Maranhão

Relações Internacionais para os demais Continentes:

Prof. Dr. José Barroso Filho - ENAJUM

Prof. Dr. Rubens Beçak - USP - São Paulo

Profa. Dra. Viviane Coêlho de Séllos Knoerr - Unicritiba - Paraná

Eventos:

Prof. Dr. Antônio Carlos Diniz Murta - Fumec - Minas Gerais

Profa. Dra. Cinthia Obladen de Almendra Freitas - PUC - Paraná

Profa. Dra. Livia Gaigher Bosio Campello - UFMS - Mato Grosso do Sul

Membro Nato - Presidência anterior Prof. Dr. Raymundo Juliano Feitosa - UMICAP - Pernambuco

D597

Direito, governança e novas tecnologias II [Recurso eletrônico on-line] organização CONPEDI

Coordenadores: Danielle Jacon Ayres Pinto; Edson Ricardo Saleme; Fernando Galindo Ayuda – Florianópolis; CONPEDI, 2023.

Inclui bibliografia

ISBN: 78-65-5648-746-5

Modo de acesso: www.conpedi.org.br em publicações

Tema: Direito e Políticas Públicas na era digital

1. Direito – Estudo e ensino (Pós-graduação) – Encontros Nacionais. 2. Direito. 3. Governança e novas tecnologias. VI Encontro Virtual do CONPEDI (1; 2023; Florianópolis, Brasil).

CDU: 34



VI ENCONTRO VIRTUAL DO CONPEDI

DIREITO, GOVERNANÇA E NOVAS TECNOLOGIAS II

Apresentação

Apresentação do CONPEDI – novas tecnologias.

O grupo constituído por DANIELLE JACON AYRES PINTO, FERNANDO GALINDO e EDSON R. SALEME presidiram o GT Direito, Governança e novas tecnologias II, que tiveram o privilégio de conduzir excelentes trabalhos apresentados, que apontaram as necessidades brasileiras mais prementes, em termos normativos, na era digital. Os trabalhos abordaram as características mais marcantes que estão sujeitos os dados, sobretudo em face da LGPD, mediante a apresentação de propostas para a governança democrática. Outros temas a destacar foram os relacionados ao uso de tecnologias da informação e comunicação nos julgados, bem como de que forma os tribunais brasileiros estão empregando programas de inteligência artificial e como se poderia encontrar limites a essa utilização.

O primeiro a apresentar o trabalho foi o doutorando Ronaldo Felix Moreira Junior acerca da disseminação de notícias falsas e os limites do uso de dados pessoais em campanhas eleitorais, que abarcou a LGPD discutindo como os dados pessoais sensíveis têm sido empregados para fins políticos, como instrumentos de ataque à democracia. O discente Lorenzo Borges de Pietro apresentou o trabalho denominado “A (in) constitucionalidade da suspensão de plataformas da internet em decorrência do descumprimento de decisão judicial: um debate a luz do princípio da proporcionalidade, discutindo o alcance das decisões judiciais em termo de internet. O tema entabulado no próximo artigo foi o “Colonialismo Digital e os entraves à proteção de direitos fundamentais na era do Capitalismo de Vigilância”, por Ronaldo Felix Moreira Junior, que apresentou o primeiro trabalho. Discutiuse que os dados pessoais foram incluídos no rol de direitos fundamentais e que grandes empresas, contratadas para lidar com dados pessoais, podem empregá-los a seu talante. Portanto, deve existir uma tecnologia própria para a proteção deles. Pedro Ribeiro Fagundes apresentou o trabalho acerca da importância da gestão de riscos para a motivação dos atos administrativos. Esta motivação, essencial em todo o ato, deve levar em consideração os riscos que o gestor pode incidir, bem como os respectivos prejuízos que esses riscos podem produzir. Tainara Conti Peres e Deise Marcelino da Silva apresentaram o trabalho “A LGPD e a sua adequação no ambiente laboral: sob a ótica de controle do empregador privado brasileiro.” As autoras inferem que a proteção de dados é própria desta época e abordaram, especificamente, as relações trabalhistas e analisam como se aplicam nas relações de trabalho, sobretudo sob a ótica do empregador privado. Valdir Rodrigues de Sá e Irineu

Francisco Barreto Júnior, que se encarregaram do tema “Liberdade de expressão nas plataformas digitais”, teve como objeto a análise da prática de crimes com a abertura da liberdade virtual existente no presente. O próximo trabalho apresentado por Gabrieli Santos Lacerda da Silva, dedicou-se ao tema “Os limites do consentimento frente ao direito fundamental de proteção dos dados pessoais”, que abordou a temática da mudança do comportamento humano diante dos avanços digitais. Nesse sentido, o grande volume de dados da internet, entre eles os dados pessoais, geram implicações na própria dinâmica social, o que fez a CF incluir dispositivos constitucionais e infraconstitucionais. Após a apresentação e aluna Triciele Radaelli Fernandes e Fernando Hoffmam trouxeram a temática “O capital e a(s) guerra(s) na era do capitalismo de vigilância e a constituição de tecnopolíticas de combate”. O trabalho reflete que pode ser uma guerra real ou de violência simbólica diante da existência de tecnologias que podem perpetuar ou resgatar fórmulas capitalistas existentes nas diversas zonas. A seguir passou-se a apresentar por Estella Ananda Neves o artigo “Análise econômica do impacto da inteligência artificial nos tribunais brasileiros.” O baixo nível de investimentos e a parca participação de empresas brasileiras refletem o desenvolvimento atual do país e afirmam que o Judiciário pode em muito auxiliar o aprimoramento do Brasil. O primeiro bloco finalizou com a apresentação do trabalho “Administração Pública na era digital: uma análise sobre a segurança de dados nas sociedades de economia mista e empresas públicas à luz da LGPD” apresentado por Jean Marcel dos Santos. Como proteger os dados no atual panorama. O primeiro bloco foi encerrado com considerações dos coordenadores do GT, sobretudo o Prof. Galindo, que observou a questão da vigilância de dados nos sistemas jurídicos, a exemplo do que se pode observar na legislação europeia, como a que estabelece regras acerca da inteligência artificial, cuja matéria continua sendo regulada pelo Parlamento Europeu que, no último 14 de junho de 2023, aprovou sua posição negociadora sobre a Lei de Inteligência Artificial. Importante recordar que esta norma inclui, entre os sistemas de alto risco os sistemas de IA que estão referidos na Administração de Justiça.

O segundo bloco de intervenções começou com o trabalho de Roseli Rêgo Santos Cunha Silva abordou no trabalho A LGPD e o tratamento de dados por agentes de pequeno porte: uma análise a partir da Resolução CD/ANPD N°2/2022. A abordagem indica que devem ser disponibilizados meios, compatíveis com as atividades de menor porte, considerando o bem que a LGPD objetiva proteger, a Resolução não exclui atores de menor porte; o discente Guilherme Elias Trevisan apresentou o trabalho “Big tech, dados, infraestruturas digitais e as universidades públicas federais brasileiras.” Restringiu-se a análise da verificação do sigilo da infraestrutura de dados e a disparidade de tecnologia que geram impactos geopolíticos, sobretudo nas universidades federais. Lidiana Costa de Sousa Trovão e Igor Marcellus Araujo Rosa apresentaram o trabalho intitulado “Cidades Inteligentes Sustentáveis,

governança e regulamentação de dados”; o trabalho analisa como essas cidades podem atingir o objetivo socioambiental e a quem são efetivamente destinadas. A seguir Luiz Fernando Mingati passou a expor o trabalho Constitucionalismo na era digital: os desafios impostos pela era informacional frente às garantias constitucionais. O artigo versa sobre como o impacto da era da informação e como ocorrem modificações na ordem interna geradas por esse fato. A seguir o Prof. Lucas Gonçalves da Silva apresentou juntamente com o aluno Reginaldo Felix “Tributação e Novas Tecnologias”, os autores indicam que há uma tributação apresenta um novo percalço pela falta de transparência que os entes tributantes possuem diante desta atividade. O próximo trabalho trouxe a temática “Das cortes físicas às cortes digitais: a transformação digital dos tribunais como instrumento de acesso à justiça”, pelo aluno Dennys Damião Rodrigues Albino; a temática se concentra na possibilidade de o Judiciário acompanhar a atual tendência digital e quais seriam as condicionantes a essas mudanças. A seguir David Elias Cardoso Camara apresentou o trabalho “Software de decisão automatizada como ferramenta de compliance no Tribunal de Justiça do Maranhão.” O artigo estabelece uma análise geral sobre alguns documentos do Banco Mundial que analisa algumas ineficiências do Poder Judiciário. A seguir o aluno Pedro Gabriel C. Passos analisa no artigo “Desafios para concretização do ODS 8: análise a partir da dinâmica da indústria 4.0” que trata das TICs no ambiente do trabalho e alguns fenômenos que este pode apresentar em termos de prestação de serviços no mundo digital. Thiago Leandro Moreno seguiu apresentando o trabalho “Direito e Tecnologia: criptoativos e tokens não fungíveis”, o trabalho versa sobre a ideia do metaverso e as transações ocorridas nos espaços virtuais. Novamente Irineu Francisco Barreto Jr e Kelly Cristina Maciel da Silva apresentaram o trabalho “O paradoxo entre a garantia constitucional do direito à informação e a preservação da privacidade em banco de dados públicos e privados.” Constata-se pelo artigo que não existe ainda proteção suficiente para eventuais ataques virtuais.

O último bloco iniciou-se com o artigo “Mercosul X União Europeia: necessária adequação da autoridade nacional de proteção de dados” por Bruno Alexander Mauricio e Kennedy Josué Grecca de Mattos. A seguir apresentou-se o artigo “Mitigação de vieses algorítmicos em processos decisórios: os impactos da diversidade na constituição de equipes desenvolvedoras de inteligência artificial”, por Airto Chaves Jr e Pollyanna Maria da Silva. O objetivo da investigação é verificar os impactos da constituição de equipes responsáveis pelas inteligências artificiais. Na sequência José Octávio de Castro Melo apresentou o trabalho “Novas tecnologias e regulação: uma análise do PL 872/2021 face ao dever de diligência do Estado na proteção do direito à privacidade.” A apresentação do trabalho “O uso da inteligência artificial no âmbito do processo judicial: desafios e oportunidades” por Jordy Arcadio Ramirez Trejo e Saulo Capelari Junior abordou de que forma deve ser implementada a inteligência artificial no âmbito do Poder Judiciário. A seguir Luciana

Cristina de Souza apresentou o trabalho “Risco no uso das inteligências artificiais e segurança digital” levando em consideração a atual forma que se aborda possíveis culpados com possível transgressão ao princípio da presunção de inocência. Na sequência, Thais Aline Mazetto Corazza, expôs o trabalho “Os riscos na tomada de decisões por máquinas”. Já existe, no âmbito dos tribunais, certa triagem para evitar repetições e assim proporcionar melhores benefícios. Deve-se ter cuidado ao aplicar essas ferramentas, pois possuem subjetividades complexas. Bruno Mello Corrêa de Barros Beuron apresentou o trabalho “Revolução tecnológica e sociedade pós-moderna: perspectivas da obsolescência programada e do direito do consumidor à luz da metateoria do direito fraterno” . Luciana Rodrigues dos Santos e Aparecida Moreira de Oliveira Paiva apresentaram o artigo “Risco no uso das inteligências artificiais e segurança digital” em que se observa a questão relacionada a inteligência artificial pelos órgãos públicos e as questões discriminatórias.

Ao final houve manifestação de todos relativamente ao conteúdo apresentado e o quanto enriquecedor o Grupo de Trabalho foi para todos com ponderações extremamente profícuas de todos os presentes.

**MITIGAÇÃO DE VIESES ALGORÍTMICOS EM PROCESSOS DECISÓRIOS: OS
IMPACTOS DA DIVERSIDADE NA CONSTITUIÇÃO DE EQUIPES
DESENVOLVEDORAS DE INTELIGÊNCIA ARTIFICIAL**

**MITIGATION OF ALGORITHMIC BIASES IN DECISION-MAKING PROCESSES:
THE IMPACTS OF DIVERSITY IN THE CONSTITUTION OF ARTIFICIAL
INTELLIGENCE DEVELOPMENT TEAMS**

**Pollyanna Maria Da Silva ¹
Aírto Chaves Junior ²**

Resumo

A utilização de algoritmos para tomada de decisão, apesar de extremamente atrativa em termos de eficiência e objetividade, também levanta preocupações em relação a possíveis vieses e discriminações, trazendo à tona discussões éticas e legais. Por isso, além de apresentar algumas maneiras de atenuar vieses algorítmicos, objetivou-se, com maior ênfase, demonstrar os impactos da diversidade na constituição de equipes responsáveis pelo desenvolvimento de inteligências artificiais de apoio a decisões humana. Para isso, utilizou-se o método indutivo, operacionalizado pela técnica da pesquisa bibliográfica. Iniciou-se abordando os vieses algorítmicos e sua relação com a temática da discriminação. Foram apresentados alguns tipos de discriminações algorítmicas e exemplos de preconceito algorítmico de raça, gênero e sexualidade. Em um segundo momento, enfocou-se na mitigação de vieses algorítmicos, sendo apresentadas estratégias como: transparência, auditoria, explicabilidade e regulação algorítmica. Verificou-se que a diversidade em equipes de desenvolvimento de IA é importante porque ajuda a garantir que os modelos sejam mais justos, precisos e úteis para todas as pessoas, independentemente de seu gênero, raça, etnia, orientação sexual ou outras características demográficas. Dessa forma, tendo como pano de fundo os valores e interesses da sociedade, acredita-se ser possível alcançar um juízo algorítmico não enviesado capaz de promover a justiça social.

Palavras-chave: Inteligência artificial, Viés, Viés algorítmico, Diversidade

Abstract/Resumen/Résumé

The use of algorithms for decision-making, despite being extremely attractive in terms of efficiency and objectivity, also raises concerns about possible biases and discrimination, bringing up ethical and legal discussions. Therefore, in addition to presenting some ways to mitigate algorithmic biases, the aim was, with greater emphasis, to demonstrate the impacts

¹ Doutoranda em Ciência Jurídica pela Universidade do Vale do Itajaí - UNIVALI. Mestre em Ciências Criminais pela Pontifícia Universidade Católica do Rio Grande do Sul - PUC-RS. Professora.

² Doutor em Ciência Jurídica pela Universidade do Vale do Itajaí – UNIVALI. Doutor em Direito pela Universidade de Alicante, Espanha. Advogado Criminalista. Professor.

of diversity in the constitution of teams responsible for the development of artificial intelligences to support human decisions. For this, the inductive method was used, operationalized by the technique of bibliographical research. It began by addressing algorithmic biases and their relationship with the issue of discrimination. Some types of algorithmic discrimination and examples of algorithmic prejudice based on race, gender and sexuality were presented. In a second moment, it focused on the mitigation of algorithmic biases, presenting strategies such as: transparency, auditing, explainability and algorithmic regulation. Diversity in AI development teams has been found to be important because it helps ensure that models are more fair, accurate, and useful to all people, regardless of their gender, race, ethnicity, sexual orientation, or other demographic characteristics. Thus, against the background of society's values and interests, it is believed to be possible to achieve an unbiased algorithmic judgment capable of promoting social justice.

Keywords/Palabras-claves/Mots-clés: Artificial intelligence, Bias, Algorithmic bias, Diversity

INTRODUÇÃO

Com a popularização da inteligência artificial e a disponibilidade de tecnologias inovadoras, o uso de algoritmos em processos decisórios está se tornando cada vez mais frequente em diversos setores da sociedade, incluindo finanças, saúde, educação, varejo, marketing, logística, recursos humanos e governança. O uso dessas ferramentas, especificamente pelo Poder Judiciário, é um assunto bastante sensível e controverso, sendo objeto de discussão em muitos países.

Apesar de extremamente atrativa em termos de eficiência e objetividade, a utilização de algoritmos para tomada de decisão também levanta preocupações em relação a possíveis vieses e discriminações, trazendo à tona discussões éticas e legais. Por isso, inicia-se abordando os vieses algorítmicos e sua relação com a temática da discriminação. Apresenta-se alguns tipos de discriminações algorítmicas e exemplos de preconceito algorítmico de raça, gênero e sexualidade. Ressalta-se que os vieses sociais e algorítmicos podem ser reciprocamente fortalecidos.

Diante dos impactos negativos que podem ser causados pelo uso da inteligência artificial no processo decisório, haja visto os vieses algorítmicos, mostra-se imperioso identificar formas de evitá-los. Atualmente, esta é uma tarefa política que envolve valores e interesses coletivos, devendo ser discutida em espaços públicos. Por isso, em um segundo momento, enfoca-se na mitigação de vieses algorítmicos, sendo apresentadas estratégias como: transparência, auditoria, explicabilidade e regulação algorítmica.

Além de apresentar algumas maneiras de atenuar vieses algorítmicos, objetiva-se, com maior ênfase, demonstrar os impactos da diversidade na constituição de equipes responsáveis pelo desenvolvimento de inteligências artificiais de apoio a decisões humana. Para isso, utiliza-se o método indutivo, operacionalizado pela técnica da pesquisa bibliográfica.

1 Vieses algorítmicos

À primeira vista, algoritmos podem parecer matematicamente excelentes. Em um segundo olhar, eles se mostram funcionalmente preocupantes. Tanto é que, Cathy O’Neil (2020) denomina modelos matemáticos nocivos de “Armas de Destruição Matemática”¹. Não

¹ Armas de destruição matemática (ADMs) é o nome criado por Cathy O’Neil (2020, p. 8) para designar modelos nocivos que tendem a punir pobres e oprimidos da sociedade, enriquecendo os mais ricos.

se pode ignorar os riscos associados ao uso de *softwares* de apoio a decisões. Um deles é o enviesamento algorítmico (FLORES, 2021, p. 99 – 113).

A palavra *bias* geralmente tem uma conotação negativa na língua inglesa. O viés é algo a ser evitado, ou que é necessariamente problemático. Entretanto, David Danks e Alex John London (2017. p. 4691- 4697) entendem o termo de uma maneira mais neutra, referindo-se, simplesmente, ao desvio de um padrão. Eles fazem uma distinção entre os vieses algorítmicos que são neutros ou inquestionáveis e aqueles que são problemáticos de alguma forma e exigem uma resposta. Em alguns casos, há ajustes tecnológicos ou algorítmicos que os desenvolvedores podem usar para “compensar” vieses problemáticos. Em outros casos, porém, as respostas requerem ajustes por parte do agente, seja humano ou sistema autônomo, que utiliza os resultados do algoritmo. Não existe uma solução “tamanho único” para o viés algorítmico.

Inclusive, há casos em que o viés algorítmico em uma dimensão pode contribuir para o desempenho adequado em uma outra mais importante, ou em que o viés ou desvio é importante para permitir que o sistema geral atinja os objetivos desejados, em conformidade com as normas éticas e legais. Para David Danks e Alex John London (2017. p. 4691- 4697), o “ecossistema” ao redor de um algoritmo contém muitas oportunidades tanto para a introdução de viés quanto para a sua “compensação”, no sentido de minimizar os danos (se houver) feitos pelo viés algorítmico.

Todavia, apesar de reconhecer a relevância e coerência do posicionamento dos autores, para fins dessa pesquisa, a categoria viés algorítmico² é interpretada como algo ruim que deve ser evitado - mesmo sabendo que a situação é significativamente mais complexa. Adota-se a expressão vieses algorítmicos para “[...] designar erros gerados por algoritmos, seus efeitos discriminatórios ou excludentes” (FERRARI, 2020. p. 85).

Estudos acerca do objeto viés algorítmico encontram-se em expansão na literatura acadêmica das Ciências Jurídicas e Sociais.³ Contudo, ainda se nota indefinição conceitual. O termo, geralmente, aparece relacionado à discriminação, hora como sinônimo, hora como um de suas causas.

Especificamente por discriminação algorítmica entende-se todas as “[...] distinções, preferências ou exclusões realizadas através de algoritmo capazes de afetar a igualdade de

² Como sinônimos, o termo *algorithmic bias* é muito utilizado. Eles também são denominados de *Coded Gaze*, olhar codificado, olhar viciado. (FERRARI, 2020. p. 85)

³ Em pesquisa publicada em 2020, com o objetivo de acompanhar o surgimento e os desdobramentos do debate sobre o viés algorítmico, constatou-se que 83,4% da produção científica sobre o tema ocorreu a partir de 2016. (SIMÕES-GOMES; ROBERTO; MENDONÇA, 2020, p.139-166)

tratamento ou de direitos entre seres humanos”. Ela pode ocorrer de forma direta ou indireta. A primeira ocorre quando um dado sensível é utilizado de forma intencional pelo agente. A segunda, quando há interferências com efeitos discriminatórios em grupos historicamente desfavorecidos (COSTA, 2022, p. 154).

Laura Schertel Mendes e Marcela Mattiuzzo (2019) sistematizam a discriminação algorítmica em quatro principais tipos: discriminação por erro estatístico⁴; discriminação por generalização⁵; discriminação pelo uso de informações sensíveis⁶ e discriminação limitadora do exercício de direitos⁷. Dessa forma, além de englobarem no termo “discriminação algorítmica”, situações estatisticamente inconsistentes, incluem também conjunturas estatisticamente lógicas, mas injustas.

Os vieses algorítmicos, como o próprio nome expressa, consistem em um enviesamento, de ordem estatística ou moral. Eles representam a discriminação decorrente da automatização de cálculos e do processamento automatizado de dados, ou seja, dos meios algorítmicos (SIMÕES-GOMES; ROBERTO; MENDONÇA, 2020, p.139-166).

Os vieses algorítmicos são graves “falhas” produzidas por algoritmos de IA ao gerarem discriminações sistemáticas e injustas. Caracterizados pelo comprometimento da neutralidade do sistema, eles refletem predileções humanas por determinados valores que podem decorrer do treinamento incorreto do algoritmo ou de desdobramentos inesperados do *machine learning*. Como exemplos de preconceito algorítmico de raça e gênero, cita-se: a) negros indevidamente classificados com altos índices de reincidência; b) preconceito no Google fotos; c) discriminação contra mulheres na concessão de crédito (ALVES; ANDRADE, 2021, p. 348 - 373).

Em relação ao viés preditivo racial, ganhou repercussão o caso do *software COMPAS*⁸ (*Correctional Offender Management Profiling for Alternative Sanction*), acusado pela instituição ProPublica de ser tendencioso em relação à réus negros. Após apuração da pontuação

⁴ Decorre basicamente de um erro cometido pelos engenheiros ou cientistas de dados responsáveis pelo desenho do algoritmo. Ex: dados coletados incorretamente e problemas no código do algoritmo.

⁵ Classificação equivocada de algumas pessoas em certos grupos, mesmo quando o algoritmo e as informações estejam corretos. O problema surge quando uma pessoa demonstra ser um caso atípico. Relaciona-se à natureza de exercícios probabilísticos.

⁶ Baseia-se em dados ou *proxies* legalmente protegidos.

⁷ O problema advém da relação entre a informação utilizada pelo algoritmo e a realização de um direito.

⁸ Ferramenta de gerenciamento de casos e suporte à decisão usada pelos tribunais dos EUA que, a partir da avaliação de 137 características do réu, avalia a probabilidade de reincidir dentro de 2 anos. O software foi desenvolvido em 1998 e, desde então, tem sido atualizado periodicamente para acompanhar os avanços tecnológicos. (SILVA; BERNARDES NETO, 2019. p. 163 - 181)

de risco de mais de 7.000 pessoas em Broward, na Flórida, nos anos de 2013 e 2014, foram identificadas disparidades raciais significativas. Devido à suscetibilidade da fórmula, réus negros foram apontados erroneamente como futuros criminosos em quase o dobro dos casos em comparação à réus brancos. O resultado foi possível pois os pesquisadores verificaram quantas pessoas foram acusadas de novos crimes nos dois anos seguintes (mesmo *benchmark* utilizado pelos criadores do algoritmo) (ANGWIN et al., 2016).

Acerca dos impactos raciais díspares dos algoritmos de avaliação de risco, Sandra Mayson (2019, p. 2218 - 2300) entende que o problema é a própria previsão que, buscando suposições para eventos futuros, olha para o passado.⁹ No âmbito dos sistemas de justiça criminal, por exemplo, corrigir as previsões acerca da disparidade racial exigirá mudanças na maneira como o se concebe e se responde aos possíveis riscos.¹⁰

No tocante à discriminação algorítmica por motivos relacionados à sexualidade, envolvendo mulheres, homossexuais e transgêneros, Diego Carneiro Costa (2020, p. 116 - 146) subdivide-os em três principais vieses: a) viés de gênero¹¹; b) viés relacionado à orientação sexual¹²; c) viés relacionado à identidade de gênero¹³.

Os algoritmos refletem a qualidade dos dados com os quais trabalham. Daí, infere-se três premissas: a) dados imperfeitos fazem com a algoritmos herdem os preconceitos manifestos em decisões anteriores (*data sets* viciados); b) dados podem refletir vieses generalizados da sociedade¹⁴; c) a mineração de dados pode descobrir padrões preexistente de exclusão e desigualdade.

⁹ “*In a racially stratified world, any method of prediction will project the inequalities of the past into the future. This is as true of the subjective prediction that has long pervaded criminal justice as it is of the algorithmic tools now replacing it. Algorithmic risk assessment has revealed the inequality inherent in all prediction, forcing us to confront a problem much larger than the challenges of a new technology. Algorithms, in short, shed new light on an old problem.*” (MAYSON, 2022, p. 2218)

¹⁰ “*As long as crime and arrest rates are unequal across racial lines, any method of assessing crime or arrest risk will produce racial disparity. The only way to redress the racial inequality inherent in prediction in a racially unequal world is to rethink the way in which contemporary criminal justice systems conceive of and respond to risk.*” (MAYSON, 2022. p. 2296)

¹¹ “[...] erros presentes nos algoritmos que podem acarretar um tratamento diferenciado comparativamente entre homens e mulheres. [...] quando tais vieses gerarem efeitos concretos violando o tratamento isonômico entre os gêneros estaremos diante de uma hipótese de discriminação algorítmica.” O viés de gênero pode ser identificado em modelos de incorporação de palavras (ex: *Google tradutor*), softwares de reconhecimento facial, algoritmos de reconhecimento de imagens, anúncios de emprego, recrutadores digitais e, em análises discriminatórias de currículos, por exemplo. (COSTA, 2022, p. 116)

¹² O viés relacionado à orientação sexual é fruto do padrão heteronormativo imposto socialmente. Manifesta-se nas distinções e diferenciações algorítmicas por conta de aspectos relacionados à sexualidade humana, especificamente em relação aos homossexuais. (COSTA, 2022, p. 132)

¹³ O viés relacionado a identidade de gênero é fruto do padrão binário e excludente adotado para a sexualidade que faz com que os transgêneros sejam mais suscetíveis às discriminações algorítmicas. (COSTA, 2022, p. 141)

¹⁴ “[...] mesmo que se use uma base de dados correta, pode haver um efeito discriminatório sendo absorvido pelo algoritmo, que corresponde, na verdade, a um *bias*, um preconceito já existente em nossa sociedade, que o

Vale mencionar ainda que, a discriminação resultante do uso de algoritmos, quase sempre, não representa uma escolha consciente de seus programadores. Talvez por isso, considera-se difícil identificar a origem do problema e explicá-lo. Tal dificuldade reflete-se no âmbito técnico, legal e político (BAROCAS; SELBST, 2016, p. 671-732). Por isso, é crucial a colaboração entre cientistas da computação, advogados, reguladores, fabricantes e agentes estatais.

Ferrari, Becker e Wolkart (2018, p.1-16) consideram perversos os efeitos do uso de algoritmos construídos a partir de *data sets* viciados. Frisa-se que a potencial aplicação em larga escala exacerba o problema. Além disso, a associação do emprego de algoritmos à pretensão científica do resultado obtido, torna o resultado mais facilmente aceitável, mesmo que inexistentes as discussões notáveis sobre a estrutura e o funcionamento do algoritmo.

Carlos Henrique Bath (2021, p. 39-68) difere um algoritmo enviesado de um algoritmo falho e argumenta que a mitigação de vieses é uma tarefa política, envolvendo valores e interesses públicos atuais.¹⁵ Para ele, o problema dos vieses está associado a questões não técnicas, devendo ser discutidas em espaços públicos. Onde, quando e como as pessoas querem fazer uso dos modelos computacionais? Em busca de respostas, propõe-se uma espécie de “governabilidade algorítmica”¹⁶. Para ele, um juízo algorítmico não enviesado é aquele que tem como pano de fundo os valores e interesses da comunidade. A regulação estabelecida pela sociedade faz com que o juízo não seja neutro nem desinteressado.

A opressão a grupos minoritário é uma característica eminentemente humana. O que a IA faz é expor este atributo, ainda não sendo capaz de corrigir discriminações e outras questões sociais profundas. Por isso, a automação dos processos decisórios pode estar reproduzindo decisões humanas negativas, caracterizadas por estereótipos e discriminações, decisões

algoritmo tende a cristalizar, já que ele sempre trabalha com o passado (ou seja, a partir de dados antigos).” (FERRARI, 2020, p. 91)

¹⁵ “A mitigação de vieses é possível, mas não é uma tarefa técnica. Não é algo que deva ficar exclusivamente nas mãos de engenheiros. A caracterização de um sistema como enviesado não se dá em função do modo como ele é computacionalmente modelado. Um algoritmo que não implemente um dado domínio fechado de modo correto não é enviesado, mas sim falho. [...] Ao engenheiro, deve caber somente a responsabilidade sobre a implementação computacional adequada de um dado domínio fechado. A constituição desse domínio, bem como a aceitação ou rejeição de seu uso para um dado fim, são temas de debate público. Deve o algoritmo tratar a todos de modo formalmente idêntico, ou deve ele compensar certas desigualdades materiais? A resposta a esse tipo de pergunta não é de ordem técnica, mas sim política. A mitigação dos vieses, portanto, depende de um pano de fundo regulatório que envolve, necessariamente, o conjunto de valores e interesses públicos vigentes. Tratá-la como uma questão técnica é chegar tarde demais.” (BARTH, 2021, p. 39-68)

¹⁶ “Trata-se de uma forma de governo, entendido aqui como um modo de lidar com as incertezas associadas à conduta dos governados, produzindo assim certas regularidades em torno do que se considera adequado ou desejável. Tal regime está presente em várias plataformas informacionais, tanto as puramente virtuais (sistemas de busca) quanto as híbridas (aplicativos de transporte) e tende a transbordar rapidamente para outras dimensões da vida, tais como a política, o direito e a saúde.” (BARTH, 2021, p. 39-68)

fundadas em dados contaminados (FREITAS; FREITAS, 2020. p. 93 - 94). Há casos em que, ao invés de o algoritmo aprimorar o processo decisório, ele acaba reforçando o status quo e, até mesmo, aprofundando a desigualdade ao sugerir que grupos vulneráveis recebam tratamento desfavorável.

Devido ao fato de que “[...] preconceitos sociais estariam sendo reificados nas máquinas – causando danos e discriminação de minorias [...]” (WOJCIECHOWSKI; ROSA, 2021. p. 100), considera-se a questão dos vieses algorítmicos de suma importância no desenvolvimento, manutenção e calibragem de inteligências artificiais.

Os vieses algorítmicos podem ir além de desvios sistemáticos de modelos computacionais. Eles consistem também no fenômeno em que a compreensão do que é ou não adequado passa a afetar a mente, exercendo influência nos *frames*¹⁷ humanos. Tanto é que, Cathy O’Neil (2020, p. 323 – 324), tomando como exemplo os modelos de reincidência criminal que juízes de muitos estados norte-americanos consultam antes de sentenciar, questiona se terão os padrões de sentenças de juízes mudado desde que passaram a receber análises de risco de “armas de destruição matemática”. Para ela, em alguns casos, “[...] provavelmente, o software pode moderar o julgamento deles. Em outro, não”.

Há uma tendência de as pessoas pensarem que os resultados gerados por algoritmos são corretos, menos questionáveis, tendo caráter de cientificidade. Esse erro de avaliação dos seres humanos é chamado de *machine bias*¹⁸, viés de máquina. O uso em larga escala de bases de dados viciadas, potencializa a presunção de legitimidade do processo heurístico de escolha da melhor decisão (VALENTINI, 2020, p. 546).

Os vieses sociais e algorítmicos podem ser reciprocamente fortalecidos. A retroalimentação (*feedback loop*) cria um círculo vicioso de injustiças, aprofundando desigualdades pré-existentes (SIMÕES-GOMES; ROBERTO; MENDONÇA, 2020, p.139-166). Os algoritmos propendem para os seus próprios vieses, assumindo incorretamente sua correção, e se afastam cada vez mais do ideal de justiça. Por isso, os algoritmos de aprendizado

¹⁷ Modo como as pessoas compreendem a estrutura de suas atividades. Os *frames* consistem em enquadramentos que caracterizam estruturas das atividades humanas, o modo como as pessoas a compreendem. Eles “[...] são uma espécie de domínio aberto, isto é, um domínio cujas bordas são borradas, mas não a ponto de se dissolverem: há sim regras que guiam, ainda que de modo vago e geral, nossa compreensão do que caracteriza um comportamento adequado ou inadequado[...]” em determinado local, por exemplo. (BARTH, 2021, p. 39-68)

¹⁸ “[...] uma das espécies dos vieses cognitivos humanos que ocorre pela propensão de favorecer sugestões de sistemas automatizados de tomada de decisão. Isto ocorre quando o humano sobrevaloriza a resposta da máquina e passa a não refletir acerca da correção de seus resultados. Tal viés conduz as pessoas a não reconhecerem quando os sistemas automatizados erram e a seguirem seus resultados quando apresentadas informações contraditórias.” (NUNES; LUD; PEDRON, 2022. p. 167)

de máquina precisam ser constantemente retreinados e informados das falhas em suas determinações iniciais.¹⁹

Apesar de se reconhecer a existência de vieses algorítmicos, tem-se em mente que “os sistemas devem promover a justiça social e devem ser desenvolvidos de modo a repudiar todos os tipos de discriminação, seja ela econômica, étnico, racial, de gênero, de origem nacional, orientação sexual, cultural e política” (MORAIS, et al., 2022. p. 76).

Acredita-se que, a partir de uma análise exploratória de dados, é possível identificar vieses e ajudar a corrigi-los. Para resolver tal questão é necessário uma “[...] estratégia coletiva complexa que depende de vontade política, de mudanças de aspectos culturais e muitos outros fatores sociais” (JUNQUILHO, 2022. p. 162).

A automação dos processos decisórios (ainda) não tem possibilitado realizar juízos mais imparciais. Entretanto, vislumbra-se que o uso de ferramentas de inteligência artificial tem potencial para “[...] mitigar os riscos inerentes aos vieses decorrentes da heurística de disponibilidade” (TAVARES, 2021, p. 105-132). Tem-se sugerido que, a adequada utilização dessas ferramentas pode contribuir para evitar vieses cognitivos.²⁰ Cabe ao programador fazer com que informações irrelevantes (e enviesantes) não influenciem na sugestão apresentada pela máquina. Exemplifica-se: a etnia de um réu, apesar de ser algo absolutamente sem importância no âmbito do processo penal, por vezes é considerada de forma (in)consciente pelos julgadores e julgadoras. Neste caso, em um sistema automatizado de dosimetria da pena, tal elemento poderia, teoricamente, ser desconsiderado. (TAVARES, 2021, p. 105-132)

Ao contrário dos seres humanos, as máquinas são capazes de abstrair ruídos que influenciam no processo decisório, como fome, vestimentas e tom de voz do interlocutor. Além disso, as máquinas não se submetem a uma “memória emotiva” (TAVARES, 2021, p. 105-132). Rômulo Soares Valentini (2020, p. 542) vislumbra que tais ferramentas possam proporcionar melhora qualitativa da prestação jurisdicional pois possibilitam reduzir ou controlar vieses de julgamentos humanos.

Mesmo que se conjecture a possibilidade de se escapar de vieses e preconceitos humanos por meio do uso de ferramentas tecnológicas, se reconhece que isso não afastaria a necessidade de constante supervisão e ampla abordagem da temática. Enquanto os vieses

¹⁹ “[...] *the need to affirmatively and continuously re-train algorithms against these biases becomes even more urgent over time.*” (ISRANI, 2017)

²⁰ Posiciona-se no sentido de que, “Infelizmente, até o momento, o objetivo de utilizar ferramentas tecnológicas de modo a escapar de vieses e preconceitos humanos não se concretizou” (WOJCIECHOWSKI; ROSA, 2021. p. 110).

algorítmicos ainda representam um problema significativo dos sistemas artificiais de apoio à decisão, mais do que nunca, é preciso elaborar estratégias para evitá-los ou, ao menos, mitigá-los.

2 Mitigação de vieses algorítmicos

O automatismo não pode colocar em risco a liberdade, a equidade e a dignidade. Por isso, o desviesamento²¹ representa uma decisão mandatória. Além do desviesamento da inteligência artificial, a própria IA pode auxiliar o desviesamento humano.

Para que este duplo movimento mostre-se viável, o direito à explicabilidade deve ser garantido. Ao invés de vedar a margem discriminatória do sistema automatizado, opta-se por fixar o grau e o calibre da autonomia com clareza e precisão, detectando e neutralizando predisposições ilícitas e automatismos inoportunos. Dessa forma, enfrentando os vieses que a detém, a IA poderá atuar na defesa do humano. (FREITAS; FREITAS, 2020. p. 93 - 100).

A supervisão humana é uma importante ferramenta para mitigação de vieses algorítmicos. A “precaução com vulneráveis” consiste em um princípio a ser observado na esfera da inteligência artificial e do Direito. Os sistemas de inteligência artificial devem ser dotados de mecanismos com o intuito de prevenir e evitar efeitos injustos e efeitos negativos em relação às pessoas vulneráveis (PEIXOTO, 2020. p. 156).

A transparência algorítmica, consistente na possibilidade de investigar a sequência de comandos dos algoritmos, também pode contribuir para evitar escolhas erradas, injustas ou discriminatórias. Algoritmos transparentes tornam os processos decisórios passíveis de *accountability*, tornando-os escrutináveis para prevenir, detectar ou eliminar possíveis vieses (MAGGIOLINO, 2019). Ressalta-se que a transparência precisa ser útil e viável pois apenas fornecer informações incompreensíveis pelas pessoas não tornará o processo, efetivamente, transparente (BUITEN, 2019, p. 41-59).

Entende-se a transparência como condição *sine qua non* e imperativo ético, sobre o qual deve ser construído qualquer tipo de processo decisório baseado em algoritmos. Trata-se de um princípio fundamental no gerenciamento ético de dados. O papel desempenhado pela transparência algorítmica proporciona a realização de auditoria e de supervisão dos dados. Dessa forma, é possível atenuar o risco de implementação de processos algorítmicos discriminatórios (BUITEN, 2019, p. 41-59).

²¹ “[...] reprogramação intencional das predisposições discriminatórias e injustas.” (FREITAS; FREITAS, 2020, p. 93)

Visando transparência algorítmica, de modo efetivamente explicável e compreensível, mecanismos para auditar algoritmos utilizados pelo Poder Público precisam ser criados (CANTALI; ENGELMANN, 2021). A auditoria algorítmica comporta instrumentos de governança, com função de evitar vieses. Ela infere a ideia de que um terceiro deve analisar o método utilizado pelo algoritmo para alcançar alguma decisão (MENDES; MATTIUZZO, 2019).

Outro desafio posto em diversos países, tanto na esfera pública quanto na privada, é promover a criação de um panorama regulatório para tratar dos algoritmos decisórios sem prejudicar a inovação, podendo aproveitar os seus respectivos benefícios. A forma de implementar a regulação que é um desafio. Tem predominado a ideia de uma regulação *soft*, experimentalista, com planos de ação, metas e ferramentas periódicas de reavaliação dos resultados atingidos.

Atualmente, não existe uma legislação internacional projetada, especificamente, para regular o uso de IA. O que se observa é uma “regulação” transversal, manifesta por inúmeras iniciativas esparsas. A abordagem que se destaca é a de ordem principiológica. Busca-se o estabelecimento de um rol de princípios mínimos ou valores éticos e morais que devem ser observados pelos operadores envolvidos no ciclo de vida da inteligência artificial, balizando a sua produção e a disponibilização.

Conforme argumentado por Tainá Aguiar Junquilha (2022. p. 178), a regulação ética em IA deve seguir uma abordagem humanocêntrica. Ela defende a existência de um *design* ético da IA com caráter de direito fundamental embasado no acrônimo TREPES (transparência, responsabilidade, equidade, prestação de contas, explicabilidade e sustentabilidade). Além desses princípios éticos serem embasados em normas é preciso verificar como se dará a sua implementação e controle. Nesse sentido, pontua-se que as normativas existentes não atribuem responsabilidades nem são claras sobre a aplicação e execução prática dos princípios. Por isso, no desenvolvimento de aplicações em IA, cabe ao Direito criar regulações efetivas, capazes de atribuir responsabilidades sem, no entanto, impedir a inovação.

Um outro possível expediente aos problemas relacionados aos vieses algorítmicos está na diversidade das equipes responsáveis pelo desenvolvimento de sistemas de decisões algorítmicas.

2.1 Diversidade em equipes responsáveis pelo desenvolvimento de sistemas de decisões algorítmicas

Vislumbra-se que uma das formas de evitar que preconceitos humanos influenciem no desenvolvimento e atuação de algoritmos pode ser encontrada na esfera social. O maior envolvimento comunitário no desenvolvimento das ferramentas, assim como, a formação de equipes diversificadas e multidisciplinares, mostram-se como estratégias para mitigação vieses.

A submissão do algoritmo à diversidade no treinamento é crucial para reconhecê-la no momento da aplicação (PEIXOTO; SILVA, 2019. p. 103). Ademais, os programadores devem possuir uma formação que leve em conta a complexidade social e as suas responsabilidades de atuação. Ou seja, um ensino também de ordem social. Inclusive, o Princípio 2 de Asilomar prevê a necessidade de investimento em pesquisas em Ciência da Computação, Economia, Direito, Ética e Estudos Sociais. Já o princípio 3 estabelece a premência de haver um intercâmbio construtivo e saudável entre pesquisadores de IA e formuladores de políticas (FUTURE OF LIFE INSTITUTE, 2017).

O desenvolvimento de algoritmos mais justos também depende que os times sejam diversificados, formados por grupos diversos de engenheiros e engenheiras. Isso diminui as chances de que vieses presentes na sociedade sejam transportados para os algoritmos (DUARTE; NEGÓCIO, 2021, p. 219 - 244).

Como os algoritmos são capazes de herdar os preconceitos sociais das pessoas que os programaram, mostra-se de suma importância a diversidade de programadores e programadoras, assim como, constante supervisão. Desenvolver algoritmos que não sejam influenciados por vieses humanos é um grande desafio para a humanidade. Acredita-se ser possível que os sistemas de inteligência artificial sejam melhorados para lidar com vieses (CASTRO; BOMFIM, 2020, p. 31- 45).

A contratação de grupos minoritários socialmente (mulheres, negros, indígenas, homossexuais, transgêneros etc.) pelas empresas de tecnologia para atuarem com programação consiste em uma estratégia para mitigação de vieses. Esta pode ser uma forma de contribuir para atenuação dos impactos da *big data*, que reflete injustiças sociais, nas decisões automatizadas. A inclusão é uma responsabilidade das empresas cujo ambiente “[...] também deve ser plural e diversificado, representando um microcosmo da própria sociedade” (COSTA, 2020, p. 155).

O Comitê de Inteligência Artificial do Reino Unido, em relatório publicado em 16 de abril de 2107, reforça a importância da diversidade de gênero, étnica e socioeconômica dos

talentos, no sentido de “democratizar a IA”. O domínio das posições por grupos já privilegiados na sociedade, além de agravar desigualdades já existentes, impacta significativamente na forma como os sistemas de IA são projetados e desenvolvidos.²² Por isso, o Governo do Reino Unido, tendo em vista a boa remuneração das carreiras nas áreas de tecnologia, tem incentivado a mobilidade social, garantindo que pessoas de famílias de baixa renda e desfavorecidas participem do desenvolvimento de inteligências artificiais. Há recomendação que os doutorados financiados publicamente sejam disponibilizados para uma população diversificada (minorias femininas e étnicas), mais representativa da sociedade em geral (UK PARLIAMENT, 2017).

Visando evitar vícios de origens dos sistemas de IA ou que eles incorram em decisões ética ou legalmente condenáveis, também é necessário que equipes interdisciplinares, baseadas em amplo espectro de diversidade, atuem na construção e revisão de parâmetros (GUTIERREZ, 2019, p. 90).

Esta é justamente a previsão do artigo 20, caput, da Resolução n. 332 do Conselho Nacional de Justiça (2020) que priorizou a diversidade²³, em seu mais amplo espectro, na composição de equipes para pesquisa, desenvolvimento e implementação das soluções computacionais que se utilizem de IA. Outrossim, a Resolução explicita no §4º do artigo 20 a necessidade do caráter interdisciplinar das equipes que, além de profissionais de Tecnologia de Informação, devem contar com a participação de agentes de outras áreas do conhecimento.

Para isso, é preciso uma abordagem transdisciplinar, que tenha em vista fatores sociais, técnicos e jurídico. Exige-se a colaboração de juristas, sociólogos, filósofos, psicólogos, cientistas (políticos e da computação), programadores, desenvolvedores, matemáticos e estatísticos (WOJCIECHOWSKI; ROSA, 2021, p. 110). Órgãos da sociedade civil organizada devem atuar na auditoria dos sistemas, investigando a existência de vieses e pressionar em prol da explicabilidade de decisões tomadas por algoritmos (FERRARI, 2020, p. 103 - 104).

Isabela Ferrari (2020, p. 102 - 103) também pontua a necessidade de inclusão de múltiplos *stakeholders* no desenvolvimento da forma de operação do algoritmo. Visões distintas dos programadores podem acarretar sugestões de aprimoramento do *software*. A própria regulação, não pode ser desenvolvida de forma isolada por juristas ou por especialistas

²² “[...] *If we are to ensure that these systems, which are exerting growing influence over our lives and societies, serve us fairly rather than perpetuate and exacerbate prejudice and inequality, it is important to ensure that all groups in society are participating in their development.* [...]” (UK PARLIAMENT, 2017)

²³ “[...] incluindo gênero, raça, etnia, cor, orientação sexual, pessoas com deficiência, geração e demais características individuais.” (CONSELHO NACIONAL DE JUSTIÇA, 2020)

em políticas públicas. “Problemas dessa dimensão só conseguem ser resolvidos quando se junta muita gente com *backgrounds* diferentes para atacar de forma holística a questão.”

De acordo com o Princípio da qualidade e da segurança, previsto na Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu ambiente, os modelos tecnológicos também devem ser concebidos de modo multidisciplinar.²⁴

A democratização da Inteligência Artificial mostra-se uma das estratégias a serem observadas para mitigação de vieses algorítmicos. Ela manifesta-se de forma multifacetada. Programadores e programadoras precisam ter também uma formação de ordem social e grupos minoritários socialmente devem participar de forma ativa das equipes de caráter multidisciplinar, de modo a garantir a representatividade da pluralidade social.

Considerações finais

A possibilidade de a inteligência artificial corrigir discriminações e outras questões sociais profundas ainda é uma utopia. O que ela tem feito atualmente, especialmente quando os *inputs* são enviesados ou refletem preconceitos da sociedade, é expor o atributo eminentemente humano de opressão a grupos minoritários.

A mitigação de vieses algorítmicos é uma tarefa contínua que requer supervisão humana e colaboração entre especialistas em dados, engenheiros de *machine learning* e defensores da justiça social. Dessa forma, observando-se o princípio da precaução com vulneráveis, vislumbra-se que as tecnologias de inteligência artificial sejam usadas de maneira ética e justa.

No que tange à necessidade de tornar o processo transparente, as informações fornecidas precisam ser efetivamente explicáveis e compreensíveis pelas pessoas, proporcionando também a realização de auditoria e de supervisão dos dados.

Para que preconceitos e discriminações não se ampliem na sociedade com a utilização de ferramentas de IA é preciso que os desenvolvedores de sistemas algorítmicos respeitem padrões éticos, de modo que direitos humanos e direitos fundamentais sejam garantidos. Para isso, é necessário a formação de um panorama regulatório com viés humanocêntrico que vá além de diretrizes principiológicas, atribuindo também responsabilidades.

²⁴ "Designers of machine learning models should be able to draw widely on the expertise of the relevant justice system professionals (judges, prosecutors, lawyers, etc.) and researchers/lecturers in the fields of law and social sciences (for example, economists, sociologists and philosophers). Forming mixed project teams in short design cycles to produce functional models is one of the organisational methods making it possible to capitalise on this multidisciplinary approach." (FRANÇA, 2018, p. 9)

A diversidade em equipes de desenvolvimento de IA é importante porque ajuda a garantir que os modelos sejam mais justos, precisos e úteis para todas as pessoas, independentemente de seu gênero, raça, etnia, orientação sexual ou outras características demográficas. Dessa forma, tendo como pano de fundo os valores e interesses da sociedade, acredita-se ser possível alcançar um juízo algorítmico não enviesado capaz de promover a justiça social.

Referências bibliográficas

ALVES, Marco Antônio Sousa; ANDRADE, Otávio Morato de. Da “caixa-preta” à “caixa de vidro: o uso da *Explainable Artificial Intelligence* (XAI) para reduzir a opacidade e enfrentar o enviesamento em modelos algorítmicos. **Revista de Direito Público**. Vol. 18. n. 100, out/dez. 2021, p. 348 – 373. Disponível em: <https://doi.org/10.11117/rdp.v18i100.5973>. Acesso em: 16 mar. 2022.

ANGWIN, Julia et al. **Machine Bias**: There’s software used across the country to predict future criminals. And it’s biased against blacks. 23 maio 2016. Disponível em: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Acesso em: 03 out. 2022.

BAROCAS, Solon; SELBST, Andrew D. Big data’s disparate impact. **California Law Review**, v. 104, p. 671-732, 2016. Disponível em: <https://www.jstor.org/stable/24758720>. Acesso em: 21 jul. 2022.

BARTH, Carlos Henrique. É possível evitar vieses algorítmicos? **Revista de Filosofia Moderna e Contemporânea**, Brasília, v. 8, n. 3, p. 39-68, 31 jan. 2021. <http://dx.doi.org/10.26512/rfmc.v8i3.34363>. Disponível em: <https://periodicos.unb.br/index.php/fmc/article/view/34363>. Acesso em: 18 jul. 2022.

BUITEN, Miriam C.. Towards Intelligent Regulation of Artificial Intelligence. **European Journal of Risk Regulation**, vol. 10, 2019, p. 41-59. doi:10.1017/err.2019.8. Disponível em: [Towards Intelligent Regulation of Artificial Intelligence | European Journal of Risk Regulation | Cambridge Core](https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/towards-intelligent-regulation-of-artificial-intelligence/10.1017/err.2019.8). Acesso em 02 out. 2022.

CABALLOL, Daniel Contreras; DENDAL, Daniel Pefaur. Transparência Algorítmica: buenas prácticas y estándares de transparência em el proceso de toma de decisiones automatizadas. **Cuadernos de Trabajo**. Santiago. Out. 2020. Disponível em: [Transparencia-Algoritmica.pdf \(consejotransparencia.cl\)](https://www.consejotransparencia.cl/Transparencia-Algoritmica.pdf). Acesso em: 11 out. 2022.

CANTALI, Fernanda Borghetti; ENGELMANN, Wilson. Do não cognitivismo dos homens ao não cognitivismo das máquinas: percursos para o uso de decisões judiciais automatizadas. **Revista Jurídica Portucalense**. n. 29, 2021.

CASTRO, Bruno Fediuk de; BOMFIM, Gilberto. A inteligência artificial, o direito e os vieses. **Revista Ilustração**, v. 1, n. 3, p. 31-45, 2020. DOI:

<https://doi.org/10.46550/ilustracao.v1i3.24>. Disponível em: [23-Texto do Artigo-71-1-10-20210105.pdf](#). Acesso em 10 out. 2022.

CONSELHO NACIONAL DE JUSTIÇA. Resolução n. 332, de 21 de agosto de 2020. Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências. Brasília, Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3429>. Acesso em: 05 set. 2020.

COSTA, Diego Carneiro. **O viés do algoritmo e a discriminação por motivos relacionados à sexualidade**. Dissertação (Mestrado) – Universidade Federal da Bahia, Faculdade de Direito, Salvador, 2020. Disponível em: [Universidade Federal da Bahia: O viés do algoritmo e a discriminação por motivos relacionados à sexualidade \(ufba.br\)](#). Acesso em: 04 out. 2022.

DANKS, David; LONDON, Alex John. Algorithmic Bias in Autonomous Systems. In: **26th International Joint Conference on Artificial Intelligence (IJCAI)**. 2017. p. 4691-4697. Disponível em: [Microsoft Word - IJCAI17-AlgorithmicBias-Distrib.doc \(cmu.edu\)](#). Acesso em: 17 jul. 2022.

DUARTE, Alan; NEGÓCIO, Ramon de Vasconcelos. Todos são iguais perante o algoritmo? Uma resposta cultural do Direito à discriminação algorítmica. **Revista Direito Público**, ano 18, n. 100, out – dez 2021, p. 219 – 244. Doi: <https://doi.org/10.11117/rdp.v18i100.5869>. Disponível em: <https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/5869>. Acesso em: 10 out. 2022.

FERRARI, Isabela. O emprego de algoritmos para a tomada de decisão II: riscos dos algoritmos decisórios. In FERRARI, Isabela (org.). **Justiça Digital**. São Paulo: Thomson Reuters Brasil, 2020. Cap. 7.

FERRARI, Isabela. O emprego de algoritmos para a tomada de decisões III: regulação de algoritmos. In FERRARI, Isabela (org.). **Justiça Digital**. São Paulo: Thomson Reuters Brasil, 2020. Cap. 8.

FLORES, Nilton Cesar da Silva; SANTOS, Raphael de Souza Almeida. Direito e inteligência artificial: metamorfose, vieses algorítmicos e decisionismo tecnológico no Brasil. **Revista Direito e Justiça: reflexões sociojurídicas**. Santo Ângelo. v. 21. maio/ago. 2021. p. 99 – 113.

FRANÇA. European commission for the efficiency of justice. Council of Europe. **European ethical charter on the use of artificial intelligence in judicial systems and their environment**. Estrasburgo. 2018. p. 9. Disponível em: <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>. Acesso em: 05 set. 2020.

FREITAS, Juarez; FREITAS, Thomas Bellini. **Direito e Inteligências Artificiais**: em defesa do humano. Belo Horizonte: Fórum, 2020.

FUTURE OF LIFE INSTITUTE. **Asilomar AI Principles**. 2017. Disponível em: [AI Principles - Future of Life Institute](#). Acesso em: 28 out. 2022.

GUTIERREZ, Andriei. É possível confiar em um Sistema de Inteligência Artificial? Práticas em torno da melhoria da sua confiança, segurança e evidências de *accountability*. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (coord.) **Inteligência Artificial e direito: ética, regulação e responsabilidade**. São Paulo: Thomson Reuters Brasil, 2019. Cap. 4.

ISRANI, Ellora. Algorithmic Due Process: mistaken Accountability and Attribution in State v. Loomis. **JOLT Digest** - Harvard Journal of Law and Technology. Ago. 2017

JUNQUILHO, Tainá Aguiar. **Inteligência artificial no direito: limites éticos**. São Paulo: JusPodivm, 2022.

MAGGIOLINO, Mariateresa. EU Trade Secrets Law and Algorithmic Transparency. **Bocconi Legal Studies Research Paper**. n. 3363178. 31 de março de 2019. DOI <http://dx.doi.org/10.2139/ssrn.3363178>. Disponível em: <https://ssrn.com/abstract=3363178>. Acesso em 2 out. 2022.

MAYSON, Sandra G. Bias in, bias out. **The Yale Law Journals**, v. 128, p. 2218 - 2300, jun. 2019. Disponível em: [Yale Law Journal - Bias In, Bias Out](http://www.yalelawjournal.org/article/yale-law-journal-bias-in-bias-out). Acesso em: 22 jul. 2022.

MENDES, Laura Schertel; MATTIUZZO, Marcela. Discriminação algorítmica: conceito, fundamento legal e tipologia. **Direito Público**, v. 16, n. 90, 2019. Disponível em: [DISCRIMINAÇÃO ALGORÍTMICA: CONCEITO, FUNDAMENTO LEGAL E TIPOLOGIA | Direito Público \(idp.edu.br\)](http://www.idp.edu.br/revista/16-90-1). Acesso em: 15 out. 2022.

MORAIS, Ana Luisa Zago de et al. **Inteligência artificial e direitos humanos: aportes para um marco regulatório no Brasil**. São Paulo: Dialética, 2022.

NUNES, Dierle; LUD Natanael; PEDRON, Flávio Quinaud. **Desconfiando da (im)parcialidade dos sujeitos processuais: um estudo sobre vieses cognitivos, o ruído, a mitigação de seus efeitos e os debiasing**. 3. ed. São Paulo: JusPodium, 2022.

O'NEIL, Cathy. **Algoritmos de destruição em massa: como o big data aumenta a desigualdade e ameaça a democracia**. Santo André: Editora Rua do Sabão, 2020. Tradução: Rafael Abraham Título original: *Weapons of math destruction: how big data increases inequality and threatens democracy*.

PEIXOTO, Fabiano Hartmann. **Inteligência artificial e direito: convergências éticas e estratégicas**. Vol. 5. Curitiba: Alteridade, 2020.

PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. **Inteligência artificial e direito**. Vol. 1. Curitiba: Alteridade, 2019.

SILVA, Pollyanna Maria da. BERNARDES NETO, Napoleão. O uso de softwares de inteligência artificial nas decisões da justiça criminal: parâmetros para criação e utilização. In: PINTO, Danielle Jacson Ayres; BASTOS, Elísio Augusto Velloso; ROVER, Aires José. (Orgs.). **Direito, governança e novas tecnologias [Recurso eletrônico on-line]**. Florianópolis: CONPEDI, 2019. p. 163 - 181. Disponível em: site.conpedi.org.br/publicacoes/048p2018/z54tvw1x. Acesso em: 03 out. 2022

SIMÕES-GOMES, Letícia; ROBERTO, Enrico; MENDONÇA, Jônatas. Viés algorítmico: um balanço provisório. **Estudos de Sociologia**, Araruama, v. 25, n. 48, p.139-166, jan.- jun. 2020.

TAVARES, João Paulo Lordelo Guimarães. Vieses implícitos e técnicas de automação decisória: riscos e benefícios. **Civil Procedure Review**, v. 12, n. 1, p. 105-132, 2021. Disponível em: <https://www.civilprocedurereview.com/revista/article/view/225>. Acesso em 17 mar. 2022.

UK PARLIAMENT. **Artificial Intelligence Committee**. AI in the UK: ready, willing and able? 2017. Disponível em: [Câmara dos Lordes - IA no Reino Unido: pronta, disposta e capaz? - Comitê de Inteligência Artificial \(parliament.uk\)](#). Acesso em: 11 out. 2022.

VALENTINI, Rômulo Soares. Para além do teste de Turing Jurídico? Breves apontamentos sobre sistemas automatizados de decisão e suas potencialidades para elevar a qualidade da prestação jurisdicional. In: NUNES, Dierle; LUCON, Paulo Henrique dos Santos; WOLKART, Erik Navarro (org.). **Inteligência artificial e o direito processual**: os impactos da virada tecnológica no direito processual. Salvador: Juspodium, 2020. Cap. 20.

WOJCIECHOWSKI, Paola Bianchi; ROSA, Alexandre Morais da. **Vieses da justiça**: como as heurísticas e vieses operam nas decisões penais e a atuação contraintuitiva. 2. ed. Florianópolis: Emais, 2021.