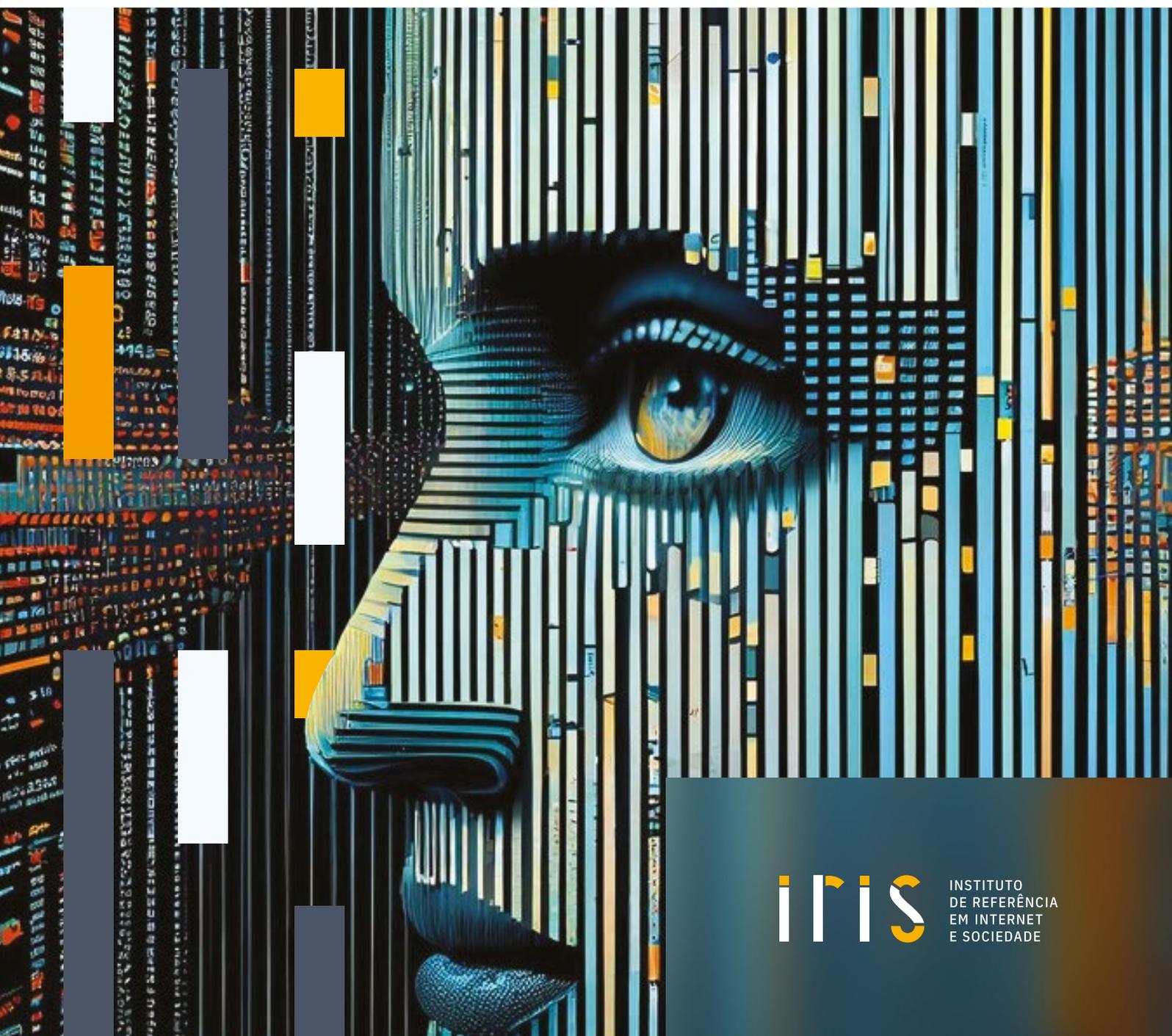


Anais

Inteligência Artificial e Segurança Internacional

desafios globais, desenvolvimento,
ética e regulação



iris

INSTITUTO
DE REFERÊNCIA
EM INTERNET
E SOCIEDADE

Anais

Inteligência Artificial e Segurança Internacional

desafios globais, desenvolvimento,
ética e regulação

ORGANIZAÇÃO

Fernanda dos Santos Rodrigues Silva
Wilson Guilherme Dias Pereira

REVISÃO

Ana Bárbara Gomes Pereira

PROJETO GRÁFICO, CAPA E DIREÇÃO DE ARTE

Felipe Duarte
Imagem de capa por Vecteezy no Freepik

DIAGRAMAÇÃO

Míriam Vitória

PRODUÇÃO EDITORIAL

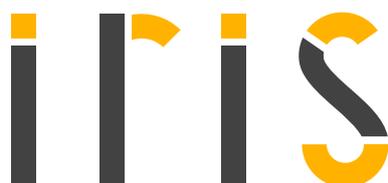
IRIS - Instituto de Referência em Internet e Sociedade

COMO REFERENCIAR EM ABNT

SILVA, Fernanda dos Santos Rodrigues; PEREIRA, Wilson Guilherme Dias (orgs.). **Anais Inteligência Artificial e Segurança Internacional**: desafios globais, desenvolvimento, ética e regulação. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2023. Disponível em: <<https://bit.ly/40M5ti3>>. Acesso em: dd mmm aaaa.

ISBN

978-65-87416-04-5



INSTITUTO
DE REFERÊNCIA
EM INTERNET
E SOCIEDADE

DIREÇÃO

Ana Bárbara Gomes

Paloma Rocillo

MEMBROS

Bruni Emanuele | Analista Administrativo

Felipe Duarte | Coordenador de Comunicação

Fernanda Rodrigues | Coordenadora de Pesquisa e Pesquisadora

Glenda Dantas | Pesquisadora

Júlia Caldeira | Pesquisadora

Luiza Correa de Magalhães Dutra | Pesquisadora

Paulo Rená da Silva Santarém | Pesquisador

Rafaela Ferreira | Estagiária de pesquisa

Thais Moreira | Analista de comunicação

Victor Barbieri Rodrigues Vieira | Pesquisador

Wilson Guilherme | Pesquisadore

irisbh.com.br

SUMÁRIO

APRESENTAÇÃO	05
PALAVRAS INICIAIS	05
GRUPO DE TRABALHO: IA, DESENVOLVIMENTO E DESAFIOS GLOBAIS	07
Algoritmos decisoriais de retropropagação no poder judiciário brasileiro: limites considerando o estado d'arte	08
Rafa 2030 - Classificação de textos jurídicos em objetivos de desenvolvimento sustentável da agenda 2030 das nações unidas	21
As implicações sob a perspectiva da justiça de dados acerca do uso de tecnologias de reconhecimento facial para gratuidade de transporte público	34
Externalidades da inteligência artificial na transformação digital das cidades	51
GRUPO DE TRABALHO: SEGURANÇA INTERNACIONAL E IA	67
Is there a state's right to security? Artificial intelligence as a tool of deepening global security and defense inequalities	68
Inteligência artificial no combate ao tráfico internacional de crianças e adolescentes: para aqueles que não tiveram a chance de buscar ajuda, de serem vistos e de serem protegidos	90
Parâmetros regulatórios do uso da inteligência artificial na segurança pública: uma análise entre brasil e união europeia	110
GRUPO DE TRABALHO: ÉTICA E REGULAÇÃO DA IA	134
Da explicabilidade à interpretabilidade: devido processo legal e modelos de inteligência artificial para tomada de decisão judicial	135
Avaliação de impacto algorítmico na regulação de inteligência artificial: elementos estruturais sobre esse instrumento de accountability	147
Algorithmic impact assessment as a tool to strengthen ethics by design in ai	164
Câmeras de reconhecimento facial na segurança pública: uma análise sobre as novas formas de controle e vigilância	178

Apresentação

O Instituto de Referência em Internet e Sociedade (IRIS) é um centro de pesquisa independente e interdisciplinar fundado em 2015 é dedicado a produzir e comunicar conhecimento científico sobre os temas de internet e sociedade, bem como a defender e fomentar políticas públicas que avancem os direitos humanos na área digital. Nossa atuação busca qualificar e democratizar os debates sobre internet, sociedade e novas tecnologias ao trazer insumos científicos aos usuários da internet e aos diferentes setores que compõem a sociedade: governo, sociedade civil, setor privado, comunidade técnica e acadêmica.

Desenvolvemos projetos de pesquisa que envolvem a publicação de livros, estudos, artigos científicos, notas técnicas e relatórios no campo da governança da internet sobre temas como privacidade e proteção de dados, criptografia e segurança pública, regulação de plataformas, inclusão digital, regulação de criptoativos, inteligência artificial e muitos outros. Posicionando a comunicação como aspecto central à atividade científica, também atuamos na criação de cursos e conteúdos em diferentes formatos nas redes sociais, assim como no oferecimento de apoio consultivo à imprensa sobre nossos temas de estudo e na promoção de campanhas envolvendo direitos humanos na área digital. Acreditamos que a incidência por uma internet democrática, aberta e livre é amparada pelo conhecimento cientificamente embasado, por relatórios, estudos nacionais e internacionais e pelo diálogo com os pares.

Palavras iniciais

O ano de 2023 foi fortemente marcado no Brasil e no mundo pela corrida em torno do desenvolvimento e utilização de tecnologias de inteligência artificial - IA, especialmente diante da inserção de IAs generativas na realidade de muitas pessoas usuárias da internet. De redes sociais a trabalhos, e até mesmo na segurança pública, a utilização de IA tem se tornado cada vez mais comum. Entretanto, quando se ampliam as oportunidades de utilização, aumenta-se também a exposição a riscos, o que demanda a construção de critérios éticos para o seu desenvolvimento.

O 18º Encontro do *Internet Governance Forum* - IGF foi marcado pelo diálogo em torno da necessidade de ampliar globalmente tanto o desenvolvimento, como a responsabilidade na utilização de tecnologias de IA. De fato, essas tecnologias podem auxiliar na automação e produtividade de determinadas áreas, mas em campos como segurança pública, saúde, justiça, dentre outros, elas podem apresentar sérios riscos à sociedade e, especialmente, a grupos historicamente vulnerabilizados.

Assim, o Instituto de Referência em Internet e Sociedade - IRIS, comprometido com o desenvolvimento científico nacional, lançou em fevereiro de 2023 a chamada de

trabalhos científicos “Inteligência Artificial e Segurança Internacional: desafios globais, desenvolvimento, ética e regulação”. A apresentação dos trabalhos selecionados ocorreu entre os dias 09 a 11 de maio de 2023 e contou com 11 artigos científicos, escritos em português ou inglês, divididos em três grupos de trabalhos, quais sejam:

- a. **IA, Desenvolvimento e Desafios Globais:** com pesquisas que analisassem, como os sistemas de IA podem ser desenvolvidos e implantados contribuindo para uma mudança positiva e uma abordagem baseada em direitos humanos para o desenvolvimento social e econômico no contexto dos desafios internacionais emergentes.
- b. **Segurança Internacional e IA:** com enfoque no desenvolvimento de pesquisas sobre o papel cada vez mais importante assumido pelas aplicações de IA em contextos de segurança nacional e internacional, como segurança cibernética, biossegurança, armas autônomas, combate ao financiamento de ilícitos, entre outros temas com implicações internacionais.
- c. **Ética e regulação da IA:** com trabalhos voltados aos debates contemporâneos sobre abordagens éticas e regulatórias para lidar com os riscos impostos por sistemas de IA e promover a inovação responsável.

A coordenação dos GTs foi composta por duas pessoas, sendo uma convidada externa e uma pesquisadora do IRIS. Aproveitamos para agradecer as pessoas coordenadoras de GT pela generosa colaboração na seleção, mediação e contribuição aos trabalhos, bem como agradecer as pessoas autoras dos artigos pela confiança no trabalho desenvolvido pelo IRIS para a divulgação de sua pesquisa científica.

Grupo de Trabalho

IA, DESENVOLVIMENTO E DESAFIOS GLOBAIS

**PESSOAS
COORDENADORAS**

Paulo Rená da Silva Santarém
e Tainá Aguiar Junquilha

ALGORITMOS DECISIONAIS DE RETROPROPAGAÇÃO NO PODER JUDICIÁRIO BRASILEIRO: LIMITES CONSIDERANDO O ESTADO D'ARTE

BACKPROPAGATION DECISION ALGORITHMS IN THE BRAZILIAN JUDICIARY: LIMITS CONSIDERING THE STATE OF THE ART

Roberta Eggert Poll¹

Aline Pires de Souza Machado de Castilhos²

Resumo

O trabalho promove uma discussão acerca do uso de sistemas de inteligência artificial no Poder Judiciário brasileiro. O problema de pesquisa procura responder a seguinte indagação: uma máquina que se utiliza da técnica de retropropagação é capaz de compreender, interpretar e aplicar o Direito considerando o sistema jurídico vigente? O objetivo da pesquisa é analisar como a revolução tecnológica é capaz de contribuir para o desenvolvimento social e econômico do país, atendendo aos direitos humanos e mitigando possíveis vieses e heurísticas negativos no uso de sistemas de IA pelo Poder Judiciário. Para a construção do texto foi utilizado o método de procedimento estruturalista e, o método de interpretação, foi o sistemático aberto. Por fim, a pesquisa, quanto à sua natureza foi teórica, com a revisão da bibliografia (nacional e estrangeira) sobre o tema, visando construir e embasar o referencial teórico para uma interpretação sistêmica do Direito. O trabalho estrutura-se da seguinte forma: inicia-se o texto abordando os principais aspectos da virada tecnológica dentro do Direito para, depois, explicar o que são e como funcionam os algoritmos de retropropagação e como eles podem auxiliar juízes e Tribunais na tomada de decisões jurisdicionais. Ao final, são tecidas as conclusões da pesquisa.

1 Doutora em Direito pela PUC/RS. Mestre em Ciências Criminais pela PUC/RS. Advogada Criminalista sócia-fundadora da Poll Advocacia. Professora de Direito Penal e Criminologia. E-mail: roberta.poll@hotmail.com.

2 Doutoranda e mestre em Ciências Criminais pela PUC/RS. Professora de Direito Penal e Psicologia Jurídica na UNIFTEC. Professora convidada do CJED – Centro de Estudos do Poder Judiciário. Funcionária Pública do TJRS. E-mail: alinepirescastilhos@gmail.com.

Palavras-chave: Inteligência artificial; Algoritmo de retropropagação; Decisões jurisdicionais; Direito.

Abstract

The paper promotes a discussion about the use of artificial intelligence systems in the Brazilian Judiciary. The research problem seeks to answer the following question: is a machine that uses the technique of backpropagation able to understand, interpret and apply the Law considering the current legal system? The objective of the research is to analyze how the technological revolution is able to contribute to the social and economic development of the country, meeting human rights and mitigating possible negative biases and heuristics in the use of AI systems by the Judiciary. For the construction of the text, the structuralist procedure method was used, and the interpretation method was the open systematic one. Finally, the research, as to its nature, was theoretical, with a review of the bibliography (national and foreign) on the theme, aiming to build and base the theoretical referential for a systemic interpretation of Law. The work is structured as follows: the text begins by addressing the main aspects of the technological change in Law, and then explains what backpropagation algorithms are and how they work, and how they can help judges and courts in making jurisdictional decisions. At the end, the conclusions of the research are drawn.

Keywords: Artificial intelligence; Backpropagation algorithm; Jurisdictional decisions; Law.

1. INTRODUÇÃO

A inteligência artificial vem sendo descrita e de certa forma “atacada”, como uma perturbação no campo do Direito e das diversas profissões jurídicas. Isto porque, ao se mostrar disruptiva, a IA modificou não só as formas de pensar, mas o dia-a-dia dos profissionais do Direito, que se viram em uma encruzilhada entre alfabetização digital e adaptação as novas tecnologias ou a permanência no ostracismo e na manufatura contraproducente as metas e necessidades dos novos tempos.

A IA no campo do direito inclui abordagens como o processamento de linguagem natural, isto é, a capacidade de comunicação natural entre homens e máquinas, (mais conhecido no Teste de Turing), o conceito de *machine learning* (ou aprendizagem automática), com melhoras exponenciais em termos de previsões algorítmicas, além é claro do raciocínio, planejamento, comunicação e percepção extraídos do *big data*.

Somente a terminologia exposta nos dois primeiros parágrafos deste texto já aponta para a interdisciplinaridade entre as ciências computacionais, robótica, tecnologia da informação e o Direito. Ou seja, os operadores do Direito nos últimos 5 anos precisaram

efetivamente modificar as suas formas de trabalho, frente a esse mundo tecnológico contemporâneo.

Normalmente, um sistema de IA recebe um conjunto de treino de dados sobre o sujeito e os seus algoritmos identificam as relações entre os dados. Este “treino” pode basear-se em humanos que corrigem as respostas da máquina (aprendizagem supervisionada) ou simplesmente pelo sistema que responde ao *feedback* do seu ambiente (aprendizagem não supervisionada).

Diante de todas essas inovações, mister discutir acerca do uso de sistemas de inteligência artificial no Poder Judiciário brasileiro. O problema de pesquisa procura responder, portanto, a seguinte indagação: uma máquina que se utiliza da técnica de retropropagação é capaz de compreender, interpretar e aplicar o Direito considerando o sistema jurídico vigente?

O objetivo da pesquisa é analisar como a revolução tecnológica é capaz de contribuir para o desenvolvimento social e econômico do país, atendendo aos direitos humanos e mitigando possíveis vieses e heurísticas negativos no uso de sistemas de IA pelo Poder Judiciário.

O problema de pesquisa interliga noções de Direito e Cibernética, daí porque um dos métodos condutores da pesquisa é o dialético. Para a construção do texto, também, foi utilizado o método de procedimento estruturalista e, o método de interpretação foi o sistemático aberto. Por fim, a pesquisa, quanto à sua natureza foi teórica, com a revisão da bibliografia (nacional e estrangeira) sobre o tema, visando construir e embasar o referencial teórico para uma interpretação sistêmica do Direito.

O trabalho estrutura-se da seguinte forma: inicia-se o texto abordando os principais aspectos da virada tecnológica dentro do Direito para, depois, explicar o que são e como funcionam os algoritmos de retropropagação e como eles podem auxiliar juízes e Tribunais na tomada de decisões jurisdicionais. Ao final, são tecidas as conclusões da pesquisa.

2. COMPREENDENDO O USO DE ALGORITMOS DECISIONAIS E OS LIMITES DE SUA UTILIZAÇÃO NO PODER JUDICIÁRIO BRASILEIRO

As últimas décadas trouxeram mudanças significativas de paradigmas promovidas pelo desenvolvimento de sistemas inteligentes. A queda das fronteiras operada pelo avanço

das novas tecnológicas como, a internet das coisas, comunicações *machine-to-machine*, robótica, *machine learning* e o *big data*,³ representa apenas uma parcela do impacto da Revolução 4.0 no campo do Direito.

No caso brasileiro, o desenvolvimento, a implementação e o uso de tecnologia da informação no âmbito do Poder Judiciário é um foco importante de atuação do Conselho Nacional de Justiça,⁴ bem como das Cortes Superiores, observado-se sempre a sua compatibilidade com os Direitos Fundamentais.⁵ Não por outro motivo que, o CNJ editou a Resolução 332, que trata sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário, foi um grande marco regulatório da matéria no país.

Nesse diapasão, verifica-se um alargamento nos últimos anos da utilização da IA no Poder Judiciário, a qual vem contribuindo com agilidade e coerência no processo de tomada de decisão. A IA, deve, por via de consequência lógica observar uma série de critérios éticos de transparência, que garantam a possibilidade de auditoria, previsibilidade de sua atuação e, também, garantia de imparcialidade na toma de decisão. Ademais, tais decisões devem preservar a igualdade e não discriminação, o pluralismo, a solidariedade e o julgamento justo.

As soluções de IA no âmbito do Direito, por outro lado, surgem, para além da inafastável revolução tecnológica, diante da impossibilidade de o ser humano gerir todos os dados que influenciam a sua vida, seja por restrição de armazenamento de informações no cérebro, seja por absoluta impropriedade do meio.

Explica-se: é biologicamente impossível ensinar tudo a alguém ou exigir que um indivíduo trabalhe de forma ininterrupta; os algoritmos de IA, por outro lado, podem ser programados para isto, realizando atividades rotineiras de um ser humano, bem como aquelas que poderiam ser inatingíveis como, verificar todos os acórdãos julgados por um Tribunal, em questão de minutos, até encontrar algum que seja relacionado ao

3 CASTILHOS, Guilherme Machado de; POLL, Roberta Eggert; CASTILHOS, Aline Pires de Souza Machado de. “E se a sua geladeira pudesse depor contra você no Tribunal?”. *Internet das Coisas e provas no processo penal brasileiro*. **Revista Brasileira de Ciências Criminais**, São Paulo, v. 163, n. 28, p. 363-391, jan. 2020, p. 365.

4 Faz parte da Estratégia Nacional do Poder Judiciário, que trata da gestão e do planejamento do Judiciário brasileiro, o Macro desafio Melhoria da Infraestrutura e Governança de TIC, cujo objetivo é: [...] garantir confiabilidade, integralidade e disponibilidade das informações, dos serviços e sistemas essenciais da justiça, por meio do incremento e modernização dos mecanismos tecnológicos, controles efetivos dos processos de segurança e de riscos, assim como a otimização de recursos humanos, orçamentários e tecnológicos. Compare em: BRASIL. Conselho Nacional de Justiça. **Relatório de acompanhamento da estratégia nacional: 2016/2017**. Brasília: CNJ, 2018. Disponível em: <https://www.cnj.jus.br/wp-content/uploads/2022/11/relatorio-de-acompanhamento-da-estrategia-nacional-2021-v2-2022-05-31.pdf>. Acesso em: 08 de mar. 2023.

5 OLIVEIRA, Fabiana Luci de; CUNHA, Luciana Gross. Os indicadores sobre o Judiciários brasileiro: limitações, desafios e o uso da tecnologia. **Revista Direito FGV**, São Paulo, v. 16, n. 01, p. 1-23, jan./abr. 2020, p. 6.

objeto de pesquisa – tarefa que demandaria anos para um ser humano.⁶

Diante desse cenário complexo, o presente trabalho evidencia a maximização da capacidade cognitiva dos magistrados na tomada de decisões jurisdicionais, de modo a aperfeiçoar o processo decisório e de atuação dos intérpretes da lei, especialmente, juízes, verificando os limites de aplicação da inteligência artificial, diante da não tão distante possibilidade de termos futuramente um processo conduzido e julgado na íntegra por um robô. Para tanto, verifica-se a hipótese de utilização de algoritmos de retropropagação pelo Judiciário brasileiro.

Com o aprimoramento dos programas de computador a vida social e comercial das pessoas foi impactada, gerando uma série de novos modelos. A Inteligência artificial também já é uma realidade. Quando algoritmos adquirem a capacidade de interagir em situações e contextos, se aproximando do pensar e agir de um ser humano, a partir de uma técnica que se convencionou chamar de *machine learning*.

O aprendizado da máquina ou *machine learning* consiste na capacidade do programa de IA realizar de forma inteligente tarefas cotidianas e profissionais que outrora eram desenvolvidas por manufatura humana. Trata-se, portanto, de uma inteligência que se assemelha ao cérebro humano que, no entanto, é feita por um programa de computador. Os mecanismos de IA são tão inteligentes que o sistema possui a capacidade de perceber o ambiente a sua volta, coletar dados da realidade e tomar decisões acertadas.

O *machine learning* supervisionado ocorre quando a IA é estruturada por meio de uma técnica de aprendizagem profunda, mas que necessita, no entanto, da supervisão humana, na medida em que o sistema não é capaz de fazer inferências por si próprio. Os modelos de IA atuais são capazes de aprender, de treinar por si próprios por meio do acúmulo de experiências anteriores próprias ou de agentes externos, chegando as suas próprias conclusões, mas ainda necessitam ao fim do apoio do ser humano em uma ou outra etapa.

Por exemplo, o sistema desenvolvido para jogar o jogo Go (*AlphaGo*) foi treinado em mais de 30 milhões de jogadas, sendo capaz de fazer jogadas que nenhum especialista em Go imaginou, especialmente quando derrotou o campeão mundial por cinco jogos a zero.⁷

O termo “IA estreita” é utilizado para sistemas concebidos para atingir um objetivo específico, como jogar xadrez ou Go, e até mesmo diagnosticar uma doença. Os atuais sistemas de IA são deste tipo. Um dos primeiros sistemas de IA altamente avançados foi o Watson da IBM, construído para aplicar um sistema de processamento de linguagem natural, recuperação de informação, representação do conhecimento, raciocínio

6 MARTINEZ, Vinício Carrilho; SCHERCH, Vinícius Alves. Relações entre direito e tecnologia no século XXI. **Revista de Direito, Viçosa**, v. 12, n. 1, p. 1-23, jan./abr. 2020, p. 3-6.

7 Knight, Will (2016), ‘AI’s language problem’. MIT Technology Review. Disponível em: <https://www.technologyreview.com/s/602094/ais-language-problem>.

automático e tecnologias de aprendizagem. Tal sistema vem sendo utilizado em domínios relacionados à área de saúde, da indústria farmacêutica, na edição e na biotecnologia, como assistente de ensino e para a previsão meteorológica.⁸

Outra aplicação específica da IA é a dos “*chatbots*” – assistentes virtuais como a Siri da Apple, o Google Assistant ou a Alexa da Amazon. Existem aqueles mais sofisticados que utilizam o processamento de linguagem natural, incluindo um *chatbot* médico alimentado por IA, como acontece com o Melody, desenvolvido pelo gigante chinês das pesquisas Baidu. O bot pergunta aos pacientes os seus sintomas, que são comparados com todo o conhecimento médico anterior armazenado pelo Melody. Depois disso, os sintomas e uma série de possíveis diagnósticos são enviados ao médico, que poderá iniciar o tratamento mais adequado.⁹

Existem, ainda, modelos mais profundos de aprendizagem ou na linguagem técnica – *deep learning* – são aqueles que realizam análise de dados sem que exista uma programação prévia; atuando de forma totalmente independente de um ser humano são capazes de atingir níveis profundos inferências, tal qual uma rede neural. “Os algoritmos do sistema não supervisionado aprendem com uma vasta quantidade de dados que estão disponibilizados imediatamente na internet ou em qualquer outra fonte (*big data*)”.¹⁰

Embora a descrição pareça simples, nas redes neurais, as etapas capturam as características do cérebro humano. Os neurônios humanos são substituídos por unidades de processamento dispostas em camadas sucessivas (*layers*), que se conectam entre si, relacionando-se sistematicamente conforme o ingresso ou a saída de informações. Em particular, cada unidade computacional calcula uma função específica de suas entradas e passa o resultado para outras unidades da rede que estão conectadas a ela: os resultados finais são produzidos pelo processamento paralelo e distribuído dessa

8 WATSON, Willis Towers (2018). Horizon Scanning forward thinking. Artificial Intelligence and the legal profession. The Law Society. Disponível em: <https://www.lawsociety.org.uk/topics/research/ai-artificial-intelligence-and-the-legal-profession>. Acesso em: 16 maio 2023.

9 Gellego, Jellor (2016), ‘**An AI-Powered Chatbot Is Helping Doctors Diagnose Patients**’. Retrieved from: <http://futurism.com/an-ai-powered-chatbot-is-helping-doctors-diagnose-patients/>. Acesso em: 16 maio 2023.

10 A principal fonte de dados é a internet, onde deixamos “pegadas digitais” ao navegar através de browsers, utilizar aplicativos, realizarmos transações on-line, enviarmos *e-mails* ou mensagens em chats, visualizar vídeos, imagens, clicarmos em *likes*, interagirmos em redes sociais, etc. Mas há outras formas de deixar pegadas digitais fora da internet, que são capturáveis e transformadas em dados que alimentam o sistema (*inputs*), como imagens captadas por câmaras de videovigilância, passes-livres rodoviários (sensores que permitem o veículo avançar sem ter de parar para pagar pedágio), cartões de desconto, etc. Em um primeiro momento, o sistema explora e prospecta todos esses dados (*data mining*), que depois servirão para um modelo de análise de dados, a partir de algoritmos que conduzem a um resultado (*outcome*) para o qual foi programado. Sobre esse processo, v. LEAL, Ana Alves. *Big data* nos setores bancário e financeiro. In: LOPES ROCHA, Manuel; SOARES PEREIRA, Rui (coord.). *Inteligência Artificial & Direito*. Coimbra: Almedina, 2020, p. 199/220, p. 201.

rede de conexões neurais a seus limiares de peso.¹¹

As redes neurais, portanto, não aprendem por meio de sentenças em linguagem simbólica. O conhecimento está implícito na estrutura e na interação desses neurônios. Logo, elas aprendem modificando o seu esqueleto global, de modo a se adaptar às contingências do sistema que habitam. Os neurônios em uma rede de multicamadas estão conectados com unidades de uma camada passando as suas ativações apenas para os neurônios de camadas superiores. Esse processamento de sinais em múltiplas camadas significa que os erros no interior da rede podem ser corrigidos, fazendo com que o algoritmo evolua em formas complexas. A retropropagação fornece, nessa atividade, um algoritmo para atribuir aos neurônios a sua parcela de culpa pelo erro da rede e ajustar os pesos de forma correspondente para que não aconteçam mais.¹²

Transportando esse processo para o Direito é possível que um algoritmo de rede neural de retropropagação, baseado em um modelo de acertos e erros, identifique sistematicamente o melhor caminho e a decisão mais acertada para o caso concreto.

Inicialmente, ao argumento cético no sentido de que a atividade judicial de análise de provas, por envolver matéria de fato, não poderia ser realizada por sistemas automatizados, responderam os técnicos com a invocação do teorema de Bayes, o qual descreve a probabilidade de um evento, baseado em um conhecimento que inicialmente pode estar relacionado ao evento.¹³ Ou seja, pouco conhecemos a respeito das potencialidades positivas ou negativas no uso de sistemas inteligentes no Judiciário e estamos a tecer comentários sem antes verificar o efetivo uso na prática forense.

Pois bem, retornando ao objeto central de discussão: uso de algoritmos de retropropagação no Judiciário, é importante esclarecer que a programação seria realizada com um objetivo de entrada (acidente de trânsito, erro médico, fato do produto, reclamatória trabalhista por horas extras etc.), e vários caminhos que serão testados para atingir o fim almejado: indenização material ou não; vínculo empregatício ou não; cumprimento de todas as

11 MEDEIROS, Nathália Roberta Fett Viana de. **Uso da inteligência artificial no processo de toma de decisões jurisdicionais**: uma análise sob a perspectiva da teoria normativa da participação. Dissertação (Mestrado em Direito) – Pós-Graduação em Direito. Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, 2019, p. 36.

12 LUGER, George F. **Inteligência artificial**. Tradução de Daniel Vieira. 6ª ed. São Paulo: Pearson Education do Brasil, 2013, p. 386.

13 Já sabemos que o nosso raciocínio jurídico-probatório é este: temos uma determinada convicção sobre a verificação de um facto; à medida que a prova vai sendo introduzida no processo, vai-se atualizando essa nossa convicção sobre a verificação do facto. O movimento é oscilatório, e pode ser para cima como para baixo, dependendo da prova. Portanto, a ideia é que o teorema de Bayes é uma fórmula matemática que capta matematicamente este movimento. (...) O teorema de Bayes tem sido visto como uma ferramenta essencial para o desenvolvimento da inteligência artificial. Isto porque um sistema de inteligência artificial tem de tomar decisões: essa é uma das características da nossa inteligência. Tem também de tomar decisões em condições de incerteza, fazer inferências, e tudo isso é, aparentemente, ou pode ser eventualmente traduzido através do teorema de Bayes. Sobre o assunto v. MARTINS, João Marques. Inteligência artificial e utilização de redes bayesianas na construção de argumentos probatórios. In: LOPES ROCHA, Manuel; SOARES PEREIRA, Rui (coord.). **Inteligência Artificial & Direito**. Coimbra: Almedina, 2020, p. 77/92, p. 78/79.

normas técnicas ou não etc.

Quando o algoritmo chega ao resultado almejado, por exemplo, analisa os pressupostos da responsabilidade civil (conduta, nexo de causalidade e dano) e verifica que o caso é de indenização material, o caminho feito até esse objetivo recebe um peso maior na conta matemática. Dessa forma, as camadas neurais internas (*hidden layers*) mais assertivas passam a comandar a atividade, entregando resultados mais precisos, vez que o algoritmo confere um peso maior àquelas conexões que apresentam resultados mais peremptórios ou pelo menos próximos do almejado.

Se, por exemplo, em outra análise o algoritmo entender que o caso é de compensação pelos danos morais suportados, porém não de dano material, esse caminho receberá um maior peso na conta e, assim por diante. A cada nova operação o sistema irá se atualizar, seja com doutrina, jurisprudência, análise de casos semelhantes, entregando cada vez mais uma resposta precisa, de acordo com o proposto inicialmente e o seu autoaprendizado. Ou seja, o algoritmo conseguirá iniciar uma camada de saída com informações mínimas, propagando o conteúdo retroativamente através das camadas ocultas até atingir o objetivo, isso, claro, em questão de minutos.¹⁴

O aprendizado em redes multicamadas permite que o algoritmo de retropropagação funcione, portanto, baseado em erros e acertos, com identificação paulatina dos caminhos e decisões mais corretas para atingir os objetivos.¹⁵ Por isso que esses algoritmos têm a capacidade de organizar amostras sem que exista uma classe predefina, sendo excelente para a prolação de decisões judiciais, bastando que o operador do Direito insira informações preliminares, através de formulários, o que já é feito em menor escala no peticionamento eletrônico, apontando também o tipo de demanda para que o algoritmo chegue à decisão final, a partir da análise do conteúdo dos autos.

Os algoritmos de retropropagação já são utilizados por instituições financeiras para a aprovação da concessão de empréstimos; ou seja, já se tem uma ideia de como o sistema se comporta em operações complexas, que envolve grande análise de dados, previamente rotulados como positivos ou negativos, podendo, dessa forma, ser adaptado para gerar uma decisão final de procedência, improcedência ou parcial procedência.

Não obstante, em um momento em que cada vez mais decisões tipicamente humanas são relegadas a algoritmos, é fundamental refletir sobre as situações em que a automação da decisão pode gerar resultados positivos, que estejam em conformidade com a preservação da democracia, a partir dos parâmetros delineados pela Constituição Federal.

14 LUGER, George F. **Inteligência artificial**. Tradução de Daniel Vieira. 6ª ed. São Paulo: Pearson Education do Brasil, 2013, p. 387.

15 BECKER, Daniel; FERRARI, Isabela. **A prática jurídica em tempos exponenciais**. JOTA: Opinião e Análise, [04 de outubro de 2017]. Disponível em: <https://jota.info/artigos/a-pratica-juridica-em-tempos-exponenciais-04102017>. Acesso em: 08 de mar. 2023.

Oportuno, apontar, que em razão dos riscos, e, às vezes, da irreversibilidade da decisão lançada por um algoritmo, a vulnerabilidade dos indivíduos envolvidos torna-se muito maior, sendo imprescindível a proposição de parâmetros jurídicos, assegurando-se, ao final, a coexistência da eficácia dos direitos humanos e fundamentais.¹⁶

Na prolação de decisões, os próprios magistrados podem ser parte do “filtro”, também aqui, o Poder Judiciário pode – e deve – exigir certas explicações de como esses algoritmos funcionam, para assegurar que são de confiança e justos. Operadores do direito, como promotores, defensores, auxiliares, também são susceptíveis de pressionar os desenvolvedores dos sistemas inteligentes para exigir explicações, ao mesmo tempo, em que se adequam a nova realidade tecnológica.¹⁷

Uma vez fornecidas as explicações, caso se entenda que elas são necessárias, o algoritmo decisional terá realizado a tarefa para o qual foi criado: a prestação jurisdicional por meio da entrega de uma decisão final.

Ou seja, no que tange a atividades jurisdicionais, a inteligência artificial pode auxiliar no trabalho de juízes e servidores, especialmente em atividades de ordem repetitiva, possibilitando uma resposta jurisdicional mais célere e efetiva, já que é possível se dedicar mais tempo a tarefas complexas.¹⁸ Ademais, não há dúvidas que despachos, decisão e sentenças de matérias de baixa complexidade e que se caracterizam como repetitivas poderão ser realizadas por meio de inteligência artificial, deixando ainda mais real a figura do chamado “juiz robô”.

A título exemplificativo, podemos citar o uso da inteligência artificial no TJDFT:

o TJDFT também vem investindo em iniciativas de inteligência artificial, com a finalidade de auxiliar o dia a dia dos servidores e magistrados - reduzindo atividades repetitivas e agregando agilidade e precisão na entrega de resultados à sociedade. Como exemplo, o projeto Hórus permitiu a classificação de documentos para distribuição de 274 mil processos de modo automático da VEF. O Amon, em fase de teste, permite reconhecimento facial a partir de imagens e vídeos atendendo algumas necessidades da segurança interna do TJ, além da possibilidade de auxiliar a VEPERA no processo de apresentação de presos em regime aberto. Já o Toth, em fase de estudos, permitirá análise da petição inicial do advogado buscando recomendar a classe e os assuntos processuais a serem

16 SARLET, Ingo Wolfgang. **A eficácia dos direitos fundamentais**. 12ª ed. Porto Alegre: Livraria do Advogado, 2017, p. 405-406.

17 DEEKS, Ashley. The judicial demand for explainable artificial intelligence. **Columbia Law Review**, Columbia, v. 119, n. 7, p. 1.829-1850, dez. 2019, p. 22.

18 TJDFT. Disponível em: <https://www.tjdft.jus.br/institucional/imprensa/campanhas-e-produtos/artigos-discursos-e-entrevistas/artigos/2020/inteligencia-artificial>. Acesso em: 20/05/2023.

cadastrados no PJE durante a autuação.¹⁹

Em suma, o Poder Judiciário brasileiro não pode ficar equidistante às inovações científicas. Ainda que a entrega da prestação jurisdicional demande um certo tempo, é possível a otimização de etapas mortas com a utilização de procedimentos digitais, bem como utilizando-se de algoritmos de retropropagação na tomada de decisões.

O momento é de ousar. Não temer a concorrência da IA nas profissões jurídicas, vendo-a como um importante recurso de apoio ao magistrado na tomada de decisões jurisdicionais.

3. CONCLUSÃO

A inteligência artificial já é uma realidade em nossa vida cotidiana, assim como na atividade jurisdicional. Consultas realizadas em sites de buscas de tribunais, compras direcionadas a partir da análise do perfil do usuário, recomendações em redes sociais com base nas interações, reconhecimento facial, processamento de voz humana, equipamentos acionados a partir de uma movimentação corporal, atendimento *on line* por softwares especialistas (*chatbots*), são apenas alguns exemplos da aplicabilidade da inteligência artificial em nossas vidas.

Não há como fugir dela e, definitivamente, a tendência é que seja cada vez mais utilizada, auxiliando em atividades simples e complexas. No entanto, a inteligência artificial pode nos auxiliar ainda mais, especialmente no âmbito do Poder Judiciário.

Por meio da inteligência artificial, seja através do aprendizado supervisionado ou não, uma máquina é capaz de realizar uma tarefa executada por humanos de forma otimizada e com maior rapidez, trazendo uma série de benefícios.

De fato, muitas questões terão que ser pensadas, em especial quanto à forma de revisão de tais decisões, o que deve passar, necessariamente, por um ser humano, bem como o estabelecimento de mecanismos que garantam e possibilitem a fiscalização de que, efetivamente, essa revisão humana foi realizada, como efetiva garantia a partes e procuradores.

Assim como em qualquer inovação, problemas surgirão, e demandarão soluções. Mas há muitos problemas a serem solucionados no Poder Judiciário brasileiro. A Justiça Brasileira enfrenta sérios óbices, e um deles é a morosidade com que as demandas são julgadas. Com mais de 80 milhões de processos em andamento, a inteligência artificial

19 Em outros países, o uso da inteligência artificial está sendo testado para fins de confeccionar sentenças judiciais. Vide: G1. Juiz robô usa ChatGPT para redigir sentença em caso de criança autista na Colômbia. Disponível em: <https://g1.globo.com/tecnologia/noticia/2023/02/03/juiz-usa-robo-chatgpt-para-redigir-sentenca-de-caso-de-crianca-autista-na-colombia.ghtml>. Acesso em: 20/05/2023.

auxiliaria neste problema, conferindo agilidade na tomada de decisão, principalmente nas demandas repetitivas.

Outro ponto em que a inteligência artificial atuaria de forma positiva seria como limitador de preconceitos na tomada de decisão judicial. O machismo e o racismo é estrutural e assim, se reflete e propaga nas decisões judiciais. Para ao menos minimizarmos este problema, basta que o *machine learning* seja alimentado com sentenças pragmáticas e que não perpetuem preconceitos, possibilitando um controle que não pode ser realizado com seres humanos. A inteligência artificial, ainda, pode atuar como forma de apontar eventuais padrões decisórios nas sentenças de magistrados, trazendo resultados comparativos que demonstrem, por exemplo, que quando um determinado julgador julga questão “x” em relação a pessoas de determinada cidade ou região, decide, em 80% das vezes, de forma pior comparando a pessoas de outra região. A mesma análise pode ser feita com marcadores raciais, etários ou de ordem social.

De fato, tais mecanismos podem causar receio aos julgadores, já que suas decisões passarão a ser monitoradas e analisadas facilmente pelas partes e procuradores. No entanto, trata-se de ferramenta de grande valia na fiscalização de decisões judiciais, atuando de forma a melhorar a qualidade de recursos eventualmente apresentados.

Por outro lado, há certas demandas que ainda permanecerão distantes do juiz robô, como as de natureza criminal e de direito de família, uma vez que se trata de dados sensíveis, assim como diante da excessiva complexidade e peculiaridades pertencentes a cada caso. Ainda assim, as decisões poderão ser analisadas à luz dos marcadores constantes no julgado, auxiliando os procuradores na fundamentação de eventuais recursos judiciais.

Sabe-se que os dados utilizados no processo de aprendizado da máquina e que direcionam a IA no processo de tomada de decisão devem ser provenientes de fontes seguras, preferencialmente governamentais, passíveis de serem rastreados e auditados. Ademais, no seu processo de tratamento, os dados utilizados devem ser protegidos contra riscos de destruição, modificação, extravio, acessos e transmissões não autorizadas.

O uso da IA deve respeitar a privacidade dos usuários, cabendo somente a estes a ciência e o controle sobre o uso de seus dados privados e pessoais. Desta forma, os dados que forem coletados pela IA devem ser utilizados de forma responsável para a proteção do usuário, garantindo-se e fomentando-se a dignidade humana.

De qualquer forma, não há dúvidas de que, negar a utilização da inteligência artificial em sede jurisdicional, não é uma alternativa a ser cogitada. Ao longo da história, constatamos que o homem sempre teve receio das novas tecnologias, no entanto, a história igualmente nos mostra que a união do homem e da máquina é inevitável e, se usada corretamente, sempre traz avanços e contribuições.

4. Referências

BECKER, Daniel; FERRARI, Isabela. **A prática jurídica em tempos exponenciais**. JOTA: Opinião e Análise, [04 de outubro de 2017]. Disponível em: <https://jota.info/artigos/a-pratica-juridica-em-tempos-exponenciais-04102017>. Acesso em: 08 de mar. 2023.

BRASIL. Conselho Nacional de Justiça. **Relatório de acompanhamento da estratégia nacional: 2016/2017**. Brasília: CNJ, 2018. Disponível em: <https://www.cnj.jus.br/wp-content/uploads/2022/11/relatorio-de-acompanhamento-da-estrategia-nacional-2021-v2-2022-05-31.pdf>. Acesso em: 08 de mar. 2023.

CASTILHOS, Guilherme Machado de; POLL, Roberta Eggert; CASTILHOS, Aline Pires de Souza Machado de. “E se a sua geladeira pudesse depor contra você no Tribunal?”. *Internet das Coisas e provas no processo penal brasileiro*. **Revista Brasileira de Ciências Criminais**, São Paulo, v. 163, n. 28, p. 363-391, jan. 2020.

CNJ. **Juízes apontam como algoritmos podem ajudar Justiça a melhorar julgamentos**. Disponível em: <https://www.cnj.jus.br/juizes-apontam-como-algoritmos-podem-ajudar-justica-a-melhorar-julgamentos/>. Acesso em: 20 de maio 2023.

DEEKS, Ashley. The judicial demand for explainable artificial intelligence. **Columbia Law Review**, Columbia, v. 119, n. 7, p. 1.829-1850, dez. 2019.

G1. **Juiz robô usa ChatGPT para redigir sentença em caso de criança autista na Colômbia**. Disponível em: <https://g1.globo.com/tecnologia/noticia/2023/02/03/juiz-usa-robo-chatgpt-para-redigir-sentenca-de-caso-de-crianca-autista-na-colombia.ghtml>. Acesso em: 20 de maio 2023.

Gellego, Jellor (2016), **‘An AI-Powered Chatbot Is Helping Doctors Diagnose Patients’**. Retrieved from: <http://futurism.com/an-ai-powered-chatbot-is-helping-doctors-diagnose-patients/>. Acesso em: 16 de maio 2023.

Knight, Will (2016), *‘AI’s language problem’*. **MIT Technology Review**. Disponível em: <https://www.technologyreview.com/s/602094/ais-language-problem>. Acesso em: 16 maio 2023.

LEAL, Ana Alves. *Big data* nos setores bancário e financeiro. In: LOPES ROCHA, Manuel; SOARES PEREIRA, Rui (coord.). **Inteligência Artificial & Direito**. Coimbra: Almedina, 2020, p. 199/220.

LUGER, George F. **Inteligência artificial**. Tradução de Daniel Vieira. 6ª ed. São Paulo: Pearson Education do Brasil, 2013.

MARTINEZ, Vinício Carrilho; SCHERCH, Vinícius Alves. Relações entre direito e tecnologia no século XXI. **Revista de Direito**, Viçosa, v. 12, n. 1, p. 1-23, jan./abr. 2020.

MARTINS, João Marques. Inteligência artificial e utilização de redes bayesianas na construção de argumentos probatórios. In: LOPES ROCHA, Manuel; SOARES PEREIRA, Rui (coord.). **Inteligência Artificial & Direito**. Coimbra: Almedina, 2020, p. 77/92.

MEDEIROS, Nathália Roberta Fett Viana de. **Uso da inteligência artificial no processo de toma de decisões jurisdicionais**: uma análise sob a perspectiva da teoria normativa da participação. Dissertação (Mestrado em Direito) – Pós-Graduação em Direito. Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, 2019.

NEVES, Daniel Amorim Assumpção. **Manual de Direito Processual Civil**. Volume único. 10^a ed. E-book. Salvador: JusPodivm, 2018.

OLIVEIRA, Fabiana Luci de; CUNHA, Luciana Gross. Os indicadores sobre o Judiciários brasileiro: limitações, desafios e o uso da tecnologia. **Revista Direito FGV**, São Paulo, v. 16, n. 01, p. 1-23, jan./abr. 2020.

SARLET, Ingo Wolfgang. **A eficácia dos direitos fundamentais**. 12^a ed. Porto Alegre: Livraria do Advogado, 2017.

TJDFT. Disponível em: <https://www.tjdft.jus.br/institucional/imprensa/campanhas-e-produtos/artigos-discursos-e-entrevistas/artigos/2020/inteligencia-artificial>. Acesso em: 20 de maio 2023.

WATSON, Willis Towers (2018). Horizon Scanning forward thinking. Artificial Intelligence and the legal profession. **The Law Society**. Disponível em: <https://www.lawsociety.org.uk/topics/research/ai-artificial-intelligence-and-the-legal-profession>. Acesso em: 16 de maio 2023.

RAFA 2030 - CLASSIFICAÇÃO DE TEXTOS JURÍDICOS EM OBJETIVOS DE DESENVOLVIMENTO SUSTENTÁVEL DA AGENDA 2030 DAS NAÇÕES UNIDAS

RAFA 2030 - LEGAL TEXT CLASSIFICATION IN UN 2030 AGENDA SUSTAINABLE DEVELOPMENT GOALS

Alexandre Ferreira de Menezes²⁰

Euler Rodrigues de Alencar²¹

Ian Ferrare Meier²²

Lucas José Gonçalves Freitas²³

Paulo Henrique Vencio Moreira²⁴

Regis Proença Picanço²⁵

Rodrigo Lobo Canalli²⁶

Resumo

O Supremo Tribunal Federal (STF), instância máxima do sistema judiciário brasileiro, produz uma imensa quantidade de dados geralmente organizados em forma de texto, por meio de decisões, petições, liminares, recursos e outros documentos legais. Categorizar

20 Especialista em Sistemas Orientados a Objetos pela Universidade Católica de Brasília (UCB); técnico judiciário do Supremo Tribunal Federal.

21 Bacharel e Mestre em estatística pela Universidade de Brasília (UnB); analista judiciário do Supremo Tribunal Federal.

22 Bacharel em Direito pelo Centro Universitário de Brasília (CEUB); pesquisador da BI Norwegian Business School.

23 Bacharel em estatística pela ENCE; mestre em estatística pela Universidade de Brasília (UnB); analista judiciário do Supremo Tribunal Federal.

24 Bacharel em processamento de dados com MBA em gestão de projetos, ambos pela Universidade Católica de Brasília (UCB); analista judiciário do Supremo Tribunal Federal.

25 Engenheiro de Redes de Comunicação pela Universidade de Brasília (UnB); especialista em cybersegurança pelo Instituto IDESP.

26 Mestre em Direito, Estado e Constituição pela Universidade de Brasília, mestre em Competition, Innovation and Information Law pela New York University, pesquisador do Ethics4AI, professor. Foi Assessor-Chefe de Inteligência Artificial do Supremo Tribunal Federal.

textos com etiquetas serve ao propósito de agrupar processos parecidos e auxiliar na recuperação de informações. Neste contexto, uma ferramenta de aprendizagem supervisionada para auxiliar na classificação de documentos jurídicos nos 17 Objetivos de Desenvolvimento Sustentável (ODS) da Agenda 2030 da Organização das Nações Unidas (ONU) é de grande utilidade para o tribunal, uma vez que essa tarefa é realizada manualmente por um grupo de funcionários da corte e a adoção da Agenda 2030 da ONU é um objetivo do tribunal atualmente. O objetivo geral deste projeto, denominado Redes Artificiais com Foco na Agenda 2030 (RAFA 2030), é gerar valor para o tribunal por meio da construção de sistemas de classificação baseados em Processamento de Linguagem Natural (PLN). Atualmente, a principal entrega deste projeto consiste em um aplicativo desenvolvido através do *framework Shiny* (linguagem R) contendo ferramentas gráficas para PLN (gráfico de co-ocorrência e nuvem de palavras), algoritmos de redes neurais para sugestões de ODS em peças jurídicas e motores para localização de leis, busca por contexto e contagem de palavras-chave em textos. Avaliações internas indicam imenso potencial para utilização da iniciativa RAFA 2030 no apoio à classificação de documentos jurídicos não apenas em temas da Agenda 2030, mas em etiquetas de outras naturezas também.

Palavras-chave: Agenda 2030 da ONU; Aprendizagem de Máquina; Aprendizagem Profunda; Classificação de Textos.

Abstract

The Federal Supreme Court, the highest court of the Brazilian judicial system, produces an immense amount of data generally organized in text form, through decisions, petitions, injunctions, appeals and other legal documents. Categorizing text with tags serves the purpose of grouping similar processes and aiding in information retrieval. In this context, a supervised learning tool to assist in the classification of legal documents in the 17 Sustainable Development Goals (SDGs) of the 2030 Agenda of the United Nations (UN) is of great use for the court, since this task is performed manually by a group of court staff and the adoption of the UN 2030 Agenda is an objective of the court today. The general objective of this project, called RAFA 2030 (acronym in Portuguese for Artificial Networks with a Focus on the 2030 Agenda), is to build classification systems based on Natural Language Processing (NLP). Currently, the main product of this project consists of an application developed using the Shiny framework (language R) containing graphical tools for NLP (cooccurrence graphs and word cloud), neural network algorithms for ODS suggestions in legal pieces, word-based context search and keyword counting in texts. Internal evaluations indicate immense potential for using the RAFA 2030 initiative to support the classification of legal documents not only in terms of the 2030 Agenda, but in other labels and tags.

Keywords: UN 2030 Agenda; Machine Learning; Deep Learning; Text Classification.

1. INTRODUÇÃO

A Agenda 2030 da Organização das Nações Unidas é um plano de ação para as pessoas, o planeta, a prosperidade, a paz universal e a liberdade.²⁷ Criado em 2015, esse plano de ação, que envolve objetivos como erradicação da pobreza, redução da desigualdade, preservação do meio ambiente e economia sustentável, está organizado em 17 ODS. Todos os ODS são integrados²⁸ - a ação em uma área afetará os resultados em outras -, mas os textos de cada um são completamente diferentes em sua essência, de forma que identificá-los como um rótulo em textos legais é um importante passo para automatizar tarefas repetitivas e melhorar fluxos de trabalho de servidores de tribunais lotados em setores de atuação, por exemplo. Incluir a Agenda 2030 da ONU no dia-a-dia dos tribunais é uma estratégia para tornar a justiça mais eficiente, pois as decisões judiciais devem afetar positivamente a vida das pessoas. A Agenda é, portanto, uma ferramenta importante nesse sentido. Na prática, rotular ações ou textos jurídicos de acordo com ela pode ser uma forma de antecipar julgamentos e decisões com importante impacto social e assim melhor organizar as entregas do poder judiciário à população como um todo e especificamente aos jurisdicionados.

O Supremo Tribunal Federal possui um *hotsite*²⁹ onde informações sobre a Agenda 2030 estão disponíveis, bem como um painel contendo os metadados de processos atualmente etiquetados com ODS na corte. Tais dados podem ser facilmente baixados em formato .xlsx (Excel) através do painel, que também apresenta links de outras iniciativas sobre a Agenda 2030 no poder judiciário, especificamente no âmbito do Conselho Nacional de Justiça (CNJ), cuja missão constitucional é controlar a atuação administrativa e financeira de mais de 90 órgãos do poder judiciário, como tribunais estaduais, tribunais regionais federais, tribunais regionais do trabalho e tribunais regionais eleitorais, por exemplo.³⁰ O Poder Judiciário Brasileiro é pioneiro, no mundo, na institucionalização da Agenda e a indexação de sua base de dados (processos judiciais) com base nos 17 Objetivos de Desenvolvimento Sustentável (ODS). O Supremo Tribunal Federal, em especial, a institucionalizou através da Resolução nº 710, de 20 de novembro de 2020.³¹ Dessa forma, o trabalho busca responder perguntas como: quais são os direitos fundamentais

27 NAÇÕES UNIDAS NO BRASIL. **Homepage dos Objetivos de Desenvolvimento Sustentável no Brasil**, 2015b. Disponível em: <https://brasil.un.org/pt-br/sdgs>. Acesso em: 29 mai. 2023.

28 NAÇÕES UNIDAS NO BRASIL. **Agenda 2030 para o Desenvolvimento Sustentável**, 2015a. Disponível em: <https://brasil.un.org/pt-br/91863-agenda-2030-para-o-desenvolvimento-sustent%C3%A1vel>. Acesso em: 29 mai. 2023.

29 SUPREMO TRIBUNAL FEDERAL. **Hotsite da Agenda 2030 para o Supremo Tribunal Federal (STF). Relação da Agenda 2030 com o STF**, 2022. Disponível em: <http://portal.stf.jus.br/hotsites/agenda-2030>. Acesso em: 29 mai. 2023.

30 CONSELHO NACIONAL DE JUSTIÇA. **Quem somos**, 2023. Disponível em: <https://www.cnj.jus.br/sobre-o-cnj/quem-somos/>. Acesso em: 29 mai. 2023.

31 SUPREMO TRIBUNAL FEDERAL. **Resolução nº 710, de 20 de novembro de 2020, 2020**. Disponível em: <https://portal.stf.jus.br/hotsites/agenda-2030/assets/img/RESOLUCAO710-2020.PDF>. Acesso em: 29 mai. 2023.

presentes nos ODS da Agenda 2030 com maior número de violações observadas pelo Poder Judiciário? É possível estabelecer, no futuro, alguma precedência nos julgamentos que envolvam ODS da Agenda 2030? A figura a seguir mostra exemplos dos objetivos:

Figura 1 - Objetivos de Desenvolvimento Sustentável da Agenda 2030



Fonte: arquivo pessoal dos autores, 2023.

Incluir a agenda ONU 2030 no cotidiano de órgãos públicos é uma maneira eficiente de fazer com que pessoas se beneficiem deste pacto global em nível individual, aumentando o caráter prático da própria agenda. Além dos 17 objetivos de desenvolvimento sustentável, a Agenda 2030 conta com 169 metas que, por essência, podem ser utilizadas como indicadores da implantação da agenda e também de políticas públicas de forma isolada.

2. OBJETIVOS

Dado que funcionários do STF em seu fluxo diário de trabalho classificam manualmente processos novos e antigos que compõem o acervo da corte, o objetivo geral deste trabalho é apresentar ferramentas capazes de auxiliar os servidores na etapa de autuação/etiquetagem. Importante notar que o caso específico da Agenda 2030 é apenas um entre tantas aplicações possíveis para classificação de textos jurídicos, dado que abordagens parecidas com a que será apresentada nas seções a seguir já foram utilizadas no próprio STF em contextos de *cibersegurança*, por exemplo. Fixando dois eixos de atuação - sugestão de etiquetas/ODS e ferramentas gráficas de apoio à atividade de classificação - a iniciativa RAFA 2030 pode ser utilizada parcialmente ou em sua totalidade por diversos setores do tribunal, dado que possui caráter geral e percorre todo fluxo de PNL para classificação textual: entrada dos dados (em formato PDF), limpeza dos textos, *embedding*, modelos de classificação e visualização em forma de aplicativo.

O objetivo específico da ferramenta é difundir o uso de estatística e aprendizagem de máquina em contextos jurídicos de tomada de decisão. Trata-se de universalizar o acesso a ferramentas de *data science*, para usuários leigos de outras áreas. O uso da ferramenta pode ser o passo inicial para que setores exclusivamente jurídicos façam uso de dados no dia a dia. No aplicativo existem objetos e resultados que não dependem de profundo conhecimento estatístico para interpretação, tais como a nuvem de palavras e grafo de coocorrência. A figura 2 mostra a interface do aplicativo em Shiny.³²

Figura 2 - Interface do aplicativo desenvolvido em Shiny



Fonte: arquivo pessoal dos autores, 2023.

A ferramenta se destaca, também, pela presença de uma busca de contexto para palavras de interesse. Funciona de maneira intuitiva, pois os usuários escrevem palavras num campo de pesquisa similar aos disponíveis em motores de busca. O aplicativo, então, retorna a ocorrência da palavra no texto com uma janela de palavras anteriores e posteriores, permitindo identificar qual é a vizinhança da palavra buscada. Esta funcionalidade contém também um buscador de leis e artigos legais, que permite a imediata identificação da legislação citada no texto.

Em 2022, através do Ato Regulamentar nº 26, de 19 de dezembro de 2022,³³ foi criada a Assessoria de Inteligência Artificial do Supremo Tribunal Federal. A partir da sua criação, a assessoria coordenaria os trabalhos do tribunal envolvendo o uso de inteligência artificial, a exemplo da RAFA 2030.

32 WICKHAM, H. **Mastering Shiny: Build Interactive Apps, Reports, and Dashboards Powered by R.** O'Reilly, 2021.

33 SUPREMO TRIBUNAL FEDERAL. **Ato Regulamentar nº 26, de 19 de dezembro de 2020**, 2020. Disponível em: <https://www.stf.jus.br/arquivo/norma/atoregulamentar026-2022.pdf>. Acesso em: 10 out. 2023.

3. BASE E PRÉ-PROCESSAMENTO DOS DADOS

Conjunto de dados formado por 1,8 mil acórdãos e petições iniciais rotuladas por especialistas do STF e cerca de 40 mil acórdãos não rotulados extraídos do acervo do tribunal. Estratégias de *data augmentation* e similaridade entre textos são aplicadas para tornar mais denso o conjunto etiquetado, usando os textos sem rótulos como entrada. Neste primeiro momento, apenas os ODS 16 (Paz, Justiça e Instituições Eficazes), 10 (Redução de Desigualdades), 8 (Trabalho Decente) e 3 (Boa Saúde) foram considerados para a sugestão de classificação via redes neurais, pois são os maiores ODS em número de processos etiquetados.

O pré-processamento é muito importante em fluxos de PLN, pois aumenta o desempenho dos modelos e reduz a complexidade computacional das rotinas de classificação textual. Alguns dos principais passos de limpeza utilizados na iniciativa RAFA 2030 são:

- extração de texto de PDFs com imagens e em arquivos nato digitais;
- remoção de palavras de parada (*stopwords*);
- remoção de caracteres especiais, como #, @ e &;
- remoção de espaços em branco desnecessários;
- remoção de termos legais e números desnecessários;
- forma minúscula em todas as palavras;
- estratégia pos-tag para manutenção de termos gramaticalmente relevantes;
- remoção de cabeçalhos e rodapés;
- anonimização de dados sensíveis;
- tratamento de leis, artigos e normas jurídicas.

A etapa de *embedding* (vetorização) é feita para transformar textos em informações numéricas e é imediatamente subsequente à limpeza dos textos. A etapa de vetorização muda de acordo com a finalidade do objeto. Para as redes ajustadas na iniciativa RAFA 2030, foram utilizados os fluxos disponíveis no *framework pytorch*, nativo da linguagem

Python e modelos ajustados em linguagem R.³⁴ Para ferramentas de visualização, foram utilizados embeddings contextuais com passo de *pos-tag* em fluxos de gráficos no pacote *ggplot2* da linguagem R. Trabalhos futuros certamente irão envolver representações de palavras e textos obtidos via arquitetura *transformers*³⁵ e modelos da família GPT, sendo este o atual estado da arte em PLN.

4. RESULTADOS

Os resultados iniciais que serão apresentados a seguir tratam apenas de aprendizagem de máquina, em razão do aplicativo estar em fase de testes, e demonstram que as redes neurais podem auxiliar no processo de rotulagem de textos jurídicos. A medida de acurácia para os 4 ODS escolhidos na amostra de teste (de tamanho 199) pode ser vista na Tabela 1:

Tabela 1 - Acurácias por ODS

	ODS 16	ODS 10	ODS 8	ODS 3
Acertos	187	189	195	197
Erros	12	10	4	2

Fonte: elaboração própria dos autores, 2020.

A matriz de confusão para o ODS 16 é dada pela Figura 3:

Figura 3 - Matriz de Confusão para o ODS 16

Prediction	1-	109	7
	0-	5	78
		1	0
		Truth	

Fonte: arquivo pessoal dos autores, 2023.

34 SILGE, J.; HVITFELDT, E. **Supervised Machine Learning for Text Analysis in R**. Chapman and Hall/CRC, 2021. Disponível em: <https://smltar.com/references.html>. Acesso em: 29 mai. 2023.

35 TUNSTALL, L.; WOLF, T.; WERRA, L. V. **Natural Language Processing With Transformers: Building Language Applications With Hugging Face**. O'Reilly, 2022.

É possível verificar que as redes performam bem nas duas categorias - com etiqueta ODS 16 e sem etiqueta ODS 16. As métricas F1-Score e Precisão também estão boas. Os resultados também são submetidos à validação humana, pelos mesmos especialistas que confeccionam as etiquetas utilizadas no treinamento supervisionado. Existem painéis automáticos (*dashboards*) desenvolvidos em *QlikSense* e *PowerBI*, dedicados ao monitoramento das métricas de interesse, além de reuniões periódicas com funcionários especialistas da área jurídica do tribunal, usuários frequentes do aplicativo em suas rotinas diárias de trabalho. Tais encontros servem para equilibrar as estratégias de classificação e realizar uma curadoria ativa dos resultados obtidos pelos classificadores embarcados no aplicativo. Os resultados e os feedbacks coletados junto de usuários do aplicativo são positivos, a experiência de internalizar a Agenda 2030 no STF foi fortemente impactada pelo uso de recursos tecnológicos. Os eventos de apresentação da ferramenta³⁶ serviram também ao propósito de divulgar os ODS da Agenda, para além do simples uso do aplicativo em vários setores do tribunal. O painel de controle das etiquetas pode ser visto na figura a seguir:

Figura 4 - Painel de controle das etiquetas de ODS

Processos Monitorados: 2108			Ocorrências de ODS: 3336		
Indicação de processos com aderência à Agenda 2030 no STF					
ODS1	ODS2	ODS3	ODS4	ODS5	ODS6
Erradicação da Pobreza	Fome Zero e Agricultura Sustent.	Saúde e Bem-Estar	Educação de Qualidade	Igualdade de Gênero	Água Potável e Saneamento
62	42	324	122	36	39
ODS7	ODS8	ODS9	ODS10	ODS11	ODS12
Energia Acessível e Limpa	Trabalho Decente e Cresc. Econ.	Indústria, Inovação e Infra	Redução das Desigualdades	Cidades e Comunidades Sust.	Consumo e Produção Respons.
53	441	104	325	92	66
ODS13	ODS14	ODS15	ODS16	ODS17	
Ação Contra Mud. Global/Clima	Vida na Água	Vida Terrestre	Paz, Justiça e Instituições Eficazes	Parcerias e Meios de Implem.	
13	14	84	1.300	219	

Fonte: Hotsite da Agenda 2030 para o Supremo Tribunal Federal, 2023.

É possível observar que alguns ODS têm maior número de etiquetas, como o ODS 16 - Paz, Justiça e Instituições Eficazes, cujo tema é fortemente relacionado com a atuação de cortes constitucionais, a exemplo do STF. Vale registrar que, com a classificação de processos estabelecida, foi aberta a possibilidade de priorização dos julgamentos de ações que impactem positivamente os objetivos e metas da Agenda 2030, tendo, já no primeiro semestre de 2021, 76% dos processos incluídos na pauta do Plenário com uma ou mais etiquetas de ODS da Agenda 2030, conforme informado pelo

36 SUPREMO TRIBUNAL FEDERAL. Hotsite da Agenda 2030 para o Supremo Tribunal Federal (STF). Relação da Agenda 2030 com o STF, 2022. Disponível em: <http://portal.stf.jus.br/hotsites/agenda-2030>. Acesso em: 29 mai. 2023.

Ministro Luiz Fux no evento Suprema Corte e Diálogos sobre a Agenda 2030. A produção de dados sobre a classificação dos processos em ODS é importante para aferir a efetiva contribuição da Corte com a pauta da Agenda 2030 (*accountability*). A alteração dos critérios de identificação dos processos submetidos ao STF e a consequente alteração na seleção de matérias a serem julgadas representam um avanço na internacionalização da Corte com equiparação a outros tribunais constitucionais de todo o mundo. Destaca-se, neste ponto, o caráter inédito e pioneiro da iniciativa de classificação dos processos por ODS em uma Corte Suprema.

As ferramentas de apoio gráfico à classificação e contagem de palavras têm maior destaque na iniciativa, dado que são de caráter geral e permitem a análise de textos jurídicos diversos. A ideia de usar aprendizagem de máquina e automação em contextos legais, por advogados e tribunais, não deve ser centrada em substituir integralmente a mão de obra humana e especializada. A experiência da iniciativa RAFA 2030, ao contrário, é centrada na integração das inteligências artificial, de negócio e humana, de forma a reduzir demandas manuais e repetitivas. Como já mencionado anteriormente, a iniciativa auxilia na classificação dos textos e absorve tais classificações para melhorar a performance dos modelos, integrando a sabedoria dos funcionários do tribunal em tarefas de atuação e o poder dos computadores para geração de sugestões, gráficos e busca por termos importantes.

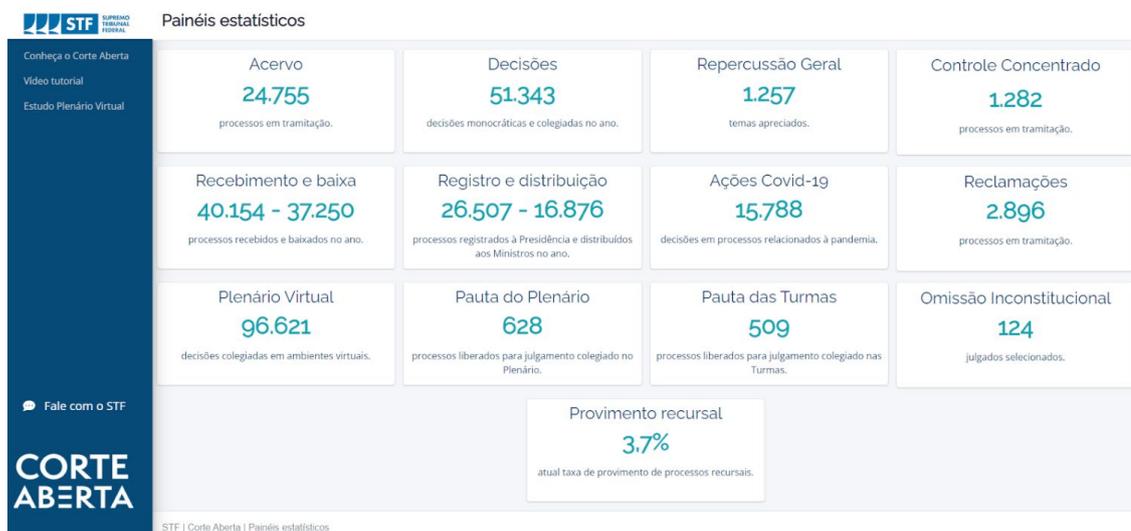
Segundo noticiado por Analista Judiciário do tribunal entrevistado pelos autores, desde o lançamento da iniciativa RAFA 2030, em maio de 2022, vários acórdãos foram avaliados com apoio do aplicativo, bem como 34 temas de repercussão geral, seja em sua relação com ODS da Agenda 2030 ou em atividades de afirmação de jurisprudência ou gerenciamento de precedentes. Trata-se de um número relevante, dado o fluxo processual da corte. Outro servidor observou que técnicos do tribunal criaram o hábito de submeter peças processuais ao aplicativo da iniciativa RAFA 2030 para observar os resultados da nuvem de palavras e contagem de termos relacionados aos ODS da Agenda, sem que a classificação nos próprios ODS seja o objetivo final da consulta.

Trabalhos futuros envolvem o aprimoramento da interface gráfica e testes de novas arquiteturas de redes neurais profundas, bem como novas abordagens semi-supervisionadas para sugerir rótulos para os 13 ODS com poucas entradas atualmente. Uma estratégia possível envolve aplicar métodos baseados em *few-shot learning* para rotular processos com base em poucos exemplos de entrada. Também podem ser utilizados métodos de *data augmentation* baseados em *back translation*, *easy data augmentation* e algoritmos de agrupamento. Neste último caso, agrupam-se peças com e sem etiquetas para, através de algum método de clusterização, reunir processos avaliados ou não por humanos. Com uso de alguma estratégia de propagação, rótulos originais são alocados para processos ainda não avaliados, como etiquetas sintéticas. O objetivo geral deste fluxo, que já está em fase de testes, é verificar se as bases aumentadas com etiquetas sintéticas melhoram as performances dos algoritmos de classificação processual. No caso específico da iniciativa RAFA 2030, redes neurais

profundas.

Como o Supremo Tribunal Federal dispõe de uma plataforma centrada na transparência de dados (Programa Corte Aberta), pode ser interessante incluir marcações de ODS da Agenda 2030 nos mesmos painéis em que aparecem estatísticas sobre o acervo processual, tempo de tramitação e repercussão geral. A ideia é totalmente aderente ao objetivo de internalizar a própria agenda no dia a dia da corte em seus mais diversos setores, administrativos e jurídicos. A figura 5 mostra o hub de painéis do programa Corte Aberta.

Figura 5 - Programa Corte Aberta



Fonte: Hotsite da Agenda 2030 para o Supremo Tribunal Federal, 2023.³⁷

Ainda existem pontos de melhoria na estratégia de utilização da ferramenta RAFA 2030, principalmente em relação aos sistemas de apoio judicial do tribunal, tornando-a parte ou módulo em meio às tarefas de autuação e deslocamento processual. As principais melhorias previstas são:

- conectar o aplicativo no sistema de apoio à atividade judicial, dispensando muitos cliques por parte do usuário. Esta melhoria permitirá que os usuários, num mesmo ambiente, façam classificações processuais em ODS da Agenda 2030 e deem sequência à tramitação dos processos dentro da corte;
- estudos sobre *User Experience* (UX) para melhor organizar as ferramentas gráficas disponíveis no aplicativo. Esta melhoria tornará a experiência de navegação no aplicativo mais fluida;

- criação de pacotes e classes auxiliares para redução de dependências no aplicativo. Esta melhoria permitirá que os textos em PDF sejam lidos mais rápido;
- embarcar maiores modelos no aplicativo. Esta melhoria permitirá que os modelos maiores ajustados em *Pytorch*, com redes neurais profundas e de arquiteturas mais complexas, sejam utilizados para sugestão automática de ODS da Agenda 2030 em novos textos;
- introduzir funcionalidades relacionadas ao uso dos modelos da família GPT, principalmente para sumarização de textos legais grandes;
- atualizar o design do aplicativo, tornando-o mais aderente ao sistema atual;
- melhorar a periodicidade de treinamento dos modelos de rede neural para sugestão de ODS.

A documentação inicial da iniciativa RAFA 2030 está disponível em um *hotsite*³⁸, onde se encontram extratos de códigos para várias tarefas de PLN, tais como similaridade, *embedding* via framework *StarSpace* e modelos de redes neurais simples desenvolvidas em *Keras* e *Tensorflow*.³⁹ A ideia principal da documentação é reunir códigos de todas as tentativas de modelos e tratamento de dados. O objetivo não é tornar a documentação excessivamente longa, mas ajudar de maneira geral. O que deu errado para a iniciativa RAFA 2030 pode ser decisivo na resolução de problemas de PLN em outros tribunais.

Internamente, existem painéis exclusivos para gerenciamento das métricas associadas a cada execução dos modelos em teste, algo que é bastante interessante para *pipelines* de aprendizagem de máquina. Na prática, cada técnico que testa modelos de redes neurais ou até mesmo aprendizagem de máquina - Naive Bayes, SVM, Random Forest e etc - inclui as métricas obtidas em painéis de controle, de forma que é fácil verificar a performance dos modelos em paralelo às ondas de litigiosidade que necessariamente tem impacto nas etiquetas de ODS da agenda 2030.

5. CONCLUSÕES

A ferramenta RAFA 2030 está em produção, isto é, em uso, em 4 unidades do tribunal, apoiando diariamente a classificação de peças em ODS da agenda 2030 da ONU. Muito embora ainda existam diversos pontos de melhoria, é possível destacar o impacto positivo da ferramenta pelos relatos dos servidores envolvidos na classificação em si e

38 SUPREMO TRIBUNAL FEDERAL. **Hotsite da Agenda 2030 para o Supremo Tribunal Federal (STF)**. Relação da Agenda 2030 com o STF, 2022. Disponível em: <http://portal.stf.jus.br/hotsites/agenda-2030>. Acesso em: 29 mai. 2023.

39 GÉRON, A. **Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems**. O'Reilly, 2022 - 3ed.

também pelo feedback recebido em eventos. Recentemente, a ferramenta foi centro de um debate sobre uso de inteligência artificial no STF, por exemplo. Vários órgãos públicos buscaram à época do lançamento e ainda buscam informações sobre a classificação em ODS da agenda, fundamentalmente com apoio computacional da ferramenta RAFA.

Parte das experiências adquiridas no desenvolvimento da ferramenta foram utilizadas na iniciativa VitorIA, do próprio Supremo Tribunal Federal. A VitorIA é um agrupador de textos cujo objetivo é tratar objetos parecidos em lote. Tal ferramenta, cujo lançamento ocorreu em maio de 2023, é capaz de identificar: a) novas oportunidades para criação de temas de Repercussão Geral; b) processos com matéria ou situação similar para serem tratadas em conjunto e sejam desde logo deslocadas para o local correto do fluxo processual; c) ondas de litigiosidade que podem reclamar o desenvolvimento de uma estratégia de enfrentamento eficiente.

A primeira fase do projeto abrange o agrupamento livre de processos recebidos por um modelo algorítmico de análise textual com base em similaridade. Em seguida, será possível buscar os processos semelhantes a determinado texto fornecido como paradigma.

As iniciativas RAFA 2030 e VitorIA são processos de trabalho da Assessoria de Inteligência Artificial (AIA - STF), cujo objetivo é desenvolver soluções em inteligência artificial aplicadas à prestação jurisdicional do STF, em conformidade com padrões de qualidade e eficiência, e com estrita observância dos direitos fundamentais e do devido processo legal. A estrutura da assessoria foi inspirada em modelos adotados em outros tribunais, tanto no Brasil quanto no exterior, e atende a uma necessidade emergente de aproveitar a capacidade da inteligência artificial na melhora da prestação jurisdicional.

6. Referências

CONSELHO NACIONAL DE JUSTIÇA. **Quem somos**, 2023. Disponível em: <https://www.cnj.jus.br/sobre-o-cnj/quem-somos/>. Acesso em: 29 mai. 2023.

GÉRON, A. **Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems**. O'Reilly, 2022 - 3ed.

NAÇÕES UNIDAS NO BRASIL. **Agenda 2030 para o Desenvolvimento Sustentável**, 2015a. Disponível em: <https://brasil.un.org/pt-br/91863-agenda-2030-para-o-desenvolvimento-sustent%C3%A1vel>. Acesso em: 29 mai. 2023.

NAÇÕES UNIDAS NO BRASIL. **Homepage dos Objetivos de Desenvolvimento Sustentável no Brasil**, 2015b. Disponível em: <https://brasil.un.org/pt-br/sdgs>. Acesso em: 29 mai. 2023.

SILGE, J.; HVITFELDT, E. **Supervised Machine Learning for Text Analysis in R. Chapman and Hall/CRC**, 2021. Disponível em: <https://smltar.com/references.html>. Acesso em: 29 mai. 2023.

SUPREMO TRIBUNAL FEDERAL. **Painéis estatísticos do Corte Aberta**, 2023. Disponível em: https://transparencia.stf.jus.br/extensions/corte_aberta/corte_aberta.html. Acesso em 10 out. 2023.

SUPREMO TRIBUNAL FEDERAL. **Hotsite da Agenda 2030 para o Supremo Tribunal Federal (STF)**. Relação da Agenda 2030 com o STF, 2022. Disponível em: <http://portal.stf.jus.br/hotsites/agenda-2030>. Acesso em: 29 mai. 2023.

SUPREMO TRIBUNAL FEDERAL. **Ato Regulamentar nº 26, de 19 de dezembro de 2020, 2020**. Disponível em: <https://www.stf.jus.br/arquivo/norma/atoregulamentar026-2022.pdf>. Acesso em: 10 out. 2023.

SUPREMO TRIBUNAL FEDERAL. **Resolução nº 710, de 20 de novembro de 2020, 2020**. Disponível em: <https://portal.stf.jus.br/hotsites/agenda-2030/assets/img/RESOLUCAO710-2020.PDF>. Acesso em: 29 mai. 2023.

TUNSTALL, L.; WOLF, T.; WERRA, L. V. **Natural Language Processing With Transformers: Building Language Applications With Hugging Face**. O'Reilly, 2022.

WICKHAM, H. **Mastering Shiny: Build Interactive Apps, Reports, and Dashboards Powered by R**. O'Reilly, 2021.

AS IMPLICAÇÕES SOB A PERSPECTIVA DA JUSTIÇA DE DADOS ACERCA DO USO DE TECNOLOGIAS DE RECONHECIMENTO FACIAL PARA GRATUIDADE DE TRANSPORTE PÚBLICO

IMPLICATIONS FROM THE DATA JUSTICE

PERSPECTIVE ON THE USE OF FACIAL RECOGNITION

TECHNOLOGIES FOR FREE PUBLIC TRANSPORTATION

Gedeão França⁴⁰

Izabella Bittencourt⁴¹

Júlia Mendonça⁴²

Paula Guedes Fernandes da Silva⁴³

Resumo

A adoção de tecnologias de reconhecimento facial (RF), baseadas em técnicas de Inteligência Artificial (IA), já é realidade no setor público. Isso ocorre inclusive para a efetivação de políticas públicas, como no caso da concessão do direito à gratuidade do transporte público. Ocorre que essa implementação pode gerar diferentes implicações sob a ótica da justiça de dados, envolvendo desde riscos à direitos fundamentais, discriminações a grupos vulnerabilizados, até a exclusão do acesso de parcela da população ao direito em questão. Desse modo, o presente trabalho irá analisar

40 Advogado. Pós-graduado em Processo Civil pela UniNassau e pós-graduando em Direito Digital e Proteção de Dados pelo IDP. Community Manager da Data Privacy Brasil. Pesquisador e Assistente no CEDIS/IDP Privacy Lab. E-mail: gedeo.franca.adv@gmail.com.

41 Advogada. Mestranda em Direito e Inovação pela UFJF e bacharela em Direito pela Universidade Federal de Juiz de Fora - Campus Governador Valadares.. E-mail: izabellajbittencourt@gmail.com.

42 Advogada. Bacharela e Mestranda em Direito pela Universidade Federal da Bahia (UFBA). Pesquisadora na Associação Data Privacy Brasil de Pesquisa. Ex-presidente e Conselheira do Laboratório de Inovação e Direitos Digitais da UFBA (LabiD²). Email: mendonca.julia@outlook.com.

43 Advogada. Doutoranda e mestre em Direito pela UCP/Porto (bolsista FCT); especialista em Direito Digital pelo ITS-Rio e UERJ. Membro do Legalite PUC/RJ. Pesquisadora na Associação Data Privacy Brasil de Pesquisa. E-mail: paulaguedes94@hotmail.com.

as problemáticas envolvendo a utilização de RF para concessão da gratuidade de transporte público a partir de três pilares de discussão: (i) vieses e discriminação; (ii) o uso desnecessário de dinheiro público e a possibilidade de desengajamento com a tecnologia; (iii) o uso de dados biométricos e a urgência na adoção de mecanismos de prestação de contas (*accountability*). Questionamos se o uso da tecnologia para os fins pretendidos é, de fato, eficiente para solução ou mitigação do problema que se propõe resolver, com atenção especial na análise da questão sob à ótica da justiça de dados. Como conclusão, verifica-se a possibilidade de utilização de IA para fins de execução de políticas públicas, desde que seus efeitos positivos superem os negativos – o que não é identificado na hipótese do RF para verificação de identidade beneficiários de gratuidade no transporte público, considerando o interesse público em efetivar políticas públicas baseadas na justiça de dados.

Palavras-chave: Inteligência Artificial; Reconhecimento Facial; Justiça de Dados; Transporte Público; Direitos Fundamentais.

Abstract

The adoption of facial recognition (FR) technologies, based on Artificial Intelligence (AI) techniques, is already a reality in the public sector. It normally occurs for the implementation of public policies, as in the case of granting the right to free public transport. It turns out that this implementation can generate different implications from the perspective of data justice, ranging from risks to fundamental rights, discrimination against vulnerable groups, to the exclusion of access by part of the population to the right discussed. Thus, the present work intends to analyze the problems involving the use of FR to grant free public transport from three discussions pillars: (i) bias and discrimination; (ii) the unnecessary use of public money and the possibility of disengagement with technology; (iii) the use of biometric data and the urgency of adopting accountability mechanisms. The main question is whether the use of the technology for the intended purposes is, in fact, efficient in solving or mitigating the problem it is intended to solve, with special attention to analyzing the issue from the perspective of data justice. In conclusion, we defend the possibility of using AI for the purpose of implementing public policies, as long as its positive effects outweigh the negative ones – which is not identified in the FR hypothesis for verifying the identity of beneficiaries of free public transport, considering the public interest in implementing public policies based on data justice.

Keywords: Artificial Intelligence; Facial Recognition; Data Justice; Public Transportation; Fundamental Rights.

1. INTRODUÇÃO

O uso de tecnologias de reconhecimento facial (RF), baseados em técnicas de Inteligência Artificial (IA), já é uma realidade cada vez mais comum, para as mais diversas funcionalidades. No setor público, seu uso vai desde tentativas de destinação para fins comerciais, como o caso da Via 04 que foi barrado no metrô de São Paulo⁴⁴, até a implementação como pré-requisito para acesso “mais seguro” à plataformas governamentais, como o Gov.br⁴⁵. O acesso à políticas públicas também vem sendo condicionado à cessão e tratamento de dados pessoais, inclusive biométricos, como, por exemplo, na implementação desse tipo de tecnologia para a concessão de gratuidade de passagem no transporte público.

Essa dinâmica pode gerar diferentes riscos a direitos individuais e coletivos da população, como discriminações contra grupos minoritários e vulneráveis, especialmente devido a falhas em processos de identificação tecnológica, além das mais variadas violações de regras de proteção de dados pessoais. Todos esses elementos, somados à imposição da cessão e tratamento de dados, inclusive biométricos, para o acesso à políticas públicas podem gerar grandes implicações sob a ótica da justiça de dados, que serão abaixo analisadas.

É nessa toada que, inicialmente, pretendemos nortear a discussão sobre como está ocorrendo a adoção de reconhecimento facial, baseado em tecnologias de inteligência artificial, para a concessão da gratuidade de passagem por diferentes estados. Em seguida, passaremos a uma análise crítica acerca dos riscos e limites, adentrando inicialmente as discriminações oriundas das possíveis falhas desse tipo de tecnologia, junto à perspectiva da justiça de dados.

Também será analisada a euforia com a qual tais tecnologias são recebidas, sendo apresentadas como o único caminho possível para alcançar uma ideia de “progresso”, em nome de uma suposta eficiência do serviço público, percepção com grandes influências de tecnosolucionismo⁴⁶. Como último tópico, será analisada a coleta e tratamento dos dados pessoais dos beneficiários da gratuidade em questão, considerando que são dados sensíveis - detentores de uma maior proteção garantida pela Lei Geral de Proteção de Dados (LGPD) -, bem como instrumentos de prestação de contas, já que grande parte das implementações foram realizadas antes da entrada em vigor da LGPD.

44 IDEC. ViaQuatro é condenada por reconhecimento facial no Metrô de SP. Instituto Brasileiro de Defesa do Consumidor. Disponível em: <https://idec.org.br/idec-na-imprensa/viaquatro-e-condenada-por-reconhecimento-facial-sem-autorizacao-no-metro-de-sp>. Acesso em: 07 de mar. 2023.

45 GOV.BR. Saiba mais sobre os níveis da conta gov.br. Disponível em: <https://www.gov.br/governodigital/pt-br/conta-gov-br/saiba-mais-sobre-os-niveis-da-conta-govbr/saiba-mais-sobre-os-niveis-da-conta-govbr>. Acesso em: 08 de mar. 2023.

46 MOROZOV, Evgeny. To save everything, click here: the folly of technological solutionism. **J. Inf. Policy**, v. 4, n. 2014, p. 173-175, 2014.

Por fim, para alcançarmos a devida prestação de contas metodológica, esclarecemos que os dados utilizados neste artigo para confirmar o uso de reconhecimento facial para fins de concessão de gratuidade de acesso ao transporte público não foram alvo de pedidos de acesso à informação por meio da Lei de Acesso à Informação (LAI), mas vindos de fontes de dados públicas, como notícias de fontes jornalísticas confiáveis e outros documentos disponibilizados pelas concessionárias do serviço de transporte público.

2. O USO DE RECONHECIMENTO FACIAL PARA CONCESSÃO DE GRATUIDADE DE PASSAGEM

De acordo com a Constituição de 1988, o direito ao transporte é um direito social fundamental, garantido a todos os brasileiros. Em razão de condições de justificação prescrita por lei, como estado ou situação de vulnerabilidade, alguns grupos possuem o direito à gratuidade na concessão de passagem, como é o caso, por exemplo, da gratuidade dos transportes coletivos urbanos para os maiores de 65 anos, fruto do § 2º do art. 230 da Constituição Federal de 1988.

Para além da CF/88, outras legislações infraconstitucionais também garantem essa gratuidade, a exemplo da Lei nº 3.339/1999 (gratuidade de transporte aos estudantes das redes públicas municipal e federal no Estado do Rio de Janeiro) e da Lei 11.250/1992 (isenção à pessoas acometidas com certas enfermidades, como pacientes com câncer, Parkinson, HIV, Demência, Esquizofrenia, Doença de Alzheimer, Retardo Mental entre outras doenças do Município de São Paulo).

O reconhecimento facial, por sua vez, é uma tecnologia de identificação biométrica, baseado em técnicas de inteligência artificial, que tem sido amplamente utilizada em áreas como segurança⁴⁷, comércio eletrônico⁴⁸, saúde⁴⁹ e entretenimento⁵⁰. Nos últimos anos, cresceu também a adoção desse tipo de tecnologia como requisito para verificação de identidade para a concessão de benefícios e serviços, como a gratuidade de passagem

47 “C MERAS que serão usadas pela polícia têm reconhecimento facial e proteção de dados. **Extra Online**, 13 de dezembro de 2021. Disponível em: <https://extra.globo.com/economia-e-financas/servidor-publico/cameras-que-serao-usadas-pela-policia-tem-reconhecimento-facial-protexcao-de-dados-25317417.html>. Acesso em: 07 de mar. 2023.

48 UNICO. Digitalização do varejo: Entenda o papel do reconhecimento facial. **Unico**, 12 de dezembro de 2022. Disponível em: <https://unico.io/unicocheck/digitalizacao-no-varejo00/>. Acesso em: 07 de mar. 2023.

49 VASCONCELLOS, Hygino. Plano Hapvida é notificado por exigir biometria de crianças com autismo. **Uol**, 25 de janeiro de 2023. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2023/01/25/hapvida-e-notificada-por-exigir-biometria-a-autistas-abusivo-diz-idec.htm>. Acesso em: 07 de mar. 2023.

50 RECONHECIMENTO facial é usado no BBB 23 para seguir os participantes pela casa. TudoCelular.com, 20 de janeiro de 2023. Disponível em: <https://www.tudocelular.com/tech/noticias/n201286/reconhecimento-facial-usado-bbb-23-participantes.html>. Acesso em: 07 de mar. de 2023.

em transporte público.

Em uma perspectiva do ponto de vida das cidades, os serviços de transporte público são essenciais para a vida cotidiana da população, seja na mobilidade urbana, seja na geração de empregos. No ano de 2022, foram, em média, 33,4 milhões de viagens realizadas por dia no Brasil⁵¹. Desse total, 27,1% das viagens são beneficiadas com descontos ou isenções tarifárias, ou seja, 1 em cada 5 viagens são realizadas sem o pagamento de tarifas⁵².

No Brasil, a tecnologia de RF já foi implementada no transporte público das principais cidades e capitais, como São Paulo⁵³, Florianópolis⁵⁴, Salvador⁵⁵, Recife⁵⁶ e Brasília⁵⁷. Tal sistema de reconhecimento é utilizado para identificar os usuários do transporte público a partir da análise de seus rostos e, assim, verificar se esta pessoa identificada tem direito à gratuidade, de acordo com as regras estabelecidas pelas autoridades competentes. Por exemplo, idosos, pessoas com deficiência, estudantes e outros grupos específicos podem ter direito à gratuidade de acordo não apenas com a Constituição Federal, mas também com as leis e regulamentos locais específicos.

Na cidade de Belém no Estado do Pará, podemos encontrar nos serviços de transporte público o cartão “Passe Fácil”⁵⁸. Nos validadores dos ônibus haverá a implantação gradativa do sistema de Biometria Facial, onde será validado o rosto do beneficiário da gratuidade e em seguida liberada a catraca. É interessante notar que se o sistema identificar fraude no uso do cartão, ou seja, se a imagem do usuário não for a mesma do titular cadastrado, ocorrerá o bloqueio. Uma vez bloqueado, será cobrado o valor integral da passagem. Para desbloquear, o titular do cartão deverá se dirigir ao posto de atendimento Passe Fácil para regularizar a situação, de acordo com os critérios definidos

51 NTU. **Os grandes números da mobilidade urbana:** cenário nacional. Março, 2023. Disponível em: <https://www.ntu.org.br/novo/AreasInternas.aspx?idArea=7>. Acesso em: 07 de mar. 2023.

52 *Ibidem*.

53 “ÔNIBUS de São Carlos passam a contar com sistema de biometria facial nas catracas. G1, 23 de abril de 2018. Disponível em: <https://g1.globo.com/sp/sao-carlos-regiao/noticia/onibus-de-sao-carlos-passam-a-contar-com-sistema-de-biometria-facial-nas-catracas.ghtml>. Acesso em: 07 de mar. 2023.

54 RECONHECIMENTO facial começa a funcionar em terminais de ônibus. G1. Santa Catarina, 21 de julho de 2014. Disponível em: <http://g1.globo.com/sc/santa-catarina/noticia/2014/07/reconhecimento-facial-comeca-funcionar-em-terminais-de-onibus.html>. Acesso em: 07 de mar. 2023.

55 PELEGI, Alexandre. Biometria facial em Salvador bloqueou 11 mil cartões de gratuidade em 2018. **Diário do Transporte**, 27 de setembro de 2018. Disponível em: <https://diariodotransporte.com.br/2018/09/27/biometria-facial-em-salvador-bloqueou-11-mil-cartoes-de-gratuidade-em-2018/>. Acesso em: 07 de mar. 2023

56 SOARES, Roberta. Ônibus da Região Metropolitana do Recife já estão com biometria facial. JC, 25 de abril de 2019. Disponível em: <https://jc.ne10.uol.com.br/colunas/mobilidade/2019/04/25/onibus-da-regiao-metropolitana-do-recife-ja-estao-com-biometria-facial>. Acesso em: 07 de mar. 2023.

57 BIOMETRIA FACIAL se torna obrigatória em todos os ônibus do DF. G1, Distrito Federal, 02 de maio de 2018. Disponível em: <https://g1.globo.com/df/distrito-federal/noticia/biometria-facial-se-torna-obrigatoria-em-todos-os-onibus-do-df.html>. Acesso em: 07 de mar. 2023.

58 SETRANS-BEL. Informativo Biometria Facial. Passe Fácil. Disponível em: https://www.passefacil.com.br/Content/ImgProject/INFORMATIVO_BIOMETRIA_FACIAL.pdf. Acesso em: 13 de jul 2023.

na legislação municipal.

Dessa maneira, a utilização de tais tecnologias para a concessão de benefícios de gratuidade pode ter diversos impactos positivos para a sociedade, principalmente em razão da importância da prestação desse serviço para a mobilidade urbana. Prevenção à fraude, segurança, eficiência e melhoria na prestação do serviço público com o avanço tecnológico e utilização da tecnologia na execução de políticas públicas são alguns exemplos dos benefícios que podem ser aferidos. Por outro lado, o que se pretende aprofundar no presente texto, é a existência de riscos e limites a serem observados na utilização desse tipo de tecnologia, os quais serão analisados a seguir sob a ótica da justiça de dados.

3. OS RISCOS E LIMITES DO USO DE RECONHECIMENTO FACIAL SOB A ÓTICA DA JUSTIÇA DE DADOS

O aumento da implementação de RF na verificação da identidade de pessoas beneficiárias de gratuidade no transporte público faz parte de um processo mais amplo de datificação da atual sociedade, no qual o Estado também se insere. Nesse contexto, são coletados e tratados dados de indivíduos, grupos e conjuntos populacionais inteiros, sem que haja sequer conhecimento, e muito menos controle sobre esses fluxos, que ainda incluem atores do setor público e privado. Em verdade, muitos países vem passando por um movimento de plataformação dos serviços públicos⁵⁹, que culminam também na datificação de políticas públicas.

Com base nestes dados e suas inferências, são realizados processos de perfilização e tomada de decisão sobre a vida das pessoas, gerando significativos impactos em processos sociais⁶⁰, ajudando a construir uma sociedade que não é apenas informada por dados (*data-informed*), mas orientada por eles, ou seja, *data-driven*.⁶¹

Um exemplo que pode ser mencionado no contexto brasileiro é a plataforma gov.br, estabelecida pelo Decreto nº 10.332, de 28 de abril de 2020, a qual possui como objetivo reunir digitalmente a totalidade dos serviços oferecidos pela União. Para ter acesso a determinadas políticas públicas, por exemplo, como “abono salarial” ou “auxílio

59 BIONI, Bruno; GARROTE, Marina; MEIRA, Marina; PASCHOALINI, Nathan. Entre a visibilidade e a exclusão: um mapeamento dos riscos da Identificação Civil Nacional e do uso de sua base de dados para a plataforma gov.br. **Associação Data Privacy Brasil de Pesquisa**, 2022. Disponível em: <https://www.dataprivacybr.org/wp-content/uploads/2022/06/identidade-digital-vf2-1.pdf>.

60 TAYLOR, Linnet. **What is data justice?** The case for connecting digital rights and freedoms globally. *Big Data & Society*, July–December 2017.

61 KITCHIN R. The ethics of smart cities and urban science. **Philosophical Transactions of the Royal Society A** 374(2083): 2016, 1–15.

gás”, é necessário realizar um processo de identificação no site dividido em três níveis de autenticação: bronze, prata e ouro. Em resumo, uma das formas de alcançar uma conta “ouro” e ter acesso a mais funcionalidades, é através da habilitação do uso de reconhecimento facial. Essa lógica é bastante semelhante ao objeto de estudo do presente texto, em que o acesso à políticas públicas está condicionado à cessão de cada vez mais dados da população, inclusive sensíveis.

A justificativa atrelada à tais fenômenos sempre estão associados a garantia de uma maior eficiência, seguindo uma lógica com nítidos requintes de tecnosolucionismo,⁶² especialmente a partir de uma justificativa de se evitar fraudes e de se garantir uma maior efetividade no cumprimento das políticas públicas. A questão é que apesar dos aparentes benefícios gerados a adoção desses tipos de tecnologias é sempre acompanhada de riscos e problemas já existentes para direitos fundamentais que precisam ser observados.

É como se o potencial das tecnologias influenciarem no contexto econômico estivesse crescendo vertiginosamente⁶³ enquanto os mecanismos para redução e consciência acerca das discriminações, riscos e danos oriundos do seu uso não estivessem acompanhando o mesmo ritmo⁶⁴.

Sendo assim, a partir de circunstâncias com panoramas parecidos, Linet Taylor propôs a construção de um conceito de justiça de dados, ou seja, de uma lógica partindo da “justeza na forma como as pessoas são visibilizadas, representadas e tratadas como resultado da produção de dados digitais”⁶⁵. Para ela, assim como a ideia de “justiça” é fator condicionante para o estabelecimento do Estado de Direito, a justiça de dados é igualmente importante para estabelecer limites combinativos à *injustiças* de dados encabeçadas por essa dinâmica constante de coleta, agrupamento e perfilização de dados da população.

Tal conceito foi enquadrado pela autora através de três camadas de análise: (i) visibilidade; (ii) (des)engajamento com a tecnologia; e (iii) combate à discriminação orientada por dados. Em resumo, o primeiro pilar sobre “visibilidade”, especialmente atrelado à privacidade informacional, tem bastante relação com a análise de como os indivíduos são representados a partir dos dados coletados, envolvendo diferentes discussões, como a sobre perfilamento coletivo. Por sua vez, o segundo pilar envolve um contrafluxo à lógica de associar a adoção de tecnologias a uma coisa positiva. No sentido de escolher a possibilidade de desengajar com determinada tecnologia, ou

62 MOROZOV, Evgeny. To save everything, click here: the folly of technological solutionism. **J. Inf. Policy**, v. 4, n. 2014, p. 173-175, 2014.

63 TAYLOR, Linnet. **What is data justice?** The case for connecting digital rights and freedoms globally. *Big Data & Society*, July–December 2017.

64 *Ibidem*.

65 TAYLOR, Linnet. **What is data justice?** The case for connecting digital rights and freedoms globally. *Big Data & Society*, July–December 2017.

mesmo diretamente escolher não utilizá-la, de forma a possibilitar um maior controle sobre os termos do seu engajamento com o mercado de dados, principalmente evitando se tornar “parte de bases de dados comerciais como subproduto de intervenções pró desenvolvimento”.⁶⁶

Por fim, o terceiro pilar trazido pela autora na formulação do seu enquadramento conceitual é o de não-discriminação, que se constitui por duas dimensões: (i) o poder de identificar e desafiar os vieses na utilização de dados, e (ii) a liberdade de não ser discriminado. Esse será um dos pontos principais a serem abordados no presente texto. Isso porque, o uso de tecnologias baseadas em inteligência artificial - como as de reconhecimento facial- são marcadas pela produção de resultados discriminatórios, principalmente contra grupos minoritários e vulneráveis, e violações de regras de proteção de dados pessoais, fatores que costumam contribuir para a violações de outros direitos fundamentais. Isso ocorre principalmente em razão de falhas nos processos de identificação tecnológica, as quais podem culminar em exclusão de cidadãos do acesso a serviços essenciais oferecidos pelo Estado.

Desse modo, conforme será abordado em tópico específico, nos casos em que o caminho do mencionado (des)engajamento com a tecnologia não seja o escolhido, a sua adoção deve vir, no mínimo, acompanhadas por avaliação de riscos, políticas claras de privacidade e proteção de dados pessoais, para garantir que os direitos e liberdades individuais e coletivos sejam respeitados – o que nem sempre é verificado na prática. Diante dos pontos abordados, a discussão aqui introduzida será aprofundada a partir de três recortes que possuem em seu cerne interlocuções intrínsecas com as noções sobre justiça de dados, quais sejam, vieses e discriminação; o uso desnecessário de dinheiro público e a possibilidade de desengajamento com a tecnologia; o uso de dados biométricos e a urgência na adoção de mecanismos de prestação de contas.

3.1. O uso desnecessário de dinheiro público e a possibilidade de desengajamento com a tecnologia

Uma primeira questão que deve ser considerada, é como novas tecnologias são recebidas com bastante euforia⁶⁷, partindo de uma expectativa geral de que “o futuro sempre será melhor”⁶⁸. Tal perspectiva acaba gerando uma atuação governamental voltada

66 TAYLOR, L. BROEDERS, D. **In the name of development: Power, profit and the datafication of the global South**, 2015. *Geoforum* 64(4): 229–237.

67 MENDONÇA, Júlia; RIELLI, Mariana. Dados, regulação e fronteiras do metaverso: a urgência por parcimônia. **JOTA**, 10 de outubro de 2022. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/dados-regulacao-e-fronteiras-do-metaverso-a-urgencia-por-parcimoniam-10102022>. Acesso em: 07 de mar. 2023.

68 BRANDÃO, Luiza. **Solucionismo tecnológico e o perigo de uma história única**. Instituto de Referência em Internet e Sociedade, 2021. Disponível em: <https://irisbh.com.br/solucionismo-tecnologico-e-o-perigo-de-uma-historia-unica/>. Acesso em: 07 de mar. 2023.

à implementação de tais tecnologias a partir de robustos investimentos⁶⁹, que muitas vezes ganham uma significativa aceitação social quando são utilizadas, por exemplo, para fins de segurança pública⁷⁰.

Com relação ao uso de RF para concessão de gratuidade de passagem, o entusiasmo também é comum, no entanto, mais voltado sob a justificativa de gerar “economia” aos cofres públicos e combate à fraudes. Um caso que pode ser mencionado é o da cidade de Salvador em que, segundo dados da Integra, o uso de passes por terceiros que não são os reais beneficiários já ultrapassou 14%⁷¹ do total de passageiros, chegando a gerar um prejuízo semestral de trinta e nove milhões de reais⁷². Desse modo, a implementação do referido tipo de tecnologia é posta como uma proposta de solução universal, sob um pretexto de garantir uma melhor eficiência e “melhor” detecção de fraudes, reduzindo as perdas financeiras.

A concepção de que a prioridade de direcionamento de dinheiro público, em um contexto que o transporte público existe em condições precárias⁷³, deve ser nesse tipo de tecnologia, em nosso sentir, por si só já é bastante problemático. Quando analisa-se que, conforme ora destacado, as mesmas tecnologias possuem tamanhos índices de erros (falsos positivos e negativos) e riscos a direitos fundamentais, dificilmente a priorização estatal poderá ser justificada em nome de uma suposta eficiência.

Sendo assim, é nessa dinâmica que a dimensão de possibilidade de (des)engajamento com a tecnologia de Linet Taylor aparece. Por que receber com tamanho entusiasmo a implementação de novas tecnologias para garantir uma suposta eficiência, sendo que a tecnologia proposta em si mesma não é eficiente e gera tantos riscos à direitos fundamentais?

As escolhas pautadas em uma lógica de tecnosolucionismo⁷⁴ podem acabar fomentando

69 RECONHECIMENTO facial deve movimentar R\$50 bilhões em 2022. **O progresso digital**, 11 de novembro de 2021. Disponível em: <https://www.progresso.com.br/tecnologia/reconhecimento-facial-deve-movimentar-r-50-bilhoes-em-2022/385360/>. Acesso em: 07 de mar. 2023.

70 FALCÃO, Cíntia. Lentes racistas: Rui costa está transformando a Bahia em um laboratório de vigilância com reconhecimento facial. **The Intercept**, 20 de setembro de 2021. Disponível em: <https://theintercept.com/2021/09/20/rui-costa-esta-transformando-a-bahia-em-um-laboratorio-de-vigilancia-com-reconhecimento-facial/>. Acesso em: 07 de mar. 2023.

71 VIGNÉ, Júlia. 11 mil passageiros perdem direito ao SalvadorCard por fraude este ano. **Correio da Bahia**, 26 de setembro de 2018. Disponível em: <https://www.correio24horas.com.br/noticia/nid/11-mil-passageiros-perdem-direito-ao-salvadorcard-por-fraude-este-ano/>. Acesso em: 07 de mar. 2023.

72 VIGNÉ, Júlia. Fraudes em meia passagem e gratuidade causam prejuízo de R\$39 milhões. **Correio da Bahia**. 26 de setembro de 2018. Disponível em: <https://www.correio24horas.com.br/noticia/nid/fraudes-em-meia-passagem-e-gratuidade-causam-prejuizo-de-r-39-milhoes/>. Acesso em: 07 de mar. 2023.

73 OVIEDO, Annie. A crise do transporte público no Brasil tem solução?. **Nexo Jornal**, 20 de setembro de 2022. Disponível em: <https://pp.nexojournal.com.br/ponto-de-vista/2022/A-crise-do-transporte-p%C3%BAblico-no-Brasil-tem-solu%C3%A7%C3%A3o>. Acesso em: 07 de mar. 2023.

74 MOROZOV, Evgeny. **To save everything, click here: the folly of technological solutionism**. J. Inf. Policy, v. 4, n. 2014, p. 173-175, 2014.

a adoção a adoção irrefletida de tecnologias pelo setor governamental, a partir de um conto da história única⁷⁵, como se esse fosse o único caminho viável para a ideia de progresso. Essa análise se complexifica quando observamos que essa tecnologia em questão ainda fomenta a coleta de dados biométricos, os quais são classificados como sensíveis, multiplicando os riscos envolvidos. Isso porque, o transporte público é feito em regime de concessão⁷⁶ e pode ser encabeçado por diferentes empresas do setor privado, o que nos faz levantar outras reflexões preliminares, ainda sem conclusão, que serão pontos de partida para trabalhos futuros: (i) quantas empresas vão ter acesso a essa base com dados biométricos da população?; (ii) como é feita a transferência de tais informações entre as empresas?; (iii) as empresas possuem e mantém o mesmo nível de proteção com relação aos dados pessoais?

3.2. Viéses e Discriminação

No atual estado de desenvolvimento da tecnologia, sistemas de RF apresentam performance diferenciada de acordo com diferentes critérios, como raça, gênero, classe social e território. Estudos comprovam que as taxas de erros na identificação (falsos positivos) ou não identificação de rostos (falsos negativos) são mais altas quando a tecnologia é aplicada em indivíduos do gênero feminino e de pele negra⁷⁷, o que é intensificado quando estes dois critérios se sobrepõem.

Esta constatação científica de maior taxa de erro na identificação de raça e gênero, especialmente quando esses critérios são combinados, podem ser compreendidos da construção tecnológica a partir de um banco de dados masculino e branco, com pouquíssima representação de grupos não hegemônicos, perpassando estruturas patriarcais, machistas, binárias e racistas.⁷⁸ Nessa linha, por exemplo, pessoas transexuais e não binárias são fortemente impactadas por essas tecnologias, já que seus corpos, por não se encaixarem no modelo dito “universal” de divisão entre homem e mulher, não são sequer reconhecidos pela tecnologia.⁷⁹

75 BRANDÃO, Luiza. **Solucionismo tecnológico e o perigo de uma história única**. Instituto de Referência em Internet e Sociedade, 2021. Disponível em: <https://irisbh.com.br/solucionismo-tecnologico-e-o-perigo-de-uma-historia-unica/>. Acesso em: 07 de mar. 2023.

76 PREFEITURA DE BELO HORIZONTE. **Minuta de contrato de concessão do serviço público de transporte coletivo de passageiros por ônibus**. 2008. Disponível em: https://prefeitura.pbh.gov.br/sites/default/files/imagens/authenticated%2C%20editor_a_bhtrans/MINUTA_CONTRATO_FINAL_ANEXO_IX_080326.pdf. Acesso em: 27 jun. 2023.

77 BUOLAMWINI, Joy; GEBRU, Timnit. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. **Conference on Fairness, Accountability and Transparency. Proceedings of Machine Learning Research**, Vol. 81, pp. 1-15, 2018. Disponível em: <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>.

78 SILVA, Paula Guedes Fernandes da. É menino! É menina! Os riscos das tecnologias de análise facial para as identidades de gênero trans e não-binárias. **Revista Direito, Estado e Sociedade**, Ed. 60, 2022. DOI: <https://doi.org/10.17808/des.0.1875>.

79 COSTANZA-CHOCK, Sasha. Design Justice, A.I., and escape from the matrix of domination. Cambridge: **Journal of Design and Science**, Jul. 2018. DOI:10.21428/96c8d426. Disponível em: <https://jods.mitpress.mit.edu/pub/costanza-chock/release/4>.

No caso específico de RF aplicado para concessão de gratuidade de transporte público, os falsos negativos – isto é, o erro de não identificação de determinado rosto – pode fazer com que um cidadão que preencha os requisitos legais não tenha acesso ao serviço público e ao respectivo benefício em decorrência de falha em sua identificação. Nesse ponto, citamos o caso ocorrido no Distrito Federal em que o Transporte Público do DF (DFTrans) bloqueou o cartão de gratuidade de transporte de uma estudante após ela assumir seus cachos. Tal bloqueio foi apenas corrigido após a jovem recorrer à Defensoria Pública, já que a correção envolve processos burocráticos.⁸⁰

Esse caso exemplifica que, mesmo em casos simples que não envolvem aspectos sensíveis, a tecnologia já apresenta falhas, ganhando camadas de sensibilidade quando características sensíveis, tais como gênero e raça, são somadas na avaliação dos rostos pela tecnologia. Isso é perceptível em outro caso também ocorrido no Distrito Federal, em que uma estudante negra e trans teve seu passe bloqueado após a implementação do sistema de reconhecimento nos ônibus da região⁸¹ – mesmo após recorrer pela suspensão do bloqueio, a estudante continuou sem acesso ao benefício que, além de seu direito por lei, a afastava do exercício de outro direito fundamental, o acesso à saúde.⁸²

Esses equívocos, por vezes acompanhados de resultados discriminatórios e consequentes situações que podem ser enxergadas como *injustiças de dados*, apesar de poderem ocorrer por diferentes razões, inclusive como consequência de baixa qualidade das imagens, ângulos e condições precárias de luminosidade. No entanto, um dos principais fatores é o desequilíbrio de representação na forma como as bases de dados são utilizadas pelos sistemas, tanto no momento de treinamento da IA, como no momento da sua implementação, os quais sempre tendem a partir de um referencial branco, masculino, cisnormativo e europeu.⁸³

Nesse ponto, resta evidente que as injustiças de dados que podem ser ocasionadas por esse tipo de tecnologia perpassam de maneira bastante significativa pelas questões atreladas à discriminação nas camadas propostas por Linet Taylor. Isso foi demonstrado com os exemplos acima descritos, em que os vieses oriundos da utilização de dados biométricos para o acesso a políticas públicas e a direitos da população possuem o grande potencial de gerar situações discriminatórias. De igual maneira, tais casos também ilustram os impactos no que tange à exclusão de parte da população - que já estão em um contexto de vulnerabilidade social- que essa utilização pode ocasionar.

80 TEIXEIRA, Isadora. Biometria facial nos ônibus não reconhece mudança visual de alunos. **Metrópoles**, 12 de maio de 2018. Disponível em: <https://www.metrolopes.com/distrito-federal/transporte-df/biometria-facial-nos-onibus-nao-reconhece-mudanca-visual-de-alunos>. Acesso em 05 de julho de 2023.

81 TV BRASIL. Biometria facial em ônibus do DF apresenta problemas. YouTube, 18 de maio de 2018. Disponível em: https://www.youtube.com/watch?v=0JYSDWtL0Jo&ab_channel=TVBrasil. Acesso em: 08 de mar. de 2023.

82 VARON, Joana; SILVA, Mariah Rafaela. **Reconhecimento facial no setor público e identidades trans:** tecnopolíticas de controle e ameaça à diversidade de gênero em suas interseccionalidades de raça, classe e território. Disponível em: <https://codingrights.org/docs/rec-facial-id-trans.pdf>. Acesso em: 08 de mar. 2023.

83 Ibidem.

Sendo assim, para o combate a tais injustiças de dados é importante uma análise crítica previamente à adoção de tecnologias de RF em contextos condicionais para o acesso às políticas públicas. Inicialmente, sempre tendo em vista a possibilidade de que a não adoção pode ser a melhor escolha, ou seja, (des)engajar com a tecnologia pode ser uma possibilidade mais adequada em determinados contextos, considerando que garantir a não-discriminação pode ser mais importante e necessário do que a adoção irrefletida de uma tecnologia que ainda não está suficientemente madura, em nome de uma suposta eficiência.

3.3. O uso de dados biométricos e a urgência na adoção de mecanismos de prestação de contas

Nos casos em que, mesmo ponderando riscos e benefícios, bem como as possibilidades de gerar injustiças de dados, ainda assim, seja escolhido o caminho de adotar essa tecnologia, é urgente que a mesma seja acompanhada de mecanismos de prestação de contas. Como já mencionado, dados biométricos são classificados como sensíveis pela LGPD, o que demanda que seu tratamento, por si só, respeite regras mais específicas de proteção, em razão de sua potencialidade em levar a resultados discriminatórios.

Desta forma, por exemplo, além da obrigatoriedade de cumprimento dos princípios e do enquadramento legal em uma das bases legais do art. 11, é necessário que os agentes de tratamento adotem determinadas salvaguardas e medidas que previnem e mitiguem os riscos provenientes da utilização desses dados. A LGPD impõe algumas obrigações, como deveres de transparência e segurança, além de indicar a realização de um Relatório de Impacto à Proteção de Dados (RIPD).⁸⁴

A concessão de gratuidade de passagem se trata de uma execução de política pública pela administração pública em razão da existência de dois aspectos materiais⁸⁵: (i) ato formal que institui a política (leis⁸⁶ e regulamentos); e (ii) definição de um programa ou ação governamental específico a executado por uma entidade ou por um órgão público⁸⁷. Assim, a hipótese de tratamento de dados para fins de execução de política

84 A LGPD não define as hipóteses de obrigatoriedade de realização de RIPD, o que será futuramente regulamentado pela Autoridade Nacional de Proteção de Dados (ANPD), de acordo com sua competência estabelecida no art. 55-J, inciso XIII. Espera-se que esta exigência ocorra para os casos em que o tratamento representar alto risco à garantia dos princípios gerais de proteção de dados pessoais, o que provavelmente incluirá o tratamento de dados pessoais sensíveis.

85 AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS. **Tratamento de dados pessoais pelo poder público**. Brasília: ANPD, 2023. Disponível em: <https://www.gov.br/anpd/pt-br/documentos-e-publicacoes/documentos-de-publicacoes/guia-poder-publico-anpd-versao-final.pdf>. Acesso em: 06 de jun. 2023.

86 Como exemplo de lei que institui a possibilidade de gratuidade de transporte podemos citar o Estatuto do Idoso (Lei 10.741/2003), que prevê em seu art. 39 que maiores de 65 anos possuem o direito à gratuidade para utilização de transportes coletivos.

87 O Estado de Pernambuco, por exemplo, institui o Programa VEM Passe Livre, o qual concede passagens gratuitas aos estudantes da rede pública estadual e aos cotistas da Universidade de Pernambuco que estejam regularmente matriculados e com a frequência comprovada. Disponível em: <https://www.granderecife.pe.gov.br/vem-passe-livre/>. Acesso em: 08 de mar. 2023.

pública, conforme o art. 23 da LGPD, ganha uma camada de obrigações, uma vez que apenas poderá ser realizado para o atendimento da finalidade pública, na persecução do interesse público, com o objetivo de executar as competências legais ou cumprir as atribuições legais do serviço público.

O problema é que, nos atuais casos de tratamento de dados biométricos por RF para verificação da identidade dos beneficiários da gratuidade, existe a urgência de se investigar se esses processos foram e continuam sendo implementados com observância das regras e salvaguardas estabelecidos na LGPD, sobretudo ao princípio da prestação de contas, tendo em vista o fato de que grande parte dessas tecnologias foram implementadas antes da entrada em vigor da lei. A não observância das obrigações legais abre margem para a violação da proteção de dados dos beneficiários e, conseqüentemente, outros direitos fundamentais por meio, por exemplo, de exclusão de acesso em decorrência de falha na identificação, inclusive com resultados discriminatórios, como já exposto anteriormente.

4. CONSIDERAÇÕES FINAIS

Logo, face à realidade de utilização de IA por meio de sistemas de reconhecimento facial para verificação de identidade em transportes públicos para fins de concessão de gratuidade de passagem, o objetivo do artigo centra-se, de acordo com a metodologia hermenêutico-doutrinal de fontes jurídicas brasileiras e internacionais, na avaliação de seus riscos, limites a partir da noção de justiça de dados.

O uso de tecnologias de reconhecimento facial no transporte público já está sendo amplamente utilizado em diversas cidades do Brasil, algumas tendo sido implementadas antes mesmo da entrada em vigor da LGPD. Diversas cidades utilizam esses sistemas para conceder gratuidade de passagem para diversos grupos da sociedade. No entanto, a utilização de RF tem gerado diversos efeitos, sob a justificativa de eficiência e combate à fraude.

A partir de uma sociedade orientada por dados, são gerados processos de perfilização e de tomada de decisões sobre a vida das pessoas, sendo que o uso desta tecnologia pode resultar em erros de identificação com impactos desproporcionais em grupos minoritários e vulneráveis, que, nesses casos, são excluídos do acesso ao benefício a que têm direito, para além dos problemas no tratamento de dados pessoais e gasto desnecessário de dinheiro público.

Foram apresentados três recortes, dentro da ótica da justiça de dados, que se relacionam com as problemáticas trazidas pelo uso da tecnologia de reconhecimento facial no setor público, sendo eles: (i) vieses e discriminação; (ii) o uso desnecessário de dinheiro público e a possibilidade de desengajamento com a tecnologia; (iii) o uso de dados biométricos e a urgência na adoção de mecanismos de prestação de contas.

Nesse sentido, questionamos se o uso da tecnologia para os fins pretendidos é, de fato, eficiente para solução ou mitigação do problema que se propõe resolver, com foco especial na análise da problemática sob à luz da justiça de dados. A partir disso, como conclusão, defendemos a possibilidade de utilização de IA para fins de execução de políticas públicas, desde que seus efeitos positivos superem os negativos. Todavia, diante dos pontos ressaltados no presente artigo, tal pressuposto não é identificado na hipótese do reconhecimento facial para verificação de identidade de pessoas beneficiárias de gratuidade no transporte público, considerando o interesse público em efetivar políticas públicas sempre pautadas na ideia de justiça de dados.

5. Referências

AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS. **Tratamento de dados pessoais pelo poder público**. Brasília: ANPD, 2023. Disponível em: <https://www.gov.br/anpd/pt-br/documentos-e-publicacoes/documentos-de-publicacoes/guia-poder-publico-anpd-versao-final.pdf>. Acesso em: 06 de jun. 2023.

Biometria facial se torna obrigatória em todos os ônibus do DF. **G1**, Distrito Federal, 02 de maio de 2018. Disponível em: <https://g1.globo.com/df/distrito-federal/noticia/biometria-facial-se-torna-obrigatoria-em-todos-os-ônibus-do-df.html>. Acesso em: 07 de mar. 2023.

BIONI, Bruno; GARROTE, Marina; MEIRA, Marina; PASCHOALINI, Nathan. Entre a visibilidade e a exclusão: um mapeamento dos riscos da Identificação Civil Nacional e do uso de sua base de dados para a plataforma gov.br. **Associação Data Privacy Brasil de Pesquisa**, 2022. Disponível em: <https://www.dataprivacybr.org/wp-content/uploads/2022/06/identidade-digital-vf2-1.pdf>.

BRANDÃO, Luiza. **Solucionismo tecnológico e o perigo de uma história única**. Instituto de Referência em Internet e Sociedade, 2021. Disponível em: <https://irisbh.com.br/solucionismo-tecnologico-e-o-perigo-de-uma-historia-unica/>. Acesso em: 07 de mar. 2023.

BUOLAMWINI, Joy; GEBRU, Timnit. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. **Conference on Fairness, Accountability and Transparency. Proceedings of Machine Learning Research**, Vol. 81, pp. 1-15, 2018. Disponível em: <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>.

Câmeras que serão usadas pela polícia têm reconhecimento facial e proteção de dados. **Extra Online**, 13 de dezembro de 2021, Disponível em: <https://extra.globo.com/economia-e-financas/servidor-publico/cameras-que-serao-usadas-pela-policia-tem-reconhecimento-facial-protacao-de-dados-25317417.html>. Acesso em: 07 de mar. 2023

COSTANZA-CHOCK, Sasha. Design Justice, A.I., and escape from the matrix of domination. Cambridge: **Journal of Design and Science**, Jul. 2018. DOI:10.21428/96c8d426. Disponível em: <https://jods.mitpress.mit.edu/pub/costanza-chock/release/4>.

FALCÃO, Cíntia. Lentes racistas: Rui costa está transformando a Bahia em um laboratório de vigilância com reconhecimento facial. **The Intercept**, 20 de setembro de 2021. Disponível em: <https://theintercept.com/2021/09/20/rui-costa-esta-transformando-a-bahia-em-um-laboratorio-de-vigilancia-com-reconhecimento-facial/>. Acesso em: 07 de mar. 2023.

GOV.BR. Saiba mais sobre os níveis da conta gov.br. Disponível em: <https://www.gov.br/governodigital/pt-br/conta-gov-br/saiba-mais-sobre-os-niveis-da-conta-govbr/saiba-mais-sobre-os-niveis-da-conta-govbr>. Acesso em: 08 de mar. 2023.

IDEC. ViaQuatro é condenada por reconhecimento facial no Metrô de SP. Instituto Brasileiro de Defesa do Consumidor. Disponível em: <https://idec.org.br/idec-na-imprensa/viaquatro-e-condenada-por-reconhecimento-facial-sem-autorizacao-no-metro-de-sp>. Acesso em: 07 de mar. 2023.

KITCHIN R. The ethics of smart cities and urban science. **Philosophical Transactions of the Royal Society A** 374(2083): 2016, 1–15.

MENDONÇA, Júlia; RIELLI, Mariana. Dados, regulação e fronteiras do metaverso: a urgência por parcimônia. **JOTA**, 10 de outubro de 2022. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/dados-regulacao-e-fronteiras-do-metaverso-a-urgencia-por-parcimonia-10102022>. Acesso em: 07 de mar. 2023.

MOROZOV, Evgeny. To save everything, click here: the folly of technological solutionism. **J. Inf. Policy**, v. 4, n. 2014, p. 173-175, 2014.

NTU. **Os grandes números da mobilidade urbana**: cenário nacional. Março, 2023. Disponível em: <https://www.ntu.org.br/novo/AreasInternas.aspx?idArea=7>. Acesso em: 07 de mar. 2023.

ÔNIBUS de São Carlos passam a contar com sistema de biometria facial nas catracas. **G1**, 23 de abril de 2018. Disponível em: <https://g1.globo.com/sp/sao-carlos-regiao/noticia/onibus-de-sao-carlos-passam-a-contar-com-sistema-de-biometria-facial-nas-catracas.ghtml>. Acesso em: 07 de mar. 2023.

OVIEDO, Annie. A crise do transporte público no Brasil tem solução?. **Nexo Jornal**, 20 de setembro de 2022. Disponível em: <https://pp.nexojornal.com.br/ponto-de-vista/2022/A-crise-do-transporte-p%C3%BAblico-no-Brasil-tem-solu%C3%A7%C3%A3o>. Acesso em: 07 de mar. 2023.

PELEGI, Alexandre. Biometria facial em Salvador bloqueou 11 mil cartões de gratuidade em 2018. **Diário do Transporte**, 27 de setembro de 2018. Disponível em: <https://diariodotransporte.com.br/2018/09/27/biometria-facial-em-salvador-bloqueou-11-mil-cartoes-de-gratuidade-em-2018/>. Acesso em: 07 de mar. 2023.

PREFEITURA DE BELO HORIZONTE. **Minuta de contrato de concessão do serviço público de transporte coletivo de passageiros por ônibus**. 2008. Disponível em: https://prefeitura.pbh.gov.br/sites/default/files/imagens/authenticated%2C%20editor_a_bhtrans/MINUTA_CONTRATO_FINAL_ANEXO_IX_080326.pdf. Acesso em: 27 jun. 2023.

RECONHECIMENTO facial começa a funcionar em terminais de ônibus. **G1**. Santa Catarina, 21 de julho de 2014. Disponível em: <http://g1.globo.com/sc/santa-catarina/noticia/2014/07/reconhecimento-facial-comeca-funcionar-em-terminais-de-onibus.html>. Acesso em: 07 de mar. 2023.

RECONHECIMENTO facial deve movimentar R\$50 bilhões em 2022. **O progresso digital**, 11 de novembro de 2021. Disponível em: <https://www.progresso.com.br/tecnologia/reconhecimento-facial-deve-movimentar-r-50-bilhoes-em-2022/385360/>. Acesso em: 07 de mar. 2023.

RECONHECIMENTO facial é usado no BBB 23 para seguir os participantes pela casa. **TudoCelular.com**, 20 de janeiro de 2023. Disponível em: <https://www.tudocelular.com/tech/noticias/n201286/reconhecimento-facial-usado-bbb-23-participantes.html>. Acesso em: 07 de mar. de 2023.

SETRANS-BEL. Informativo Biometria Facial. **Passe Fácil**. Disponível em: https://www.passefacil.com.br/Content/ImgProject/INFORMATIVO_BIOMETRIA_FACIAL.pdf. Acesso em: 13 de jul 2023.

SILVA, Paula Guedes Fernandes da. É menino! É menina! Os riscos das tecnologias de análise facial para as identidades de gênero trans e não-binárias. **Revista Direito, Estado e Sociedade**, Ed. 60, 2022. DOI: <https://doi.org/10.17808/des.0.1875>.

SOARES, Roberta. Ônibus da Região Metropolitana do Recife já estão com biometria facial. **JC**, 25 de abril de 2019. Disponível em: <https://jc.ne10.uol.com.br/colunas/mobilidade/2019/04/25/onibus-da-regiao-metropolitana-do-recife-ja-estao-com-biometria-facial>. Acesso em: 07 de mar. 2023.

TAYLOR, L. BROEDERS, D. **In the name of development: Power, profit and the datafication of the global South**, 2015. *Geoforum* 64(4): 229–237

TAYLOR, Linnet. **What is data justice?** The case for connecting digital rights and freedoms globally. *Big Data & Society*, July–December 2017.

TEIXEIRA, Isadora. Biometria facial nos ônibus não reconhece mudança visual de alunos. *Metrópoles*, publicado em 12 de maio de 2018. Disponível em: <https://www.metropoles.com/distrito-federal/transporte-df/biometria-facial-nos-onibus-nao-reconhece-mudanca-visual-de-alunos>. Acesso em 05 de julho de 2023.

TV BRASIL. Biometria facial em ônibus do DF apresenta problemas. YouTube, 18 de maio de 2018. Disponível em: https://www.youtube.com/watch?v=0JYSDWtL0Jo&ab_channel=TVBrasil. Acesso em: 08 de mar. de 2023.

UNICO. Digitalização do varejo: Entenda o papel do reconhecimento facial. *Unico*, 12 de dezembro de 2022. Disponível em: <https://unico.io/unicocheck/digitalizacao-no-varejo00/>. Acesso em: 07 de mar. 2023.

VARON, Joana; SILVA, Mariah Rafaela. **Reconhecimento facial no setor público e identidades trans:** tecnopolíticas de controle e ameaça à diversidade de gênero em suas interseccionalidades de raça, classe e território. Disponível em: <https://codingrights.org/docs/rec-facial-id-trans.pdf>. Acesso em: 08 de mar. 2023.

VASCONCELLOS, Hygino. Plano Hapvida é notificado por exigir biometria de crianças com autismo. *Balneário Camboriu*, 25 de janeiro de 2023. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2023/01/25/hapvida-e-notificada-por-exigir-biometria-a-autistas-abusivo-diz-idec.htm>. Acesso em: 07 de mar. 2023.

VIGNÉ, Júlia. 11 mil passageiros perdem direito ao SalvadorCard por fraude este ano. *Correio da Bahia*, 26 de setembro de 2018. Disponível em: <https://www.correio24horas.com.br/noticia/nid/11-mil-passageiros-perdem-direito-ao-salvadorcard-por-fraude-este-ano/>. Acesso em: 07 de mar. 2023.

VIGNÉ, Júlia. Fraudes em meia passagem e gratuidade causam prejuízo de R\$39 milhões. *Correio da Bahia*. 26 de setembro de 2018. Disponível em: <https://www.correio24horas.com.br/noticia/nid/fraudes-em-meia-passagem-e-gratuidade-causam-prejuizo-de-r-39-milhoes/>. Acesso em: 07 de mar. 2023.

EXTERNALIDADES DA INTELIGÊNCIA ARTIFICIAL NA TRANSFORMAÇÃO DIGITAL DAS CIDADES

EXTERNALITIES OF ARTIFICIAL INTELLIGENCE IN THE DIGITAL TRANSFORMATION OF CITIES

Gabrielle Bezerra Sales Sarlet⁸⁸

Dionis Mauri Penning Blank⁸⁹

Resumo

O propósito da pesquisa é estudar a inteligência artificial na transformação digital das cidades identificando a ocorrência de externalidades na sua utilização. O método de abordagem utilizado é o hipotético-dedutivo, e a pesquisa foi dividida em três eixos principais – desafios para a inteligência artificial; cidades e transformação digital; e externalidades relacionadas à inteligência artificial. O uso de ferramentas tecnológicas na transformação digital das cidades, como no caso da inteligência artificial, é tema atual que desafia o combate à exclusão digital e à lesão a direitos humanos e fundamentais, especialmente a violação da privacidade e da dignidade da pessoa humana. Nesse contexto, a teoria do desenvolvimento explica que externalidades positivas e negativas constituem a estrutura de uma economia, de forma que a transformação digital apresenta soluções para a melhoria das condições de vida, mas também implica riscos ao bem-estar e à ordem social justa. Portanto, a transformação digital das cidades deve ter nas aplicações relacionadas à inteligência artificial um estado de confiabilidade e segurança, de justiça e de igualdade que propicie a tomada de decisões que garantam a efetiva melhora nas condições de vida dos habitantes, inclusive com o enfrentamento da exclusão digital por meio de políticas públicas imediatas.

Palavras-chave: inteligência artificial; transformação digital; cidades; teoria do desenvolvimento; externalidades.

88 Doutora em Direito pela Universidade de Augsburg. Pós-doutora em Direito pela Universidade de Hamburgo. Professora na PUCRS.

89 Doutor em Direito pela PUCRS. Pós-doutorando pela PUCRS.

Abstract

The purpose of the research is to study artificial intelligence in the digital transformation of cities, identifying the occurrence of externalities. The method of approach used is the hypothetical-deductive and the research was divided into three main axes – challenges for artificial intelligence; cities and digital transformation; and externalities related to artificial intelligence. The use of technological tools in the digital transformation of cities, as in the case of artificial intelligence, is a current theme that challenges to prevent of digital exclusion and the violation of human and fundamental rights, especially the violation of privacy and human dignity. In this context, development theory explains that positive and negative externalities make up the structure of an economy, so that digital transformation offers solutions for improving living conditions, but also entails risks to well-being and a fair social order. Therefore, the digital transformation of cities must rely on applications related to artificial intelligence to provide a state of reliability and security, of justice and equality that will enable decisions to be made that guarantee an effective improvement in the living conditions of the inhabitants, including tackling the digital divide through immediate public policies.

Keywords: artificial intelligence; digital transformation; cities; development theory; externalities.

1. INTRODUÇÃO

A inteligência artificial, reconhecida como inteligência não biológica, relaciona-se ao armazenamento e capacidade das máquinas, aos algoritmos e aos dados, tendo as mais diversas aplicações cotidianas (economia, geografia, saúde etc.). Existem muitos desafios a serem enfrentados pela utilização contemporânea da inteligência artificial, mas é imprescindível que a sua adoção seja segura, confiável, explicável, justa e isonômica.

A utilização de novas tecnologias traz consigo o aumento das desigualdades e dos diferentes tipos de riscos. A sociedade nunca esteve tão conectada, mas, ao mesmo tempo, desconectada da realidade coletiva. Mais de 1/3 da população mundial está digitalmente excluída, revelando a presença de muralhas digitais a serem superadas. No Brasil, quase 1/5 dos domicílios não têm acesso à internet. A pessoa humana necessariamente precisa estar no centro das preocupações relacionadas ao uso das novas tecnologias.

As cidades, no tema referente à transformação digital, têm o compromisso de oportunizar aos seus habitantes condições de bem-estar dignas e compatíveis com uma ordem social igualitária. A promoção de ferramentas tecnológicas na transformação digital das cidades, como no caso da inteligência artificial, já vem ocorrendo, impondo-se que a sua utilização não seja palco para estruturas ensejadoras de exclusão digital e de lesão

a direitos humanos e fundamentais, como a violação da privacidade e da dignidade da pessoa humana.

O estudo da teoria do desenvolvimento revela que externalidades positivas e negativas estão presentes na estrutura espacial da economia, demandando que os desafios sejam enfrentados especialmente com iniciativas regulatórias e políticas públicas efetivas. O controle exercido em relação à população, tanto pelo Estado como pelas *big techs*, deve, necessariamente, ser remodelado, de maneira que sejam afastados impactos diretos e indiretos sobre direitos e garantias fundamentais dos cidadãos.

Portanto, o objetivo desta pesquisa é identificar externalidades na adoção da inteligência artificial na transformação digital das cidades, utilizando, para isso, o método de abordagem hipotético-dedutivo, o método de procedimento monográfico e as técnicas de pesquisa bibliográfica e documental, com a divisão do trabalho em três partes principais, a saber: desafios para a inteligência artificial; cidades e transformação digital; e externalidades relacionadas à inteligência artificial.

2. DESAFIOS PARA A INTELIGÊNCIA ARTIFICIAL

Em relatório que trata sobre os riscos globais, o Fórum Econômico Mundial classificou-os por categorias (riscos econômicos, ambientais, geopolíticos, sociais e tecnológicos) e gravidade no curto (dois anos) e no longo prazo (dez anos), ocupando o 8º lugar, dentre os riscos apontados, tanto na gravidade de curto como de longo prazo, o único risco tecnológico identificado: crimes cibernéticos de forma geral e insegurança cibernética⁹⁰. De qualquer sorte, o relatório é categórico ao descrever que a tecnologia aumentará as desigualdades e que os riscos da segurança cibernética continuarão sendo motivo de preocupação constante⁹¹.

Thomas L. Friedman argumenta que o mais empolgante, relativamente à aceleração na globalização dos fluxos, “[...] é o fato de que [...] rios digitais correm agora por toda parte com igual energia, e, com os celulares e tablets em qualquer lugar, as pessoas buscam material para competir, conectar, colaborar e inventar”⁹²; porém, dados mostram que a afirmação não pode ser adotada como integralmente verdadeira. O próprio autor destaca que “somos a geração mais tecnologicamente conectada da história humana – e ainda assim nunca foi tão grande o número de pessoas que se sentem isoladas”⁹³. Na verdade,

90 FÓRUM ECONÔMICO MUNDIAL. **The global risks:** report 2023. 18. ed. Genebra, 11 jan. 2023. Disponível em: <https://www.weforum.org/reports/>. Acesso em: 9 mar. 2023, p. 6.

91 *Ibidem*, p. 8.

92 FRIEDMAN, Thomas L. **Obrigado pelo atraso:** um guia otimista para sobreviver em um mundo cada vez mais veloz. Rio de Janeiro: Objetiva, 2017, p. 161.

93 *Ibidem*, p. 519.

vai além de mero sentimento, visto que a exclusão digital é um fato (preocupante).

A União Internacional de Telecomunicações, por meio de relatório publicado em 30/11/2021, apresentou a estimativa de que 37% da população mundial – equivalente a 2,9 bilhões de pessoas – nunca tenham utilizado a internet, dos quais 96% viveriam em países em desenvolvimento⁹⁴. Significa dizer que mais de 1/3 da população mundial permanece off-line (sem conexão), estando, portanto, digitalmente excluída, revelando o abismo e a importância do tratamento do tema da conectividade em âmbito global.

O Comitê Gestor de Internet no Brasil publicou relatório, em 21/11/2022, noticiando a existência de 59 milhões de domicílios com internet no país – equivalente a 82% dos domicílios brasileiros⁹⁵. A pesquisa mostrou que a presença de computadores permaneceu estável – 39% dos domicílios –, mas sua existência é marcadamente discrepante entre os estratos da população: “[...] a proporção foi menor em domicílios das áreas rurais (20%), das regiões Norte (29%) e Nordeste (27%) e entre domicílios das classes DE (10%)”⁹⁶. O trabalho revelou que a televisão se tornou o segundo dispositivo mais usado no país para a obtenção da conexão à rede (50%), ficando atrás apenas do telefone celular (99%)⁹⁷.

A inteligência artificial que, resumidamente, pode ser caracterizada como uma integração tecnológica de algoritmos, capacidade computacional e dados traduzida por meio de máquinas, conforme sumarizam Gabrielle Bezerra Sales Sarlet e Cristina Maria de Gouveia Caldeira⁹⁸, introduz-se nesse cenário desafiador, sendo utilizada: no desenvolvimento de sistemas de aprendizagem algorítmica, em sistemas de busca, no reconhecimento facial e de voz, em sistemas de orientação, em decisões administrativas e judiciais, em diagnósticos e terapias médicas, na indústria 4.0, em plataformas e robôs de comunicação.⁹⁹

94 UNIÃO INTERNACIONAL DE TELECOMUNICAÇÕES. **Measuring digital development: fact and figures: 2021**. Genebra, 30 nov. 2021. Disponível em: <https://www.itu.int/en/mediacentre/Pages/PR-2021-11-29-FactsFigures.aspx>. Acesso em: 9 mar. 2023, p. 1.

95 COMITÊ GESTOR DA INTERNET NO BRASIL (CGI.br). **Pesquisa sobre o uso das tecnologias de informação e comunicação nos domicílios brasileiros: TIC Domicílios 2021**. São Paulo: CGI.br, 2022. Disponível em: <https://cetic.br/pt/publicacao/resumo-executivo-pesquisa-sobre-o-uso-das-tecnologias-de-informacao-e-comunicacao-nos-domicilios-brasileiros-tic-domicilios-2021/>. Acesso em: 9 mar. 2023, p. 3.

96 COMITÊ GESTOR DA INTERNET NO BRASIL (CGI.br). **Pesquisa sobre o uso das tecnologias de informação e comunicação nos domicílios brasileiros: TIC Domicílios 2021**. São Paulo: CGI.br, 2022. Disponível em: <https://cetic.br/pt/publicacao/resumo-executivo-pesquisa-sobre-o-uso-das-tecnologias-de-informacao-e-comunicacao-nos-domicilios-brasileiros-tic-domicilios-2021/>. Acesso em: 9 mar. 2023, p. 4.

97 *Ibidem*.

98 SARLET, Gabrielle Bezerra Sales; CALDEIRA, Cristina Maria de Gouveia Caldeira. A inteligência artificial e o ecossistema industrial no contexto pandêmico: uma abordagem jurídica e antropocêntrica do atual desafio das patentes na área da saúde sob o paradigma europeu. **R. Dir. Gar. Fund.**, Vitória, v. 22, n. 1, p. 131-176, jan./abr. 2021, p. 132.

99 HOFFMANN-RIEM, Wolfgang. **Teoria geral do direito digital: transformação digital: desafios para o direito**. Rio de Janeiro: Forense, 2021, p. 14-15.

Martha Gabriel, ao realizar uma analogia com o corpo humano, descreve que a inteligência artificial equivale à capacidade de processamento do cérebro biológico humano, o *big data* à memória humana, que alimenta o processamento do cérebro para virar inteligência, a internet das coisas aos sentidos, que capturam dados do mundo, e o 5G ao sistema nervoso, que possibilita a transmissão do fluxo informacional pelo corpo para ser processado no cérebro: “quanto mais desenvolvemos essas tecnologias, mais poderoso se torna o cérebro artificial do planeta e a sua inteligência [...]”¹⁰⁰.

Machado Segundo, ao trabalhar implicações da inteligência artificial, especialmente no campo do Direito, refere a necessidade de observância: a) no curto prazo: do uso por autoridades do Poder Público, dever de fundamentação, necessidade de transparência e não discriminação, erros passíveis de detecção e correção; b) no médio prazo: do manejo autônomo de máquinas, artefatos e veículos, de escolhas dramáticas (aquelas relacionadas a alternativas viáveis envolvendo prejuízos, danos e sacrifícios a bens e valores passíveis de proteção); e c) no longo prazo: das consequências jurídicas e morais do surgimento de entes artificiais com inteligência ampla (e talvez de consciência)¹⁰¹.

Roger Bootle, ao redigir sobre o significado da inteligência artificial para a estrutura política, destaca que seu impacto sobre a democracia tem relevância imediata, considerando a existência de várias ameaças à liberdade dos indivíduos e ao direito de privacidade, ainda que dentro do setor privado: “[...] os criadores das grandes empresas da era digital que dominam o mundo moderno – Amazon, Google, Apple etc. – são decididamente de um tipo: os ocidentais são todos norte-americanos, brancos e do sexo masculino”¹⁰². O autor salienta que “a revolução da IA e dos robôs precisa ser o catalisador para repensar completamente a educação – para que ela serve e como alcançar melhor os objetivos educacionais de consenso”¹⁰³ –, de forma que sejam aumentados seus aspectos criativos.

Scott Galloway, que estudou quatro gigantes de tecnologia (Apple, Amazon, Facebook e Google), denominando-as de os “Quatro Cavaleiros”, sustenta que as empresas “[...] são uma representação de deus, amor, sexo e consumo, e agregam valor à vida de bilhões de pessoas todos os dias”¹⁰⁴; no entanto, o referido autor enfatiza que elas acumulam grande poder, evitam pagar impostos, invadem a privacidade das pessoas e violam a relação empregatícia para incrementar seus lucros¹⁰⁵.

100 GABRIEL, Martha. **Inteligência artificial: do zero ao metaverso**. Barueri: Atlas, 2022, p. 13.

101 SEGUNDO, Hugo de Brito Machado. **Direito e inteligência artificial: o que os algoritmos têm a ensinar sobre interpretação, valores e justiça**. Indaiatuba: Editora Foco, 2023, p. 9-11.

102 BOOTLE, Roger. **A economia da inteligência artificial: como a IA está transformando o trabalho, a riqueza e progresso**. Rio de Janeiro: Alta Books, 2021, p. 203-204.

103 *Ibidem*, p. 223.

104 GALLOWAY, Scott. **Os quatro: Apple, Amazon, Facebook e Google**. Rio de Janeiro: Alta Books, 2019, p. 257.

105 *Ibidem*, p. 257.

Na obra 2041, Kai-Fu Lee e Chen Qiufan apontam que Facebook e Amazon têm dados que frequentemente são automaticamente rotulados pela ação dos usuários, que, por sua vez, está diretamente relacionada a métricas do negócio para maximização, ou seja, “Conforme a plataforma coleta mais dados, ela ganha mais dinheiro. Não espanta que as gigantes da internet como Google, Amazon e o Facebook tenham tido um crescimento fenomenal na última década e tenham se tornado potências de IA”¹⁰⁶.

O aprendizado profundo, que é inspirado por redes entremeadas de neurônios do cérebro humano, edifica camadas de *software* originadas de redes neurais artificiais com camadas para entrada e saída, cujas vantagens são vistas no dia a dia; no entanto, existem desvantagens, como o risco de que uma inteligência artificial conheça alguém melhor do que ela mesma, o afastamento das pessoas, o aumento da desigualdade e da discriminação, como pontuam Kai-Fu Lee e Chen Qiufan¹⁰⁷.

Amy Webb esclarece que as pessoas devem estar sempre no centro do desenvolvimento da inteligência artificial, seus sistemas devem ser confiáveis e seguros, o processo de governança deve ser capaz de analisar e mitigar eventuais danos, a inteligência artificial deve ser explicável (como um rótulo nutricional), todos os envolvidos no sistema devem se pautar por decisões constantemente éticas, deve existir um código de conduta publicado e de fácil acesso, todos devem ter o direito de questionar os sistemas de inteligência artificial, deve existir um sistema de responsabilização pública e todos os dados devem ser manejados de forma justa e igualitária¹⁰⁸.

3. CIDADES E TRANSFORMAÇÃO DIGITAL

Milton Santos sugere o espaço urbano com dois recortes: um lado, em que há extensões constituídas por pontos que se vinculam de forma contínua, lembrando a tradicional definição de região – são as chamadas horizontalidades –, e, outro lado, em que existem pontos que, separados uns dos outros, possibilitam o funcionamento global da sociedade e da economia – são as chamadas verticalidades¹⁰⁹. “De um modo geral, as cidades são o ponto de intersecção entre verticalidades e horizontalidades”¹¹⁰.

Milton Santos enfatiza que, de forma paralela, forças centrípetas e centrífugas atravessam

106 LEE, Kai-Fu; QIUFAN, Chen. **2041**: como a inteligência artificial vai mudar sua vida nas próximas décadas. Rio de Janeiro: Globo Livros, 2022, p. 48-49.

107 *Ibidem*, p. 44-52.

108 WEEB, Amy. **Os nove titãs da IA**: como os gigantes da tecnologia e suas máquinas pensantes podem subverter a humanidade. Rio de Janeiro: Alta Books, 2020, p. 246-248.

109 SANTOS, Milton. **A natureza do espaço**: técnica e tempo, razão e emoção. 4. ed. 10. reimp. São Paulo: Editora da Universidade de São Paulo, 2020, p. 284.

110 *Ibidem*, p. 286.

o território, com tendências de contrastes e de confluência, atuando em inúmeros níveis e escalas¹¹¹. “Forças centrípetas conduzem a um processo de horizontalização, forças centrífugas conduzem a um processo de verticalização. Mas, em todos os casos, sobre as forças centrípetas, vão agir forças centrífugas”¹¹², gerando uma ideia de fluxo universal, de união vertical dos lugares, a qual, entretanto, traz desordem aos subespaços nos quais se instalam, com a criação de benefícios próprios: “[...] a união vertical – seria melhor falar em unificação – está sempre sendo posta em jogo e não sobrevive senão à custa de normas rígidas”¹¹³.

Milton Santos afirma que “Cabem, pelo menos, duas perguntas em um país onde a figura do cidadão é tão esquecida. Quantos habitantes, no Brasil, são cidadãos? Quantos nem sequer sabem que não o são?”¹¹⁴. Nesse sentido, não é suficiente que a cidadania seja um estado de espírito ou mesmo uma declaração de intenções, considerando a ameaça constante de um cotidiano cruento, é indispensável que, “[...] para ter eficácia e ser fonte de direitos, ela deve ser inscrever na própria letra das leis, mediante dispositivos institucionais que assegurem a fruição das prerrogativas pactuadas e, sempre que haja recusa, o direito de reclamar e ser ouvido”¹¹⁵.

Milton Santos destaca que, com diversidades de grau e magnitude, todas as cidades no Brasil mostram problemáticas parecidas – são elementos de diferenciação a área, a região, etc., mas em todas há problemas de saúde, transporte, emprego, educação, etc. –, de forma que “Quanto maior a cidade, mais visíveis se tornam essas mazelas”¹¹⁶. Há o questionamento, então, sobre qual a maneira de definir a organização das cidades brasileiras, sobrevivendo a resposta pelo referido autor: “Quanto menor a aglomeração, menor a diversidade de sua ecologia social; quanto mais populosa e mais vasta, mais diferenciadas a atividade e a estrutura de classes, e mais o quadro urbano é compósito, deixando ver melhor as suas diferenciações”¹¹⁷.

Nesse contexto, Milton Santos é enfático ao afirmar que argumentar sobre o futuro da urbanização e das cidades “[...] é coisa temerária. Mas não falar sobre o futuro é deserção. Não se trata do futuro como certeza, porque isso seria desmentir a sua definição, mas como tendência”¹¹⁸. Nesse aspecto, ele descreve ser cada vez maior a contribuição da ciência, da tecnologia e da informação – que denomina de meio técnico científico, o qual resulta “[...] da adição, ao território, de ciência, de tecnologia, de informação, e

111 *Ibidem*, p. 286.

112 *Ibidem*, p. 287.

113 *Ibidem*, p. 287.

114 SANTOS, Milton. **O espaço do cidadão**. 4. ed. 3. reimp. São Paulo: Editora da Universidade de São Paulo, 2020, p. 19.

115 *Ibidem*, p. 20.

116 SANTOS, Milton. **A urbanização brasileira**. 5. ed. 6. reimp. São Paulo: Editora da Universidade de São Paulo, 2023, p. 105.

117 *Ibidem*, p. 105.

118 *Ibidem*, p. 129.

cria espaços inteligentes numa parte do Brasil, deixando que em outras permaneçam os espaços opacos”¹¹⁹.

De acordo com Wolfgang Hoffmann-Riem, “A transformação digital traz consigo oportunidades para melhorar as condições de vida, mas também riscos para o bem-estar dos indivíduos e para a preservação de uma ordem social justa”¹²⁰. Dionis Mauri Penning Blank sugere a superação do uso da expressão “cidades inteligentes” no Brasil para ser reconhecida a relevância de um estado de transição voltado às cidades transformadas digitalmente, de modo a serem consideradas as suas circunstâncias e particularidades, visando a promoção de recursos tecnológicos hábeis ao bem-estar de seus habitantes¹²¹.

Klaus Schwab afirma ser indispensável que as cidades e os países reúnam esforços para garantir o acesso e o uso de Tecnologias de Informação e Comunicações (TICs), as quais influenciam fortemente a 4ª Revolução Industrial¹²². Nesse sentido, o referido autor ilustra que os governos devem implementar mecanismos para acabar com o fosso digital em todas as etapas de desenvolvimento, garantindo que as cidades sejam dotadas de “[...] infraestrutura básica necessária para criar oportunidades econômicas e prosperidade compartilhada que é possível por meio de novos modelos de colaboração, eficiência e empreendedorismo”¹²³.

A Carta Brasileira para Cidades Inteligentes, conhecida como Agenda Brasileira para Cidades Inteligentes, produzida pelo Ministério do Desenvolvimento Regional, é um documento que envolve objetivos estratégicos e recomendações de enfrentamento da transformação digital nas cidades brasileiras. Dentre eles, a Carta refere a utilização das TICs para o diagnóstico e a gestão urbana, de forma que sejam usadas ferramentas de geoprocessamento para melhor compreensão dos fenômenos urbanos e para o aperfeiçoamento da gestão governamental, além da incorporação de mecanismos inovadores da ciência de dados, como a inteligência artificial e o exame de grandes quantidades de dados anonimizados (*big data*), com respeito à Lei Geral de Proteção de Dados (LGPD)¹²⁴.

De acordo com Francisco Balaguer Callejón (2021), a transformação digital da sociedade põe a Constituição Federal em uma posição delicada em relação aos fatores de poder configurados na nova ordem global, no sentido de que não se cuidam apenas de novas

119 *Ibidem*, p. 131.

120 HOFFMANN-RIEM, Wolfgang. Big data e inteligência artificial. **Revista Estudos Institucionais**, v. 6, n. 2, p. 431-506, mai./ago. 2020, p. 435.

121 BLANK, Dionis Mauri Penning Blank. **O sandbox regulatório como ferramenta para a transformação digital das cidades no Brasil**. 2022. Tese (Doutorado em Direito) – Faculdade de Direito, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2022, p. 6.

122 SCHWAB, Klaus. **A quarta revolução industrial**. São Paulo: Edipro, 2016, p. 80-81.

123 *Ibidem*, p. 81.

124 MINISTÉRIO DO DESENVOLVIMENTO REGIONAL (MDR). **Carta Brasileira para Cidades Inteligentes**. Brasília: MDR, 2020. Disponível em: <https://www.gov.br/mdr/pt-br/assuntos/desenvolvimento-urbano/carta-brasileira-para-cidades-inteligentes>. Acesso em: 9 mar. 2023, p. 43-44.

estruturas, procedimentos e técnicas, que vem bem representadas, semiologicamente, por meio dos algoritmos, mas de novos paradigmas e padrões culturais que estão modificando a maneira de pensar e os valores da sociedade¹²⁵. O referido autor afirma que a reconstrução da ordem jurídica demanda uma “Constituição do Algoritmo”, de forma que haja não apenas uma digitalização da Constituição, mas uma constitucionalização da tecnologia, para contemplar uma inteligência artificial inclusiva, a qual esteja a serviço da sociedade e controle os interesses financeiros das grandes corporações¹²⁶.

Nesse contexto, as cidades têm a transformação digital como o seu grande desafio, já que tão discrepantes entre si. Mas é a cidade que, necessariamente, deve ordenar o pleno desenvolvimento de suas funções sociais e garantir o bem-estar de seus habitantes (art. 182, *caput*, da Constituição Federal) à luz da 4ª Revolução Industrial, tendo como grande aliada a pesquisa tecnológica, que deve se voltar, preponderantemente, para a solução dos problemas brasileiros (art. 218, § 2º, da Constituição Federal)¹²⁷.

4. EXTERNALIDADES RELACIONADAS À INTELIGÊNCIA ARTIFICIAL

Em obra escrita originalmente em 1958, a respeito da formação econômica do Brasil, Celso Furtado já apontara que “[...] o Brasil [...] ainda figurará como uma das grandes áreas da terra em que maior é a disparidade entre o grau de desenvolvimento e a constelação de recursos potenciais”¹²⁸. “A teoria do desenvolvimento de Furtado aponta o subdesenvolvimento no Brasil como um processo que sofre constante metamorfose”¹²⁹. O desafio se alinhava na direção da utilização mais racional dos fatores e da distribuição mais igualitária da renda social¹³⁰.

Nesse sentido, de forma antecipada, a teoria do desenvolvimento identificou “[...] a necessidade de reformas de base e distribuição de renda para equalizar o planejamento do Estado em igualdade e direitos sociais e políticos, proporcionando o desenvolvimento

125 CALLEJÓN, Francisco Balaguer. La constitución del algoritmo. El difícil encaje de la constitución analógica en el mundo digital. In: GOMES, Ana Cláudia Nascimento; ALBERGARIA, Bruno; CANOTILHO, Mariana Rodrigues (coord.). **Direito Constitucional: diálogos em homenagem ao 80º aniversário de J. J. Gomes Canotilho**. Belo Horizonte: Fórum, 2021. E-book.

126 *Ibidem*.

127 BRASIL. **Constituição da República Federativa do Brasil de 1988**. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 9 mar. 2023.

128 FURTADO, Celso. **Formação econômica do Brasil**. 34. ed. 21. reimp. São Paulo: Companhia das Letras, 2007, p. 335.

129 GUMIERO, Rafael Gonçalves. **A teoria do desenvolvimento de Celso Furtado: diálogos entre as teses do subdesenvolvimento**. Curitiba: Appris, 2020, p. 146.

130 FURTADO, Celso. **Desenvolvimento e subdesenvolvimento**. 5. ed. 1. reimp. Rio de Janeiro: Contraponto, 2009, p. 234.

condicionado às estruturas econômicas e sociais peculiares do Brasil”¹³¹. Sendo assim, a distribuição de renda, por si, não é suficiente, sendo importante fazer a combinação com o aumento da participação popular “[...] de forma deliberativa, nas tomadas de decisões em âmbito local, para que as políticas de desenvolvimento sejam condizentes com os problemas gerados pelas especificidades das desigualdades regionais do Brasil”¹³².

Nessa perspectiva, esclarece André Luiz Cardoso Coelho que Paul Krugman, ganhador do Prêmio Nobel de Economia de 2008, produziu trabalho representativo das demarcações da origem da Nova Geografia Econômica, estabelecendo nele importantes pontos de análise: desenvolvimento desigual, custos de produção/transporte, variação do tamanho do mercado, “efeito mercado doméstico”, variação da produtividade, variação do produto agregado, variação da divisão do trabalho, retornos crescentes de escala, acontecimento histórico fortuito, causação circular cumulativa e aglomeração num ambiente de competição imperfeita¹³³.

Sendo assim, o raciocínio da Nova Geografia Econômica se ampara numa ideia de dinâmica espacial: “Haveria uma interação entre forças centrípetas (tamanho do mercado e de seus encadeamentos, densidade do mercado de trabalho, e externalidades positivas) que promovem a concentração geográfica das atividades econômicas e forças centrífugas (fatores fixos, rendas da terra e externalidades negativas) que operam na direção oposta”¹³⁴. Conforme essa interação, “[...] seriam consolidadas, ao longo do tempo, as possibilidades de desenvolvimento socioeconômico do espaço local. Assim, para essa vertente, da ação recíproca dessas duas naturezas de forças - centrípetas e centrífugas - seria formada a estrutura espacial de uma economia”¹³⁵.

Dora Kaufman identifica que a experiência humana é reivindicada pela nova ordem econômica como matéria-prima gratuita para práticas ocultadas ligadas à extração, previsão e vendas, acrescentando que “[...] o novo capitalismo, em vez de produzir produtos, gera lucro extraíndo, analisando e vendendo dados (as plataformas tecnológicas em geral, não vendem os dados, vendem publicidade direcionada/hipersegmentada)”¹³⁶.

Nessa linha de raciocínio, “Os algoritmos de IA não apenas são instrumentos comerciais, mas também permitem prever e interferir, de maneira inédita, em nossa conduta em todas as esferas da vida social. O grau e eficácia dessa intervenção, contudo, é controverso”¹³⁷.

131 *Ibidem*, p. 146-147.

132 *Ibidem*, p. 147.

133 COELHO, André Luiz Cardoso. A “velha” geografia econômica da nova geografia econômica. **Revista de Desenvolvimento econômico**, a. XV, n. 27, p. 67-74, jun.2013.

134 PELLEGRINO, Anderson César Gomes Teixeira. **O nordeste de Celso Furtado: sombras do subdesenvolvimento brasileiro**. 2003. Dissertação (Mestrado em História Econômica) – Instituto de Economia, Universidade Estadual de Campinas, Campinas, 2003, p. 32.

135 *Ibidem*, p. 32.

136 KAUFMAN, Dora. **Desmistificando a inteligência artificial**. Belo Horizonte: Autêntica, 2022, p. 158.

137 *Ibidem*, p. 159.

Dora Kaufman destaca que “As iniciativas regulatórias ainda são tímidas (algumas equivocadas) em comparação com a dimensão dos desafios, mas sinalizam as primeiras reações positivas da sociedade”¹³⁸.

O Fórum Econômico Mundial, ao examinar o risco tecnológico, menciona que avanços, especialmente, em inteligência artificial, computação quântica e biotecnologia, para países que podem custear tais tecnologias, vão fornecer uma série de soluções para crises emergentes (como novas ameaças à saúde e à segurança alimentar, mudanças climáticas), mas, por outro lado, para aqueles que não podem pagar, a desigualdade e as diferenças tendem a crescer consideravelmente¹³⁹.

Chama atenção, nesse particular, o plano – lançado em 2017 e em andamento – da Arábia Saudita de construir uma cidade inteira – chamada de Neom ou *The Line* –, que percorrerá 170km de deserto, terá 500m de altura e, com somente 200m de largura, tenderá a abrigar até 9 milhões de pessoas, com a inteligência artificial como ferramenta tecnológica central para a cidade, que tem o escopo de funcionar sem carros, ser sustentável e ter microclima regulado¹⁴⁰.

O projeto, que, por si só, ensejaria raciocínios favoráveis e desfavoráveis, lança dúvida sobre os seus reais beneficiários, tendo em conta que cidades teriam sido evacuadas à força para a realização das obras e que, sob o argumento da atração de turistas para a diversificação da economia saudita, a nova cidade, na verdade, atenderia apenas camadas elitizadas, inclusive ações relacionadas a ela estariam com acusações de *greenwashing* (propaganda/discurso falseado envolvendo temas relacionados ao meio ambiente), segundo apurado por Merlyn Thomas e Vibeke Venema¹⁴¹.

Nessa direção, Ingo Wolfgang Sarlet e Gabrielle Bezerra Sales Sarlet referem a ocorrência do tecnoautoritarismo, que consistiria no uso de recurso de ordem tecnológica, de maneira a aumentar “[...] o controle exercido pelo Estado e pelas grandes corporações sobre a população, impactando direta e indiretamente os direitos e garantias fundamentais dos cidadãos”¹⁴². Os autores citam como exemplo de sua ocorrência: a) a desinformação online, com destaque para as fake news em períodos de eleição; b) o aumento do discurso

138 *Ibidem*.

139 FÓRUM ECONÔMICO MUNDIAL. **The global risks:** report 2023. 18. ed. Genebra, 11 jan. 2023. Disponível em: <https://www.weforum.org/reports/>. Acesso em: 9 mar. 2023, p. 8.

140 BARRETO, Vitória. **O plano da Arábia Saudita de construir uma cidade dentro de arranha-céu.** Veja, 27 jul. 2022. Disponível em: <https://veja.abril.com.br/mundo/o-plano-da-arabia-saudita-de-construir-uma-cidade-dentro-de-arranha-ceu/>. Acesso em: 9 mar. 2023.

141 THOMAS, Merlyn; VENEMA, Vibeke. **A verdade sobre a cidade ecológica que a Arábia Saudita planeja no deserto.** BBC News Brasil, 25 mar. 2022. Disponível em: <https://www.bbc.com/portuguese/curiosidades-60486476>. Acesso em: 9 mar. 2023.

142 SARLET, Ingo Wolfgang; SARLET, Gabrielle Bezerra Sales. Tecno-autoritarismo, tecno-facismo societal, democracia e proteção de dados. **Revista Consultor Jurídico**, nov. 2022. Disponível em: <https://www.conjur.com.br/2022-nov-13/direitos-fundamentais-tecno-autoritarismo-tecno-facismo-societal-protexao-dados>. Acesso em: 9 mar. 2023.

de ódio e do extremismo; c) a hiperconexão e os problemas relacionados a ela, como a perda de vínculos sociais; d) vigilantismo; e e) concentração de poder informacional¹⁴³. Ainda, os referidos autores destacam que a “[...] exclusão digital — desigualdade na esfera do acesso à internet e às TICs —, [...] favorece em muito e reforça as estruturas de dominação e o tecnoautoritarismo”¹⁴⁴.

Além disso, Ingo Wolfgang Sarlet e Gabrielle Sales Sarlet revelam que os problemas mais intrigantes relacionados à utilização da inteligência artificial envolvem o risco aos direitos humanos e fundamentais, pois podem produzir e incrementar a discriminação (generalização, estereotipação, prática abusiva de perfilamento, de análise preditiva e de vigilantismo), aumentando, igualmente, assimetrias de informação, e, por consequência, “[...] violando a esfera da privacidade e a dignidade da pessoa humana”¹⁴⁵.

Sendo assim, a adoção de um Estado Democrático Digital de Direito deve ser voltada ao cidadão mediante “[...] um ecossistema inovador e suficientemente seguro, transparente e confiável, que minimize os danos e riscos causados e gerados pelas Tecnologias de Informação e Comunicação e pela inteligência artificial [...]”¹⁴⁶. Os autores descrevem que isso é possível mediante a salvaguarda técnica e jurídica orientada para “[...] proteção e promoção da dignidade da pessoa humana, de sua autodeterminação informacional, da proteção de seus dados pessoais e outros direitos e garantias fundamentais, vedando-se qualquer tipo de prática abusiva, inclusive no que diz respeito à segurança cibernética”¹⁴⁷.

Nessa linha de raciocínio, novas formas de intervenção no espaço devem ter como base a pessoa humana e o seu desenvolvimento, e não a violação de seus direitos e garantias fundamentais. Os espaços devem ser pensados de forma inclusiva, a fim de que os fossos digitais não se sobressaiam. A conectividade, sem a qual a aplicação de novas tecnologias não é possível, tem que ser viabilizada no mais amplo espectro de maneira integrada à oferta de mecanismos aptos ao alcance de habilidades digitais pela população.

5. CONCLUSÃO

A adoção de novas tecnologias intensifica desigualdades e vem atrelada a diferentes espécies de riscos (ambientais, econômicos, políticos etc.). O grande desafio está relacionado à gestão das mudanças: conectar sem excluir e progredir sem violar. A

143 *Ibidem*.

144 SARLET, Ingo Wolfgang; SARLET, Gabrielle Bezerra Sales. Tecno-autoritarismo, tecno-facismo societal, democracia e proteção de dados. **Revista Consultor Jurídico**, nov. 2022. Disponível em: <https://www.conjur.com.br/2022-nov-13/direitos-fundamentais-tecno-autoritarismo-tecno-facismo-societal-protexao-dados>. Acesso em: 9 mar. 2023.

145 SARLET, Ingo Wolfgang; SARLET, Gabrielle Bezerra Sales. **Separação informacional de poderes no direito constitucional brasileiro**. São Paulo: Associação Data Privacy Brasil, 2022, p. 33.

146 *Ibidem*, p. 37.

147 *Ibidem*.

conectividade é pressuposto básico para que se fale em transformação digital das cidades, e a utilização de sistemas de inteligência artificial demanda que a população, além de conexão, tenha acesso a ferramentas que possibilitem o aprendizado de habilidades digitais.

A transformação digital oportuniza formas de melhorar as condições de vida dos habitantes das cidades, no entanto, por outro lado, traz riscos a esses mesmos habitantes e ao próprio plano democrático. A ordem constitucional deve conter a abertura necessária à tecnologia, de maneira a contemplar a inteligência artificial voltada à sociedade, com rígido controle dos interesses econômico-financeiros das grandes empresas, ao mesmo tempo em que precisa considerar as diferentes características das regiões.

Nessa perspectiva, a teoria do desenvolvimento foi capaz de identificar a presença de diferentes externalidades na ordem econômica e a necessidade de promoção de mudanças, as quais hoje passam pelo uso da biotecnologia, da computação quântica, da inteligência artificial, tendendo à resolução de uma série de problemas. Novos projetos estão sendo produzidos e colocados em prática, como é exemplo o projeto Neom ou *The Line*, porém o cuidado necessário deve ser com a ocorrência do tecnoautoritarismo, que deve ser obstado.

A adoção da inteligência artificial na transformação digital das cidades, ainda que seja carregada de inúmeros aspectos positivos, é responsável, igualmente, por impactar negativamente especialmente direitos humanos e fundamentais, violando temas sensíveis como a privacidade e a dignidade da pessoa humana. A aplicação de ferramentas tecnológicas, como é o caso da inteligência artificial, necessariamente demanda harmonia com garantias de ordem técnica e jurídica que protejam, de forma responsável, efeitos causados pelas Tecnologias de Informação e Comunicação (TICs).

A transformação digital das cidades avança intensamente e deve ter nas aplicações relacionadas à inteligência artificial um estado de confiabilidade e segurança, de justiça e de igualdade que propicie a tomada de decisões conscientes, garantidoras da efetiva melhora nas condições de vida dos habitantes das cidades, voltadas, ao avanço da conectividade, à superação das muralhas digitais e ao enfrentamento da exclusão digital por meio de políticas públicas imediatas e com apoio em marcos regulatórios.

6. AGRADECIMENTO

Agradecimento à FAPERGS e ao CNPq pelo apoio oriundo do Edital FAPERGS/CNPq 07/2022 – Programa de Apoio à Fixação de Jovens Doutores no Brasil e à PUCRS pelo estágio pós-doutoral.

7. Referências

BARRETO, Vitória. **O plano da Arábia Saudita de construir uma cidade dentro de arranha-céu**. Veja, 27 jul. 2022. Disponível em: <https://veja.abril.com.br/mundo/o-plano-da-arabia-saudita-de-construir-uma-cidade-dentro-de-arranha-ceu/>. Acesso em: 9 mar. 2023.

BLANK, Dionis Mauri Penning Blank. **O sandbox regulatório como ferramenta para a transformação digital das cidades no Brasil**. 2022. Tese (Doutorado em Direito) – Faculdade de Direito, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2022.

BOOTLE, Roger. **A economia da inteligência artificial: como a IA está transformando o trabalho, a riqueza e progresso**. Rio de Janeiro: Alta Books, 2021.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 9 mar. 2023.

CALLEJÓN, Francisco Balaguer. La constitución del algoritmo. El difícil encaje de la constitución analógica en el mundo digital. In: GOMES, Ana Cláudia Nascimento; ALBERGARIA, Bruno; CANOTILHO, Mariana Rodrigues (coord.). **Direito Constitucional: diálogos em homenagem ao 80º aniversário de J. J. Gomes Canotilho**. Belo Horizonte: Fórum, 2021. E-book.

COELHO, André Luiz Cardoso. A “velha” geografia econômica da nova geografia econômica. Revista de Desenvolvimento econômico, a. XV, n. 27, p. 67-74, jun. 2013.

COMITÊ GESTOR DA INTERNET NO BRASIL (CGI.br). **Pesquisa sobre o uso das tecnologias de informação e comunicação nos domicílios brasileiros: TIC Domicílios 2021**. São Paulo: CGI.br, 2022. Disponível em: <https://cetic.br/pt/publicacao/resumo-executivo-pesquisa-sobre-o-uso-das-tecnologias-de-informacao-e-comunicacao-nos-domicilios-brasileiros-tic-domicilios-2021/>. Acesso em: 9 mar. 2023.

FÓRUM ECONÔMICO MUNDIAL. **The global risks: report 2023**. 18. ed. Genebra, 11 jan. 2023. Disponível em: <https://www.weforum.org/reports/>. Acesso em: 9 mar. 2023.

FRIEDMAN, Thomas L. **Obrigado pelo atraso: um guia otimista para sobreviver em um mundo cada vez mais veloz**. Rio de Janeiro: Objetiva, 2017.

FURTADO, Celso. **Formação econômica do Brasil**. 34. ed. 21. reimp. São Paulo: Companhia das Letras, 2007.

FURTADO, Celso. **Desenvolvimento e subdesenvolvimento**. 5. ed. 1. reimp. Rio de Janeiro: Contraponto, 2009.

- GABRIEL, Martha. **Inteligência artificial: do zero ao metaverso**. Barueri: Atlas, 2022.
- GALLOWAY, Scott. **Os quatro: Apple, Amazon, Facebook e Google**. Rio de Janeiro: Alta Books, 2019.
- GUMIERO, Rafael Gonçalves. **A teoria do desenvolvimento de Celso Furtado: diálogos entre as teses do subdesenvolvimento**. Curitiba: Appris, 2020.
- HOFFMANN-RIEM, Wolfgang. **Teoria geral do direito digital: transformação digital: desafios para o direito**. Rio de Janeiro: Forense, 2021.
- HOFFMANN-RIEM, Wolfgang. Big data e inteligência artificial. **Revista Estudos Institucionais**, v. 6, n. 2, p. 431-506, mai./ago. 2020.
- KAUFMAN, Dora. **Desmistificando a inteligência artificial**. Belo Horizonte: Autêntica, 2022.
- LEE, Kai-Fu; QIU-FAN, Chen. **2041: como a inteligência artificial vai mudar sua vida nas próximas décadas**. Rio de Janeiro: Globo Livros, 2022.
- MINISTÉRIO DO DESENVOLVIMENTO REGIONAL (MDR). **Carta Brasileira para Cidades Inteligentes**. Brasília: MDR, 2020. Disponível em: <https://www.gov.br/mdr/pt-br/assuntos/desenvolvimento-urbano/carta-brasileira-para-cidades-inteligentes>. Acesso em: 9 mar. 2023.
- PELLEGRINO, Anderson César Gomes Teixeira. **O nordeste de Celso Furtado: sombras do subdesenvolvimento brasileiro**. 2003. Dissertação (Mestrado em História Econômica) – Instituto de Economia, Universidade Estadual de Campinas, Campinas, 2003.
- SANTOS, Milton. **A natureza do espaço: técnica e tempo, razão e emoção**. 4. ed. 10. reimp. São Paulo: Editora da Universidade de São Paulo, 2020.
- SANTOS, Milton. **O espaço do cidadão**. 4. ed. 3. reimp. São Paulo: Editora da Universidade de São Paulo, 2020.
- SANTOS, Milton. **A urbanização brasileira**. 5. ed. 6. reimp. São Paulo: Editora da Universidade de São Paulo, 2023.
- SARLET, Gabrielle Bezerra Sales; CALDEIRA, Cristina Maria de Gouveia Caldeira. A inteligência artificial e o ecossistema industrial no contexto pandêmico: uma abordagem jurídica e antropocêntrica do atual desafio das patentes na área da saúde sob o paradigma europeu. **R. Dir. Gar. Fund.**, Vitória, v. 22, n. 1, p. 131-176, jan./abr. 2021.
- SARLET, Ingo Wolfgang; SARLET, Gabrielle Bezerra Sales. Tecno-autoritarismo, tecno-

facismo societal, democracia e proteção de dados. **Revista Consultor Jurídico**, nov. 2022. Disponível em: <https://www.conjur.com.br/2022-nov-13/direitos-fundamentais-tecno-autoritarismo-tecno-fascismo-societal-protECAO-dados>. Acesso em: 9 mar. 2023.

SARLET, Ingo Wolfgang; SARLET, Gabrielle Bezerra Sales. **Separação informacional de poderes no direito constitucional brasileiro**. São Paulo: Associação Data Privacy Brasil, 2022.

SCHWAB, Klaus. **A quarta revolução industrial**. São Paulo: Edipro, 2016.

SEGUNDO, Hugo de Brito Machado. **Direito e inteligência artificial: o que os algoritmos têm a ensinar sobre interpretação, valores e justiça**. Indaiatuba: Editora Foco, 2023.

THOMAS, Merlyn; VENEMA, Vibeke. **A verdade sobre a cidade ecológica que a Arábia Saudita planeja no deserto**. BBC News Brasil, 25 mar. 2022. Disponível em: <https://www.bbc.com/portuguese/curiosidades-60486476>. Acesso em: 9 mar. 2023.

UNIÃO INTERNACIONAL DE TELECOMUNICAÇÕES. **Measuring digital development: fact and figures: 2021**. Genebra, 30 nov. 2021. Disponível em: <https://www.itu.int/en/mediacentre/Pages/PR-2021-11-29-FactsFigures.aspx>. Acesso em: 9 mar. 2023.

WEEB, Amy. **Os nove titãs da IA: como os gigantes da tecnologia e suas máquinas pensantes podem subverter a humanidade**. Rio de Janeiro: Alta Books, 2020.

Grupo de Trabalho

SEGURANÇA INTERNACIONAL E IA

PESSOAS COORDENADORAS

Luiza Correa Dutra Magalhães
e Thallita Lima

IS THERE A STATE'S RIGHT TO SECURITY? ARTIFICIAL INTELLIGENCE AS A TOOL OF DEEPENING GLOBAL SECURITY AND DEFENSE INEQUALITIES

EXISTE UM DIREITO DO ESTADO À SEGURANÇA? A INTELIGÊNCIA ARTIFICIAL COMO FERRAMENTA DE APROFUNDAMENTO DAS DESIGUALDADES GLOBAIS EM SEGURANÇA E DEFESA

João Vitor Sales Zaidan¹⁴⁸

Wisley Simon de Lima Cunha¹⁴⁹

Abstract

This study aims to explore the consequences of the use of artificial intelligence (AI) systems to deepen defense and security inequalities and its interaction with a right to security. The possible future effects of a bigger adoption of automated systems in the strategic area are notable. One can realize, though, that, as in other moments of history, the states' capacity of assuring its own security associates itself with the level of technological advancement of the nation, which implies that Global South countries' technological inequalities also have strategic asymmetry as a consequence. Hence, a realist logic of observing the international system is intensified, as there is an even greater affirmation of those who detain the force, or, in this case, technological dominance. Although, a notion of a right to security is based on the Theory of the International Society, recognizing shared expectations from the common agreements on norms, even if they are crossed by issues such as capacity and context. Regulation proposals on legal automated systems are also explored, in a sense that rule-making and enforcing are constructed in a way so as to diminish or to break the raising of technological-strategic asymmetries. As results, it was possible to notice that the right to security is being compromised by autonomous weapons draft regulations that are vague, which are even being proposed

148 Graduando em Direito na Universidade Federal de Pernambuco. Diretor de Pesquisa da Liga de Direito Digital, Tecnologia e Sociedade da Faculdade de Direito do Recife. Pesquisador em Direitos Humanos.

149 Graduando em Direito na Universidade Federal de Pernambuco. Presidente da Liga de Direito Digital, Tecnologia e Sociedade da Faculdade de Direito do Recife.

by South Global countries.

Keywords: Inequality; artificial intelligence; security; power; regulation.

Resumo

Esta pesquisa busca explorar as consequências do uso de sistemas de inteligência artificial (IA) para o aprofundamento de desigualdades em matéria de segurança e defesa e a sua interação com um direito à segurança. São notáveis os possíveis efeitos em vista de um maior uso de sistemas automatizados na área estratégica. Percebe-se, porém, que, assim como em outros momentos da história, a capacidade de garantir a própria segurança dos Estados associa-se ao seu grau de avanço tecnológico, de modo que as desigualdades tecnológicas de países do Sul Global em relação às nações desenvolvidas implicam também em uma assimetria em matéria estratégica. Com isso, uma lógica realista de observar o sistema internacional é intensificada, na medida em que se tem uma afirmação ainda maior daqueles que detêm a força, ou, neste caso, domínio das aplicações tecnológicas. Contudo, uma noção de direito à segurança decorre de uma abordagem das relações internacionais calcada na Teoria da Sociedade Internacional, sendo baseada em expectativas compartilhadas a partir do comum acordo de normas, ainda que atravessadas por questões como a capacidade e o contexto. São também exploradas propostas de regulação de sistemas automatizados letais, no sentido de a normatização ser um caminho para diminuir ou frear o aumento das assimetrias tecnológico-estratégicas. Como resultados, pôde-se notar que a asseguarção do direito à segurança está sendo posta em jogo com projetos de regulação de armas autônomas vagos, os quais estão propostos até mesmo por países do Sul Global.

Palavras-chave: Desigualdade. Inteligência Artificial. Segurança. Poder. Regulação.

1. INTRODUCTION

Artificial intelligence (AI) is a key technological tool which already is changing societies, the world economy and how political decisions are made; it has, although, an even bigger potential of transformation. One area that certainly is impacted is the military field, as well as strategic studies and the decision-making process on conflicts and disputes between states.

One can say that the remarkable potential offered by AI in the strategic field creates a new dimension of power, since it expands the range of possibilities with which an action can be taken¹⁵⁰. More than innovations on technology itself, what has the potential of

150 GARCIA, E. V. **The Technological Leap of AI and the Global South: Deepening Asymmetries and the Future of International Security**, 2022. Available at: <https://ssrn.com/abstract=4304540>. Reached on: Feb., 25th, 2023.

changing societies — creating limits and channeling those technologies — is innovating on the use of these systems and computers, which is a result of political debate and decisions, which generate law.

Accordingly, it is important to underscore that IA has an important difference in comparison to other technologies, which is the absence of a direct causal link, which diminishes the predictability of it. In this sense, the notion of risk is crucial when it comes to regulating this kind of system. It also dictates the relations of power involving it, something that remains at the center of the discussion here proposed.

If the exact impact of these technologies is not yet known, it is possible to analyze the political and strategic dynamics of this phenomenon to current and future international relations, as it can deepen inequalities and intensify a realist logic of world politics. One question comes to mind as well, which this paper aims to discuss: is it possible to say there is a states' (and, in consequence, nations') right to security and how the technology gap is an obstacle to its assurance?

The whole international legal framework created in particular after World War II and with the United Nations gives states — especially those which detain less hard power¹⁵¹ than the others — expectations on how the international community will act¹⁵². Thus, it is possible to understand that the international system, as organized, gives nations a right to security, although it is crossed by the capacity to assure a secure strategic position.

In face of those theoretical debates, this paper has the following structure in order to discuss the presented questions. Initially, the first section explores the specific transformations of AI and related technologies in the strategic/military field and inequalities on this matter. The second section, in turn, discusses the notion of a right to security considering the interaction between legal and political issues. Finally, the third section explores the negotiations to regulate automated weapons and what that means to the Global South.

The methodology used to conduct this research is hypothetical-deductive, with data collection in the literature about the themes and in the documents from the United Nations Disarmament Affairs Office on the negotiations about automated weapons. Following this path, the research aims to describe the state in which arms that use IA and other similar systems and also to understand how the asymmetries in the access to these technologies are expressed in the dynamics of negotiating its regulation.

151 KEOHANE, R O.; NYE, J. S. **Power & Interdependence**. New York: Pearson, 2011.

152 BULL, H. **A Sociedade Anárquica**. Trad. Sérgio Bath. Brasília: Editora Universidade de Brasília; São Paulo: Imprensa Oficial do Estado de São Paulo, 2002.

2. AI POTENTIALS AND RISKS IN STRATEGY AND POWER

AI has a multitude of possible applications, including in the strategic field, armed conflicts, defense and surveillance. According to Horowitz *et al.*¹⁵³, there are some elements that may be the main ones to control in order to maintain a hegemonic position within the 21st century and AI. Data is one of them, as it is one of the bases of how automated systems make decisions and consider different scenarios. Other elements include training people to control and develop these solutions, having computing resources and incentivizing adoption of AI by organizations and branches of public and private sectors.

These systems are also sensibly useful to the decision-making process. A system that uses AI can calculate possibilities, identify vulnerabilities and make projections in such a way that it expands exponentially a country's range of action. This can lead to a security dilemma, as it mutually stimulates states to develop their AI capabilities, but it is important to underline¹⁵⁴ that leadership in that area will be defined by how societies manage the technology. In such a wide field, to define where it can be applied is a challenge itself.

Although it is not possible to measure exactly how much IA will change combat¹⁵⁵, there are already some tests, as Garcia¹⁵⁶ underscores, that show how automated systems can act. In Iraq, the United States of America has tested the Special Weapons Observation Reconnaissance Detection System (SWORDS), used to patrol streets. Another example is the Algorithmic Warfare Cross-Functional Team, or Project Maven, which applies AI in an extremely controversial area: facial recognition and in war zones. The initiative analyzes drone footage using machine learning programs to identify targets, which raises several

With this scenario being constructed, it is possible to say that, with more widespread use of AI solutions in the strategic area, a realist logic of international relations is intensified. Thus, the international system will be even more defined by those who have control and

153 HOROWITZ, M.; KANIA, E. B.; ALLEN, G. C.; SCHARRE, P. Strategic Competition in an Era of Artificial Intelligence. **Artificial Intelligence and International Security**. Washington, D. C.: Center for a New American Security, 2018. Available at: <https://www.cnas.org/publications/reports/strategic-competition-in-an-era-of-artificial-intelligence>. Reached on: March, 1st, 2023.

154 HOROWITZ *et al.*, 2018, *op. cit.*

155 GILL, A. S. Artificial Intelligence and International Security: The Long View. **Ethics & International Affairs**, v. 33, n. 02, 2019. pp. 169–179 Available at: <https://www.cambridge.org/core/journals/ethics-and-international-affairs/article/artificial-intelligence-and-international-security-the-long-view/4AB181EAF648501422257934982A4DD5> Reached on: Jun, 2nd, 2023.

156 GARCIA, E. V. **The Technological Leap of AI and the Global South: Deepening Asymmetries and the Future of International Security**, 2022. Available at: <https://ssrn.com/abstract=4304540>. Reached on: Feb., 25th, 2023.

a better application of the technology and, as such, more relative power¹⁵⁷. Risks can also be foreseen, as some AI systems can be used with few scrutiny necessities and almost secretly, which reduces room for the civil society and other institutions to oversee security policies.

As countries act more closely to a realist perspective — and, as such, far less closely to international law and regulations that limit state power —, it is also possible to imagine that human rights and international humanitarian law will suffer great violations. With a lack of regulations with a minimum level of efficiency, military powers will probably feel more comfortable to impose their interests without any constraint and civilian populations — mostly from less developed countries — often pay the price.

Whether one considers offensive¹⁵⁸ or defensive realism¹⁵⁹, dominating technological resources is crucial to guaranteeing a secure position in the international arena. This is true when considering simply keeping a nation from being attacked and specially applies to situations in which states are pursuing power projects and to occupy a hegemonic position in the world, something that increasingly depends on dominating the use of innovations.

Henceforth, an “ethics by design”¹⁶⁰ framework is very difficult to be applied in the military area because of the strategic urging of not being surpassed in the development of new techniques. As in the intelligence dilemma, in which a country not spying on others means it will simply let itself be spied on without gathering information on its adversaries, strategic technologies work under a logic of competition.

In this complex geopolitical thread, it is very sensible that some countries are constantly on the margins of these disputes: the Global South. As Horowitz *et al*¹⁶¹ remember, technological and economic development were always connected to establishing a hegemonic position in the balance of power. Navy advancements made the naval dominance of the United Kingdom possible, such as tanks, trucks and radios helped Germany’s blitzkrieg to happen at the beginning of World War II.

157 MORGENTHAU, H. J. **A política entre as nações**: a luta pelo poder e pela paz. Brasília, DF: Ed. UnB; São Paulo: Imprensa Oficial do Estado de São Paulo, 2003.

158 MEARSHEIMER, J. J. *The Tragedy of Great Power Politics*. New York: W. W. Norton & Company, 2014.

159 WALTZ, K. N. Structural Realism after the Cold War. **International Security**, Vol. 25, N° 1, pp., 2000 pp. 5–41. Available at: http://www.columbia.edu/itc/sipa/U6800/readings-sm/Waltz_Structural_Realism.pdf. Reached on: Jun, 20th, 2023.

160 EUROPEAN COMISSION. **Ethics By Design and Ethics of Use Approaches for Artificial Intelligence**, 2021. Available at: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf. Reached on Feb, 20th, 2023.

161 HOROWITZ, M.; KANIA, E. B.; ALLEN, G. C.; SCHARRE, P. *Strategic Competition in an Era of Artificial Intelligence. Artificial Intelligence and International Security*. Washington, D. C.: Center for a New American Security, 2018. Available at: <https://www.cnas.org/publications/reports/strategic-competition-in-an-era-of-artificial-intelligence>. Reached on: March, 1st, 2023.

With the Third and the Fourth Industrial Revolutions, technologies such as semiconductors, the internet and the Global Positioning System changed how societies work, but also created much ground for strategic competition to loom. Data became one of the most important concepts of this time, as things like intelligence, algorithms and AI depend on it to work and to be defined in some cases. A then unimaginable amount of data is processed daily in several different fields of work, which helps people and organizations to make decisions.

Although this narrative is not false, it certainly does not depict the whole world. As said before, many countries simply do not participate in this complex scheme of the 21st century. To these nations, the internet has a whole different meaning, as it — in some cases — may not be present at all, or does not have the same quality and cannot be trusted.

The same can be said to the states and armies of these nations, which function under an outdated logic, since they do not have the same access to the technologies that the most powerful countries do. Garcia¹⁶² writes that, for example, due to a lack of expertise, many countries gather in blocks, such as the Non-Aligned Movement, to negotiate in the Group of Governmental Experts (GGE) that discuss the matter, which works within discussions from the United Nations Convention on Certain Conventional Weapons (1980) in the UN Office on Disarmament Affairs.

It is important to recognize that colonialist perspectives have interfered in the even distribution and development of the technologies, as South Global countries' economic development is based on commodities and primary products. On the other hand, the countries of the North have specialized their productive systems on technological products, reaffirming their control of the global economy¹⁶³. In this sense, the colonies of other times served industries from developed countries with commodities.

Furthermore, it is important to underline that most colonized countries have followed the same perspective to the current days, but one of them has escaped this trajectory: China. According to Skinner¹⁶⁴, the Asian country was wise enough to gather some favorable elements: a favorable international scenario, the attractiveness of its enormous market with a cheap working force. This was important to attract companies in exchange for access to technology and the external market.

Thus, the balance of power — which had been more clearly divided and unequal — started

162 GARCIA, E. V. **The Technological Leap of AI and the Global South: Deepening Asymmetries and the Future of International Security**, 2022. Available at: <https://ssrn.com/abstract=4304540>. Reached on: Feb., 25th, 2023.

163 FAUSTO, Boris. *História do Brasil* — 2. ed. São Paulo: Universidade de São Paulo, 1995, p. 41.

164 SKINNER, M. A guerra pela supremacia tecnológica ou poder global?. **Tecnologia e Geopolítica**. *Jornal dos Economistas*. Rio de Janeiro: Conselho Regional dos Economistas, n. 378., 2021, p. 3. Available at: <http://www.corecon-rj.org.br/anexos/96B681AA2DA63EC417A02C41E5327585.pdf>. Reached on: March, 3rd, 2023.

to become more complex with other players occupying positions of power¹⁶⁵. Hence, countries have invested substantially in strategic Research and Development (R&D) to reach AI advancements, looking for these tools to tackle problems in areas such as security and military necessities. The development of these technologies allows a relentless level of technological maturity that makes the base to its mass commercial production.

Horowitz *et al*¹⁶⁶ underline that the ease with which AI inventions will be able to be replicated will also be important. The mutual relationship between commercial and military technology is also relevant to companies to enhance capabilities like image recognition, which may be used commercially, but also to strategic means.

The U.S.A's technological development has always been connected to a military perspective and the future transformation of these technologies onto marketing¹⁶⁷; many of those commercialized creations, for example, have military origins. Considering the Brazilian outlook, one can assume that military institutions are training centers and, further, developers of technology with geopolitical impact. To Duarte¹⁶⁸, military advancements are as social and subordinate to society interests as any other technology type. Beyond that, these innovations have the capacity of becoming economic vectors when they are transformed into civil application.

Beyond that, specifically in the IA field, it is impossible to forget underlying a crucial matter to the discussion about technology and power: risk control. When it comes to automated systems — notably when they are applied to public policies and conflicts —, the causality nexus is sensibly unclear and who controls the level of this risk has the power of choosing what is the acceptable damage. There is not any mutually assured destruction either, which has repercussions on the balance of power and on the asymmetries that technologies create and/or intensify.

Henceforth, the Global South — which generally does not possess this kind of technology, at least not on the same level as North countries — ends up in a position of vulnerability. If catching up is something more distant to reality¹⁶⁹, another option is to fight for

165 HOROWITZ, M.; KANIA, E. B.; ALLEN, G. C.; SCHARRE, P. *Strategic Competition in an Era of Artificial Intelligence. Artificial Intelligence and International Security*. Washington, D. C.: Center for a New American Security, 2018. Available at: <https://www.cnas.org/publications/reports/strategic-competition-in-an-era-of-artificial-intelligence>. Reached on: March, 1st, 2023.

166 *Ibidem*.

167 DUARTE, E. *Tecnologia militar e desenvolvimento econômico: Uma análise histórica*. Brasília: Institute of Applied Economic Research (Ipea), 2012. Available at: <https://www.econstor.eu/bitstream/10419/90920/1/719400392.pdf>. Reached on: Feb. 21th, 2023.

168 DUARTE, E. *Tecnologia militar e desenvolvimento econômico: Uma análise histórica*. Brasília: Institute of Applied Economic Research (Ipea), 2012. Available at: <https://www.econstor.eu/bitstream/10419/90920/1/719400392.pdf>. Reached on: Feb. 21th, 2023, p. 16.

169 GARCIA, E. V. *The Technological Leap of AI and the Global South: Deepening Asymmetries and the Future of International Security*, 2022. Available at: <https://ssrn.com/abstract=4304540>. Reached on: Feb., 25th, 2023.

regulations. It is possible, in this sense, to understand that international institutions as conceived since the end of the II World War, and especially after 1990 decade may have created a notion that could be called right to security, to be discussed in the next section.

3. INTERFACES OF THE RIGHT TO SECURITY

Since the foundation of the United Nations in 1945, there has been a deep international legal and political system, which can be theoretically compared to what the English School of the International Relations' proposal about the world. Although the norms that have been products of this framework must deal with power asymmetries, it is noticeable they created organization paradigms to countries' relations among each other.

Thus, even with certain difficulties and issues, the procedures and customs conceived in UN practice create expectations on what may happen in the near future. Bleicher¹⁷⁰, for example, analyzes how the re-citation of UN General Assembly Resolutions contribute to the expectations of it being followed. Another aspect that this work shows is that the most recited texts until the publication of it were about issues like Human Rights and decolonization.

These most recited subjects show that the international system in formation in the 20th century tried to bring obligations related to fundamental rights and the right to self determination. Both questions, but especially the latter, were and still are crucial to countries which were colonized to have a minimal level of security that they would keep being sovereign and independent from their former colonizers.

In this sense, it is possible to understand international law produced in the auspices of institutions — in special legally binding instruments such as the UN Charter and Security Council resolutions — to construct what can be called a notion of security. In the international context, one can understand it as shared expectations on the future and the capability to predict and keep their position protected from threats, with also a notion of fairness and righteousness given by international legal texts.

As Reale¹⁷¹ underscores, to have a subjective right means to maintain a relation of alterity. This means, according to the author, that, in order for someone to exercise a right, it is necessary that the provision is given on the other end. Applied to the notion of security here discussed, it is possible that all countries are holders and also providers of this right, since they must act by the norms and their assurance depends on how much countries respect them.

170

171

Is it true, however, that the notion of security is also crossed by political and strategic issues. Some regions of the world, for example, have local insecurity issues and it is not possible to forget that balance of power arrangements are also crucial for nations to stay secure and to keep themselves out of conflicts. Nevertheless, one must underscore the role of international legal principles and regulations to create obligations in order to protect those principles.

This is noticeably important to countries that possess less hard power than the others. It is not new in History the idea of those who are weaker in a group or in a society to be the ones that most demand regulations that limit the use of force¹⁷². Thus, in the international society, inequality is a characteristic that makes regulations even more important, since they provide some level of predictability and a notion of right and fair that the absence of law would not.

Another aspect of this entanglement is the interaction between an idea of security and sovereignty. The expectations created by the norms that are legally and politically binding are related in a direct way to sovereignty assurance. Although it is a basic principle when it comes to Theory of the State¹⁷³, one can understand that the constraints put by the international framework on invasions and aggressions are essential to the weakest countries to keep their integrity.

That done, a dimension in which one can understand the existence of a right to security is reached. With the creation of a framework from the UN, a very important element could be seen, which was some level of normativity to states'. That is to say all member states must comply with the UN Charter determinations, since it is an international treaty, which, being ratified, generates international obligations.

If one analyzes security through the lens of the nations – which are the “reflex” of the state –, using the classification of generations of fundamental rights created by Paulo Bonavides¹⁷⁴, it is also possible to compare a right to security to what he calls right to peace. To the praised constitutional law expert, the latter is what is defined as a fifth generation right, more advanced than transindividual rights, which are diffused through collectivity. The assurance of it is undoubtedly a challenge, as it is dependent from other generations and from other factors, as discussed in this study.

Even if sanctions related to violating those rules are crossed by power asymmetries, one can say that they are a way to, as seen, guarantee some level of predictability,

172 GARCIA, E. V. **The Technological Leap of AI and the Global South: Deepening Asymmetries and the Future of International Security**, 2022. Available at: <https://ssrn.com/abstract=4304540>. Reached on: Feb., 25th, 2023.

173 JELLINEK, G. **Teoría General del Estado**. Trad. Fernando de los Ríos. México: Fondo de Cultura Económica, 2000.

174 BONAVIDES, P. A Quinta Geração de Direitos Fundamentais. **Direitos Fundamentais & Justiça**, n. 3, abr-jun, 2008. Available at: <https://dfj.emnuvens.com.br/dfj/article/view/534>. Reached on: Jun., 10th, 2023.

creating codes and expected conducts from states. Thus, according to the International Society Theory, from the English School of International Relations¹⁷⁵, it is possible to say that states with less bellicose power and material capacities are the most profited by regulations that create constraints to the use of force.

When it comes to capacity building — especially talking about the one that happens in what is called “high politics” —, something that may come up is the concept of security dilemma. As seen in contexts like World War I, it means the situation in which a country arms itself up because of the threat of its adversary also arming itself and becoming more powerful. It was shown in the previous section how it is possible to understand that this may be occurring with IA among powerful countries.

It is impossible not to underscore, however, this reality may not be applied to the Global South. If superpowers get preoccupied with maintaining their influence zones and to catch up to the others’ capacity, the countries in the periphery of these disputes concentrate more of their thoughts on how to stay safe and secure, as well as how it is possible to keep potential conflicts out of their territory. This difference is crucial to understand how to instrumentalize this discussion.

Apart from capacity building and differences in hard power, another important factor in this complicated assurance of the right to security are military alliances. It would be possible to say this phenomenon breaks what Luhmann¹⁷⁶ calls *autopoiesis*, which means self-referencing. That is to say the system has its paradigms broken, since countries cannot be safe by only trusting in its right assurance and have to resort to the protection of a more powerful state.

With AI, a phenomenon that may be seen is a less clear distinction between Law for the Peace and for the War. Mello¹⁷⁷ underlines that International Law has this division, considering it has provisions to the “normal” organization of the international society, but also limitations and regulations for when two or more states are in a declared armed conflict. Anyhow, it is important to keep in mind, as Bull¹⁷⁸ remembers, that war is a social phenomenon — and a regulated one — and this should keep being true even with the inclusion of technologies.

Another aspect that must be considered is that states are represented by people and that its decisions, which end up being the acts of the state, are the consequence of a

175 ALVES RIBEIRO, M. M. L. A Escola Inglesa das Relações Internacionais como Instrumental Teórico Adequado de Análise das Intervenções Humanitárias. **Examâpaku Revista Eletrônica de Ciências Sociais, História e Relações Internacionais**, v. 6, n. 2, 10 mar. 2014. Available at: <https://revista.ufr.br/examapaku/article/download/2017/1269/0>. Reached on: Jun, 10th, 2023.

176 LUHMANN, Niklas. **O Direito da Sociedade**. São Paulo: Martins Fontes, 2016.

177 MELLO, Celso. D. de Albuquerque. Noções Preliminares. In: **Curso de Direito Internacional Público**. Rio de Janeiro e São Paulo: Renovar, 2000.

178 BULL, H. **A Sociedade Anárquica**. Trad. Sérgio Bath. Brasília: Editora Universidade de Brasília; São Paulo: Imprensa Oficial do Estado de São Paulo, 2002.

series of considerations. So, these regulations are one of the things considered in the phase of input. It is important that they are strong so that the output — and action to be taken — is something that complies with those rules. To consider that a level of security is existent means the risk of this output being a conflict is low — and that its cost is high.

Having discussed the complicated interfaces of a notion of a right to security, it is important to understand the concrete cases of regulation and how the discussed assurance is being treated. Another aspect to be analyzed is the asymmetries and their repercussions to the negotiations of the use of automated weapons and similar technologies, as it is going to be seen.

4. REGULATION EFFORTS: AN INTERNATIONAL SOCIETY PERSPECTIVE?

Observing the dynamics of AI, defense and technology gaps, it draws attention to the fact that power disputes are not only reinforced but also potentialized, which intensifies a realist logic of asymmetries. The inequality on defense leads the Global South to a position of insecurity because of the lack of conditions to innovate, which raises the question about how can a right to security — if existent — be assured.

Hence, one very important dimension of this inequality is the effort of regulating AI technologies, especially the military applications of them. Although one way of the Global South to level the playing field on the matter is trying to catch up with the development reached by developed countries, the idea that developing nations will be able to invest the same amount of budget on AI innovations is something distant and difficult¹⁷⁹. Hence, discussing regulations to the field is an option to counterbalance the power distribution, but it also is sensibly intricate and complex.

It is possible to say this logic comprehends the rule-making process that Bull¹⁸⁰ describes as the one adopted by the international society to create institutions — legal, practical, political or customary. It comprises steps such as creation and communication, since the norms come from discussions between states and negotiation and can only be binding if parties accept it. This also means limiting the use of force, creating constraints to it that can be put in place through mechanisms such as naming and shaming, when some actions generate political consequences to actors because of humanitarian and ethical

179 GARCIA, E. V. **The Technological Leap of AI and the Global South: Deepening Asymmetries and the Future of International Security**, 2022. Available at: <https://ssrn.com/abstract=4304540>. Reached on: Feb., 25th, 2023.

180 BULL, H. **A Sociedade Anárquica**. Trad. Sérgio Bath. Brasília: Editora Universidade de Brasília; São Paulo: Imprensa Oficial do Estado de São Paulo, 2002.

questions¹⁸¹.

Even though these phases are important, other steps acknowledged by Bull¹⁸² and crucial to their effectiveness still depend on negotiations and enforcement. This — which includes elements such as administration and protection of the rules — is, as with other international formal institutions, a challenge and can be referred to two theoretical tendencies in the English School. These are pluralism, more connected to realism and the idea of several states interacting in a selfish way, and solidarism, which is closer to liberalism/rationalism and a basis of cooperation.

It is possible to understand that the application of norms in the international context ranges between these two ways of acting that can be adopted by parties of an agreement or even to a customary rule. Although it is better for those with less hard power that solidarism prevails, cooperation is not always possible especially when its costs are higher than a position of isolationism and this is an unfavorable scenario to ensure the right to security.

Furthermore, it is important to remember that regulation does not always exactly benefit the Global South and it can also be a power instrument. This was the case of the nuclear Non-Proliferation Treaty, which former Brazilian chancellor João Augusto de Araújo Castro called an expression of “world power frosting”¹⁸³. That was because it allowed those who possessed nuclear arms to keep it, but forbidden those who did not have this kind of guns to develop it, so as to maintain power equilibrium.

Anyhow, if it is possible to establish some comparison between nuclear weapons rules and the discussion on limiting automated arms and the use of AI for military applications, one difference is clear. When the NPT was reached, nuclear technology was already pretty advanced and in use, but IA is not in this stage. Nevertheless, it is noticeable that non-proliferation agreements succeeded in creating an “anti-nuclear weapons appropriateness”. Hence, the timing is favorable for new technologies to be regulated and to prevent their abuse.

It is true, although, this analogy must be done with caution, given the distinctions in the nature of IA and nuclear power. They have implications, e. g., to the way both tools can be verified and how and even why instrumentalize limitations to it. Yet, non-proliferation is a good example of an arms regulation that, even though there are some issues, was

181 Terman, R.; Voeten, E. The relational politics of shame: Evidence from the universal periodic review. *The Review of International Organizations*, v. 13, p. 1–23, 2018. Available at: <https://link.springer.com/article/10.1007/s11558-016-9264-x>. Reached on: Jun., 12th, 2023.

182 Bull, H. *A Sociedade Anárquica*. Trad. Sérgio Bath. Brasília: Editora Universidade de Brasília; São Paulo: Imprensa Oficial do Estado de São Paulo, 2002.

183 Vargas, J. A. C. *Um Mundo Que Também é Nosso*: o pensamento e a trajetória diplomática de Araújo Castro. Brasília: Fundação Alexandre de Gusmão, 2013, p. 175.

able to create a norm of appropriateness¹⁸⁴, as can be seen with the slamming reactions to not complying with rules¹⁸⁵. The main challenge is to reach this status when it comes to powerful countries, which are, as shown, the most advanced on IA development and such.

The main instance in which the matter is being discussed is the GGE on lethal autonomous weapons systems (LAWS). It is an important multilateral forum of discussing how to regulate these technologies that enable the theoretical idea to be seen in practice. Structurally, powerful countries — that dominate the use and innovation of military AI applications — tend to be against these kinds of norms, as they are a way of imposing limits to their power, which the Global South does not have.

Hence, it is possible to analyze these disputes as a fight for a “right to security”. A state’s ability of assuring its sovereignty and a secure position in the international community is the product of several variables, which increasingly include R&D and economic capacities. So, in order to diminish their vulnerability, an interesting strategy to developing nations is to negotiate regulations that establish minimum levels of military AI development and use.

The Additional Protocol II of the Geneva Conventions (1977), in its article 36, establishes that, when a new weapon is acquired or developed, a state has the obligation of verifying if it complies with International Law. It is possible to say that this norm assures a right to security, since it binds new creations to the rules already agreed and known. The notion that LAWS must be used only in accordance with IHL is what is mostly written on proposals and what guides the debate, although it is, to current times, vague and insufficient to provide security.

That done, one can realize that the right to security’s assurance is something that ranges depending on time and, as one can imagine, on conditions and social, political and economic contexts. Besides capacity to build power to dissuade threats, it is necessary that norms are rediscussed and even newly constructed if those existing are no longer enough to assure fundamental rights and principles agreed before.

According to Garcia¹⁸⁶, although most of the 97 countries surveyed by Human Rights Watch¹⁸⁷ expressed concern with removing human control of the decision-making process

184 MARCH, J. G.; OLSEN, J. P. The logic of appropriateness. In: GOODIN, R. **The Oxford Handbook of Public Policy**. Oxford: Oxford Academic, 2009. pp. 689–708. Available at: <https://academic.oup.com/edited-volume/35474/chapter/303820733>. Reached on: Jun, 20th, 2023.

185 GOLDEMBERG, J. O TNP e o Protocolo Adicional. **Política Externa**, v. 19, nº 2, set/out/nov 2010.

186 GARCIA, E. V. **The Technological Leap of AI and the Global South: Deepening Asymmetries and the Future of International Security**, 2022. Available at: <https://ssrn.com/abstract=4304540>. Reached on: Feb., 25th, 2023.

187 HUMAN RIGHTS WATCH. **Stopping killer robots: country positions on banning fully autonomous weapons and retaining human control**. London: Human Rights Watch, 2020. Available at: <https://www.hrw.org/news/2020/08/10/killer-robots-growing-support-ban>. Reached on March 1st., 2023.

on the use of force, only 30 countries have called for a full ban on totally autonomous weapons and all of them are from the Global South¹⁸⁸. This shows the political challenge of reaching a military AI international regulation and how inequalities affected it.

One stance which may result in some level of efficacy is to raise awareness about guaranteeing fundamental rights with AI and to establish a minimum level of rules that assure human rights and humanitarian international law onto the military use of automated systems. This is the approach taken by Global South countries, which propose some principles to be adopted by most countries in order to protect people and, therefore, international security and also secure a notion of right to security¹⁸⁹.

Hence, it may be effective to create an appropriateness¹⁹⁰ rule to shape stances on military IA, such as it happens on non-proliferation. This would be important to construct limits about its application, although there are dissenting opinions about international responsibility, for instance. One should also keep in mind the need of combating what is called ethics washing¹⁹¹, which means to establish standards and pretend to meet them with a supposedly ethical rhetoric.

In the GGE negotiations, the Non-Aligned Movement, for example, suggested some elements to discussion in a 2017 statement. They include the enforcement of international law on AI military tools, responsibility of states which commit unlawful acts using LAWS, as well as the impact of those machines on security and a legally binding instrument about them. There are also more abstract — and substantially important — points, such as ethics and moral questions on LAWS, common understandings about them and the technology gap between states and its potential risks¹⁹².

Furthermore, another position that deserves to be underscored is the one from the African Group, which includes all 54 countries from the continent. Benin, according to Garcia¹⁹³, highlighted four points from their position: (i) the need of codifying legal principles based on international law to guide human control of those technologies; (ii) the importance of principles of humanity to be enunciated and take seriously; and the proposals of (iii) a

188 GARCIA, 2022, *op. cit.*

189 BODE, I. Bode, Ingvild. Norm-making and the Global South: attempts to regulate lethal autonomous weapons systems. **Global Policy**, v. 10, i. 3, 2019, p. 359–64. Available at: <https://doi.org/10.1111/1758-5899.12684>. Reached on: March, 2nd, 2023.

190 MARCH, J. G.; OLSEN, J. P. The logic of appropriateness. In: GOODIN, R. **The Oxford Handbook of Public Policy**. Oxford: Oxford Academic, 2009. pp. 689–708. Available at: <https://academic.oup.com/edited-volume/35474/chapter/303820733>. Reached on: Jun, 20th, 2023.

191 VAN MAANEN, G. AI Ethics, Ethics Washing, and the Need to Politicize Data Ethics. **Digital Society**, v. 1, n. 2, set. 2022. p. 9. Available at: <https://link.springer.com/article/10.1007/s44206-022-00013-3>. Reached on: 20th, Jun., 2023.

192 GARCIA, E. V. **The Technological Leap of AI and the Global South: Deepening Asymmetries and the Future of International Security**, 2022. Available at: <https://ssrn.com/abstract=4304540>. Reached on: Feb., 25th, 2023.

193 *Ibidem.*

ban on fully automated systems and of (iv) the conclusion of legally binding instrument on the matter.¹⁹⁴

The Brazilian government — representing the position of the biggest and most important country in Latin America — has developed prolific discussions and meetings focusing on the humanistic impact of activities on AI, following the general Global South’s position. According to Garcia¹⁹⁵, the country’s delegation submitted to the GGE a roadmap¹⁹⁶ containing bases for a legal framework on LAWS and a plan on how to put in practice 11 principles proposed by its diplomacy, with four paths. They include links between national and international regulation, cooperation, training, compliance with international law, setting up an international network on the theme, organizing international conferences on it and promoting a strategic agenda for LAWS and defense, science and technology themes.

In the Brazilian proposal, it is noticeable, from the first page, the definition of “codification through specific International Humanitarian Law (IHL) rules in a legally binding instrument” as “the ultimate goal”¹⁹⁷. Yet, the delegation does not forget the importance of multi stakeholder debate and cooperation, also proposing for the establishment of networks of experts from several areas that can contribute with study on and control of autonomous weapons.

Technical cooperation — especially when it also involves international organizations, such as the UN — is very important for an autonomous weapons control regime, if built, to work. As Gill¹⁹⁸ underscores, there is a need of reforming arms control systems, since algorithms and data are different and a very distinct application when compared to other kinds of guns, such as nuclear weapons.

One shall not forget the importance of the role of civil society in shaping those agendas

194 UNITED NATIONS OFFICE FOR DISARMAMENT AFFAIRS. Group of Governmental Experts of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems. **Statement by The African Group**. Geneva: UNODA, 2017. Available at: https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2018/gge/statements/9April_African-Group.pdf. Reached on: March, 2nd, 2023.

195 GARCIA, E. V. **The Technological Leap of AI and the Global South: Deepening Asymmetries and the Future of International Security**, 2022. Available at: <https://ssrn.com/abstract=4304540>. Reached on: Feb., 25th, 2023.

196 UNITED NATIONS OFFICE FOR DISARMAMENT AFFAIRS. CCW/GGE.1/2020/WP.3. **Operationalizing the Guiding Principles: a roadmap for the GGE on LAWS. Submitted by Brazil**. Geneva: UNODA, 2020. Available at: <https://documents.unoda.org/wp-content/uploads/2020/08/CCW-GGE.1-2020-WP.3-.pdf>. Reached on: Jun, 10th, 2023.

197 UNITED NATIONS OFFICE FOR DISARMAMENT AFFAIRS. CCW/GGE.1/2020/WP.3. **Operationalizing the Guiding Principles: a roadmap for the GGE on LAWS. Submitted by Brazil**. Geneva: UNODA, 2020. Available at: <https://documents.unoda.org/wp-content/uploads/2020/08/CCW-GGE.1-2020-WP.3-.pdf>. Reached on: June, 10th, 2023.

198 GILL, A. S. Artificial Intelligence and International Security: The Long View. **Ethics & International Affairs**, v. 33, n. 02, 2019, pp. 169–179. Available at: <https://www.cambridge.org/core/journals/ethics-and-international-affairs/article/artificial-intelligence-and-international-security-the-long-view/4AB181EAF648501422257934982A4DD5>. Reached on: Jun, 2nd, 2023.

and pressuring states not to take measures which represent violations of principles and rights. Therefore, it is vital for conferences about this theme – which also are important to analyze the evolution of the discussion – to include social groups, such as experts, non-profit organizations, think-tanks, universities and professors and young students interested in AI, defense and strategic studies.

When one analyzes a Working Paper¹⁹⁹ submitted by Australia, Canada, Japan, Poland, the Republic of Korea, the United Kingdom, and the United States, it is possible to notice some differences in tone. The focus of the proposals are in national processes in order to comply with IHL and with regulations on LAWS. It also has some worrying mentions, such as “these technologies could be used to improve the protection of civilians” (p. 1) and “damage to civilian objects excessive in relation to the concrete and direct military advantage anticipated” (p. 2). Even accountability is thought of as an internal responsibility, to be ensured by national mechanisms.

An internal regulations approach was also noticed in a Draft Protocol²⁰⁰ submitted by countries such as Argentina, Costa Rica, Sierra Leone, nations from the Global South that do not exactly possess a great level of arms development. The proposed text even has a provision saying it shall not “hamper progress in, or the inherent right of every State to the access, development, research, production, procurement, transfer and use of, emerging technologies for peaceful purposes” (p. 2). The draft, however, contains propositions related to creating obligations on transparency, review of weapons and compliance, yet stimulating states to internally apply them and “identify and share, on a voluntary basis, with other High Contracting Parties, information and good practices on the conduct of review of autonomous weapon systems” (p. 3).

In most projects submitted to the GGE, one of most frequent themes is a human-centered approach, so as to avoid a context in which humans do not have control of LAWS. It is possible to understand, hence, there is an agreement on guaranteeing what is called human on-the-loop²⁰¹, which means a human supervision, combatting out-of-loop models, that would allow fully autonomous weapons. A focus on an ethics perspective can also be seen, with a call to respect human rights and IHL while using LAWS .

199 UNITED NATIONS OFFICE FOR DISARMAMENT AFFAIRS. CCW/GGE.1/2023/WP.4/Rev.2. **Draft articles on autonomous weapon systems – prohibitions and other regulatory measures on the basis of international humanitarian law (“IHL”)**. Geneva: UNODA, 2023. Available: [https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_\(2023\)/CCW_GGE1_2023_WP.6_2.pdf](https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_WP.6_2.pdf). Reached on: Jun, 20th, 2023.

200 UNITED NATIONS OFFICE FOR DISARMAMENT AFFAIRS. CCW/GGE.1/2023/WP.6. **Draft Protocol on Autonomous Weapon Systems (Protocol VI)**. Geneva: UNODA, 2023. Available at: [https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_\(2023\)/CCW_GGE1_2023_WP.6_2.pdf](https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_WP.6_2.pdf). Reached on: Jul, 1st, 2023.

201 GARCIA, E. V. Inteligência Artificial, Paz e Segurança: Desafios para o Direito Internacional Humanitário. **Cadernos de Política Exterior**, v. 5, n. 8, 2019. pp. 99–121. Available at: https://www.academia.edu/40091908/Inteligência_Artificial_Paz_e_Segurança_Desafios_para_o_Direito_Internacional_Humanitário. Reached on: 21st, Jul, 2023.

The Report of the 2023 session²⁰² has embodied many of these points related to respecting IHL, stating that any autonomous weapon should comply with existing regulations, but without further elaborating on new rules specific to LAWS. There are, however, proposals of limitations with an ethical background, such as “limit the types of targets that the system can engage” (p. 3), although there are not many details, which may be a sign of a lack of deep consensus in negotiations. The conclusions of the report are closed with an indication that “the voluntary exchange of relevant best practices between States is encouraged”, yet with the caveat “bearing in mind national security considerations or commercial restrictions on proprietary information” (p. 3).

This sheds light to a very entangled and known issue in International Law: how to make the most powerful not use their force when it is understood as wrong by international rules. As shown by Garcia²⁰³, there were few agreements between parties in 2019 and, four years later, it is possible to say that the situation has not changed very much. The main challenge is to establish real obligations, the scope of them and decide what are the limits to autonomous weapons.

In spite of that, current texts are still vague and show that it is possible that incidents involving LAWS may happen and, without clear regulations, the world is more prone to them. This context is clearly more advantageous to the countries with the most advanced technology sector, since they are the ones that control the risk, as stated before. Hence, there is a clear threat to the assurance of a right to security, especially for the Global South. Uncertainty is the consequence of a lack of regulations, which demonstrates the complicated relation between International Law and power.

That done, there is an important actor that may help to reach some level of consensus: the civil society. Groups from the third sector can exercise pressure in order to elevate the costs of insecurity. After an obligation was taken, they also can do what Toope & Brunnée²⁰⁴ redefine as a state’s international obligation, in a sense that it can only be fulfilled with its incorporation as a duty into the society. Non-profits, the academia and other agents can contribute to this, although it is still important to maintain international control of compliance.

202 UNITED NATIONS OFFICE FOR DISARMAMENT AFFAIRS. CCW/GGE.1/2023/2. **Report of the 2023 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems**. Available at: [https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_\(2023\)/CCW_GGE1_2023_CRP.2.pdf](https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_CRP.2.pdf). Reached on: 20th, Jul, 2023.

203 GARCIA, E. V. Inteligência Artificial, Paz e Segurança: Desafios para o Direito Internacional Humanitário. **Cadernos de Política Exterior**, v. 5, n. 8, 2019. pp. 99–121. Available at: https://www.academia.edu/40091908/Inteligência_Artificial_Paz_e_Segurança_Desafios_para_o_Direito_Internacional_Humanitário. Reached on: 21st, Jul, 2023, p. 111.

204 TOOPE, Stephen J.; BRUNNÉE, Jutta. **Legitimacy and legality in international law: An interactional account**. Cambridge: Cambridge University Press, 2010.

5. FINAL REMARKS

The discussion proposed with this study shows how complicated it is to establish a definition of a right to security. Its assurance is undoubtedly included in the challenge of ensuring effectiveness to International Law, since the right to security can be considered an analogic consequence of other international obligations and recognitions, such as formal equality among all states. With LAWS and IA, more factors enter the equation of this guarantee, turning even more complex something that already does not depend only on legal issues.

This is an extremely deep discussion with more doubts than answers, which emphasizes the importance of meetings and conferences on this theme. The immediate challenge is to produce common understandings and efficient rules that establish a minimum ground on fundamental rights, LAWS and AI. The wide concept of security also includes those rights, showing how interconnected those issues are. Certainly, more can and should be discussed and theoretically and legally built on what is being proposed in this study, which is undoubtedly desired in order to deepen understandings on the right to security, new IHL regulations and so on.

Even though cooperation may be difficult in general due to the lack of trust in strategic areas, the Global South can challenge that understanding by coming together to reach technological advancements. Relations between developing and developed countries are also an intricate field of study, as it has traditionally been used for exploration and not cooperation, but some could say that it represents a path of win-win relations.

Although the latter is true, this study showed that one cannot homogenize the Global South's position. Yet some countries, such as Brazil, defend a more codified regulation, there are countries that do not have high military capacities that defend — such as powerful states — proposals focused on internal norms and laconic propositions when it comes to compliance, international sanctions, cooperation and mutual verifying.

It is always easier to analyze something that has already passed, since there is the privilege of knowing the consequences and better understanding the causes of the phenomenon. However, to study current movements is crucial for formulating questions and attempting to create ways of sorting the huge number of actors, interests and relations that can be seen in the world. That certainly applies to the deep impacts AI may offer to defense and also to inequalities on this subject.

Henceforth, it is clear how the discussions contained in this paper can surely go further, especially because its object is still in development. Not even the GGE — a small instance — has reached full agreements on deep discussions and the theme needs to be discussed in bigger forums, such as the UN General Assembly Disarmament and Security Committee and, surely, the UN Security Council. If it reaches the UNSC, it will mean the international

community recognized LAWS as a threat to international security, emphasizing the urgency of regulating them to assure a right to security.

If this right is what some would call a “fifth generation right”, such as the right to peace, the path to assure it also takes a lot of time and processes that depend on the guarantee of the predecessor rights generations. As seen, the fight for development, independence, capacities, international law norms and its efficacy are interconnected and interdependent; hence, it is possible that countries are striving in more that way to achieve their interests, even if this is done through a different way than intended.

6. Bibliographic References

ALVES RIBEIRO, M. M. L. A Escola Inglesa das Relações Internacionais como Instrumental Teórico Adequado de Análise das Intervenções Humanitárias. **Examãpaku Revista Eletrônica de Ciências Sociais, História e Relações Internacionais**, v. 6, n. 2, 10 mar. 2014. Available at: <https://revista.ufr.br/examapaku/article/download/2017/1269/0>. Reached on: Jun, 10th, 2023.

BLEICHER, S. A. The Legal Significance of Re-Citation of General Assembly Resolutions. **The American Journal of International Law**, v. 63, n. 3, Jul., 1969, pp. 444-478. Available at: <https://www.cambridge.org/core/journals/american-journal-of-international-law/article/abs/legal-significance-of-recitation-of-general-assembly-resolutions/CD86EC45DDCDD48E3DEAB435DEDF593A>. Reached on: Apr, 12th, 2023.

BODE, I. Bode, Ingvild. Norm-making and the Global South: attempts to regulate lethal autonomous weapons systems. **Global Policy**, v. 10, i. 3, 2019, pp. 359–64. Available at: <https://doi.org/10.1111/1758-5899.12684>. Reached on: March, 2nd, 2023.

BONAVIDES, P. A Quinta Geração de Direitos Fundamentais. **Direitos Fundamentais & Justiça**, n. 3, abr-jun, 2008. Available at: <https://dfj.emnuvens.com.br/dfj/article/view/534>. Reached on: Jun., 10th, 2023.

BULL, H. **A Sociedade Anárquica**. Trad. Sérgio Bath. Brasília: Editora Universidade de Brasília; São Paulo: Imprensa Oficial do Estado de São Paulo, 2002.

DUARTE, E. **Tecnologia militar e desenvolvimento econômico**: Uma análise histórica. Brasília: Institute of Applied Economic Research (Ipea), 2012. Available at: <https://www.econstor.eu/bitstream/10419/90920/1/719400392.pdf>. Reached on: Feb. 21th, 2023.

EUROPEAN COMMISSION. Ethics By Design and **Ethics of Use Approaches for Artificial Intelligence**, 2021. Available at: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf. Reached on Feb, 20th, 2023.

FAUSTO, Boris. **História do Brasil** — 2. ed. São Paulo: Editora Universidade de São Paulo, 1995.

GARCIA, E. V. Inteligência Artificial, Paz e Segurança: Desafios para o Direito Internacional Humanitário. **Cadernos de Política Exterior**, v. 5, n. 8, 2019. pp. 99–121. Available at: https://www.academia.edu/40091908/Inteligência_Artificial_Paz_e_Segurança_Desafios_para_o_Direito_Internacional_Humanitário. Reached on: 21st, Jul, 2023.

GARCIA, E. V. **The Technological Leap of AI and the Global South: Deepening Asymmetries and the Future of International Security**, 2022. Available at: <https://ssrn.com/abstract=4304540>. Reached on: Feb., 25th, 2023.

GILL, A. S. Artificial Intelligence and International Security: The Long View. **Ethics & International Affairs**, v. 33, n. 02, 2019, pp. 169–179. Available at: <https://www.cambridge.org/core/journals/ethics-and-international-affairs/article/artificial-intelligence-and-international-security-the-long-view/4AB181EAF648501422257934982A4DD5>. Reached on: Jun, 2nd, 2023.

GOLDEMBERG, J. O TNP e o Protocolo Adicional. **Política Externa**, v. 19, nº 2, set/out/nov 2010.

GOLDFARB, A.; LINDSAY, J. R. Prediction and Judgment: Why Artificial Intelligence Increases the Importance of Humans in War. **International Security**, v. 46, i. 3, 2022. Available at: https://doi.org/10.1162/isec_a_00425. Reached on 2nd, March, 2023.

HOROWITZ, M.; KANIA, E. B.; ALLEN, G. C.; SCHARRE, P. Strategic Competition in an Era of Artificial Intelligence. **Artificial Intelligence and International Security**. Washington, D. C.: Center for a New American Security, 2018. Available at: <https://www.cnas.org/publications/reports/strategic-competition-in-an-era-of-artificial-intelligence>. Reached on: March, 1st, 2023.

HUMAN RIGHTS WATCH. **Stopping killer robots: country positions on banning fully autonomous weapons and retaining human control**. London: Human Rights Watch, 2020. Available at: <https://www.hrw.org/news/2020/08/10/killer-robots-growing-support-ban>. Reached on March 1st., 2023.

JELLINEK, G. **Teoría General del Estado**. Trad. Fernando de los Ríos. México: Fondo de Cultura Económica, 2000.

KEOHANE, R O.; NYE, J. S. **Power & Interdependence**. New York: Pearson, 2011.

LUHMANN, Niklas. **O Direito da Sociedade**. São Paulo: Martins Fontes, 2016.

MARCH, J. G.; OLSEN, J. P. The logic of appropriateness. In: GOODIN, R. **The Oxford**

Handbook of Public Policy. Oxford: Oxford Academic, 2009. pp. 689–708. Available at: <https://academic.oup.com/edited-volume/35474/chapter/303820733>. Reached on: Jun, 20th, 2023.

MEARSHEIMER, J. J. **The Tragedy of Great Power Politics.** New York: W. W. Norton & Company, 2014.

MELLO, Celso. D. de Albuquerque. Noções Preliminares. *In: Curso de Direito Internacional Público.* Rio de Janeiro e São Paulo: Renovar, 2000.

MENDONÇA, B. M. Sociedade Internacional: a construção de um conceito. **Revista de Sociologia e Política**, v. 20, n. 43, out. 2012. pp. 5–22

MORGENTHAU, H. J. **A política entre as nações:** a luta pelo poder e pela paz. Brasília, DF: Ed. UnB; São Paulo: Imprensa Oficial do Estado de São Paulo, 2003.

REALE, M. **Lições Preliminares de Direito.** São Paulo: Saraiva, 2013.

SKINNER, M. A guerra pela supremacia tecnológica ou poder global?. **Tecnologia e Geopolítica. Jornal dos Economistas.** Rio de Janeiro: Conselho Regional dos Economistas, n. 378., 2021, p. 3. Available at: <http://www.corecon-rj.org.br/anexos/96B681AA2DA63EC417A02C41E5327585.pdf>. Reached on: March, 3rd, 2023.

TERMAN, R.; VOETEN, E. The relational politics of shame: Evidence from the universal periodic review. **The Review of International Organizations**, v. 13, p. 1–23, 2018. Available at: <https://link.springer.com/article/10.1007/s11558-016-9264-x>. Reached on: Jun., 12th, 2023.

UNITED NATIONS OFFICE FOR DISARMAMENT AFFAIRS. CCW/GGE.1/2020/WP.3. **Operationalizing the Guiding Principles: a roadmap for the GGE on LAWS. Submitted by Brazil.** Geneva: UNODA, 2020. Available at: <https://documents.unoda.org/wp-content/uploads/2020/08/CCW-GGE.1-2020-WP.3-.pdf>. Reached on: June, 10th, 2023.

UNITED NATIONS OFFICE FOR DISARMAMENT AFFAIRS. CCW/GGE.1/2023/WP.4/Rev.2. **Draft articles on autonomous weapon systems – prohibitions and other regulatory measures on the basis of international humanitarian law (“IHL”).** Geneva: UNODA, 2023. Available: [https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_\(2023\)/CCW_GGE1_2023_WP.6_2.pdf](https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_WP.6_2.pdf). Reached on: Jun, 20th, 2023.

UNITED NATIONS OFFICE FOR DISARMAMENT AFFAIRS. CCW/GGE.1/2023/WP.6. **Draft Protocol on Autonomous Weapon Systems (Protocol VI).** Geneva: UNODA, 2023. Available at: https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons

[Systems \(2023\)/CCW_GGE1_2023_WP.6_2.pdf](#). Reached on: July, 1st, 2023.

UNITED NATIONS OFFICE FOR DISARMAMENT AFFAIRS. CCW/GGE.1/2023/2. **Report of the 2023 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems**. Available at: [https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_\(2023\)/CCW_GGE1_2023_CRP.2.pdf](https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_CRP.2.pdf). Reached on: 20th, Jul, 2023.

UNITED NATIONS OFFICE FOR DISARMAMENT AFFAIRS. Group of Governmental Experts of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems. **Statement by The African Group**. Geneva: UNODA, 2017. Available at: https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2018/gge/statements/9April_African-Group.pdf. Reached on: March, 2nd, 2023.

TOOPE, Stephen J.; BRUNNÉE, Jutta. **Legitimacy and legality in international law: An interactional account**. Cambridge: Cambridge University Press, 2010.

VAN MAANEN, G. AI Ethics, Ethics Washing, and the Need to Politicize Data Ethics. **Digital Society**, v. 1, n. 2, set. 2022. p. 9. Available at: <https://link.springer.com/article/10.1007/s44206-022-00013-3>. Reached on: 20th, Jun., 2023.

VARGAS. J. A C. **Um Mundo Que Também é Nosso: o pensamento e a trajetória diplomática de Araújo Castro**. Brasília: Fundação Alexandre de Gusmão, 2013.

WALTZ, K. N. Structural Realism after the Cold War. **International Security**, Vol. 25, Nº 1, pp., 2000 pp. 5–41. Available at: http://www.columbia.edu/itc/sipa/U6800/readings-sm/Waltz_Structural_Realism.pdf. Reached on: Jun, 20th, 2023.

INTELIGÊNCIA ARTIFICIAL NO COMBATE AO TRÁFICO INTERNACIONAL DE CRIANÇAS E ADOLESCENTES: PARA AQUELES QUE NÃO TIVERAM A CHANCE DE BUSCAR AJUDA, DE SEREM VISTOS E DE SEREM PROTEGIDOS

ARTIFICIAL INTELLIGENCE IN THE COMBAT AGAINST INTERNACIONAL CHILD AND ADOLESCENT TRAFFICKING: FOR THOSE WHO DIDN'T HAVE THE CHANCE TO SEEK FOR HELP, TO BE SEEN AND TO BE PROTECTED

Geórgia Bertoldi Verzi Silva²⁰⁵

Resumo

O tráfico internacional de pessoas caracteriza uma prática criminosa e degradante que transcende milhares de anos, qualificando uma realidade preocupante e colocando uma situação de crise global. Essa realidade apavora em grau aumentado quando se fala em crianças e adolescentes vítimas dessa prática, demonstrando a carência por mecanismos de combate efetivos e que vislumbrem um direito protetivo aplicável sobre diversos campos de atuação. A comunidade internacional desenvolveu diversos instrumentos normativos que vislumbram os direitos desse grupo, todavia, esses instrumentos ainda não são suficientes no combate a essa prática, sendo necessária a convocação dos Estados para que atuem na efetivação do cumprimento das obrigações assumidas quanto a buscar mecanismos e formas de atuação que garantam a proteção dos direitos que estão sendo violados. Nesse sentido é aberto espaço para novos campos se interligarem ao Direito e viabilizar novas abordagens, permitindo que meios como o uso de Inteligência Artificial possam se colocar como ferramentas viáveis no controle ou prevenção. Assim,

205 Graduada em Direito pela Universidade Estadual de Ponta Grossa. Pós-graduanda em Processo Civil pela Faculdade IBMEC e pela Fundação Escola do Ministério Público do Estado do Paraná. Membro do Grupo de Pesquisa “Mind The Gap – Direito para o futuro”.

resta necessário verificar tal aplicação no plano fático e quais os caminhos que foram percorridos até chegar a essa possibilidade. Essa análise se desenvolveu através da alcinha do método dedutivo utilizando-se de pesquisa bibliográfica. Acredita-se que ao fomentar a busca e a pesquisa por novos meios de atuação protetiva tem-se uma corrente de interesses unidos com um foco comum, a garantia da aplicabilidade dos direitos inerentes a esses indivíduos.

Palavras-chave: Inteligência Artificial; tráfico internacional infantil; novo direito; direito protetivo.

Abstract

International human trafficking characterizes a criminal and degrading practice that transcends thousands of years, qualifying a worrying reality and posing a situation of global crisis. This reality is terrifying to an increased degree when it comes to children and adolescents who are victims of this practice, demonstrating the lack of effective combat mechanisms that attain a protective right applicable to different fields of action. The international community has developed several normative instruments that envisage the rights of this group, however, these instruments are still not enough to combat this practice, requiring States to act in order to effectively fulfill the obligations assumed with regard to seeking mechanisms and forms of action that guarantee the protection of the rights that are being violated. In this purpose, space is opened for new fields to interconnect with Law and enable new approaches, allowing means such as the use of Artificial Intelligence to become viable tools in the control or prevention of this situation. Thus, it remains necessary to verify such an application in the factual level and which paths were taken to reach this possibility. This analysis was developed through the nickname of the deductive method using bibliographic research. It is believed that by encouraging the pursuance and research for new means of protective action, there is a chain of interests united with a common interest, guaranteeing the applicability of the rights inherent to these individuals.

Keywords: Artificial Intelligence; international child trafficking; new law; protective right.

1. INTRODUÇÃO

Tráfico na acepção clássica e simples do dicionário *Oxford Languages* traduz-se em “1. trato mercantil, negócio, comércio; tráfico. 2. negócio clandestino, ilícito, ilegal”. Dando continuidade ao significado em termos jurídicos, o tráfico caracteriza-se por ser uma prática de negócio clandestino, fraudulento que tem aspectos em dois prismas, gera lucro ao criminoso e viola direitos de uma sociedade como um todo²⁰⁶.

Partindo para o tráfico internacional de pessoas, depara-se com uma prática de comércio voltada à negociação de seres humanos com vistas à obtenção de resposta de cunho econômico, sucumbindo direitos básicos e fundamentais de todas as pessoas. Esse ato criminoso ainda se mostra muito recorrente em todo o mundo, sendo uma prática que perdura há muito tempo e que se aproveita da escassez de informações e dados sobre seu funcionamento para manter-se em atuação. Mesmo com a existência de números camuflados e não fidedignos à real situação que o tráfico de pessoas controla, pode-se perceber um número altíssimo de casos retratados e noticiados, bem como constatar que essas informações apenas florem uma realidade muito mais grave, tornando esse assunto uma calamidade social de altíssima importância a ser tratada.

Realiza-se, então, um recorte rigoroso quando ao tráfico internacional, perfazendo a análise sobre crianças e adolescentes, por tratar-se de uma demanda ainda mais sensível e que caracteriza mecanismos protetivos de caráter específico, tanto a nível nacional, quanto internacional. Para isso, fez-se o uso do método científico dedutivo de abordagem, norteando-se pelo plano doutrinário, com direcionamento bibliográfico.

Nessa esteira, verificou-se outro aspecto temático importante que merece destaque. A tecnologia segue em desenvolvimento constante, apresentando ferramentas modernas e de alto grau de utilidade social, facilitando a vida das pessoas, gerando mecanismos de segurança mais efetivos, disseminando informações em curto espaço de tempo, dentre diversos outros avanços contemporâneos.

Nesse cenário, mas do lado oposto, a tecnologia também viabilizou a facilidade no uso de ferramentas ao mundo do crime, na praticidade em esconder dados, camuflar números, gerenciar operações globais de um só lugar, ou simplesmente dificultar a proteção dos indivíduos que sofrem as consequências desses atos criminosos. Isto é, restava ausente a união de tecnologia e direito no combate ao tráfico internacional.

À vista disso, busca-se conseguir identificar mecanismos trazidos pelo desenvolvimento da tecnologia e associá-los a uma lacuna protetiva relacionada à criança e ao adolescente, de modo individualizador, como também fomentar o debate sobre o assunto e, conseqüentemente, o aumento da pesquisa sobre o tema.

2. TRÁFICO INTERNACIONAL DE CRIANÇAS E ADOLESCENTES - BIG DEAL

“Eu tinha medo de ser espancada ou morta, caso resistisse. Estava apavorada e traumatizada, queria somente morrer. Não tinha controle algum.”

“No começo eu resisti, então fui forçada a beber a urina de um homem. Em seguida, fui amarrada, coberta com formigas e chicoteada com um cabo elétrico. Depois disso não ofereci mais resistência.”

Já Sasha foi violado em grupo. “Vários homens me amarraram junto com outros meninos e nos forçaram a ingerir drogas. Sedados, fomos filmados e fotografados enquanto éramos estuprados.” (tradução da autora)²⁰⁷.

A prática do tráfico de seres humanos engloba o recrutamento, o transporte, a transferência, o alojamento e a recepção de indivíduos, mediante a utilização de ameaça, coerção física ou outras formas de pressão, tais como rapto, fraude, engano, abuso de autoridade ou exploração de vulnerabilidades, além do oferecimento ou recebimento de remuneração, ou benefícios com o objetivo de obter consentimento para o controle de uma pessoa sobre outra, visando à exploração. Essa definição é estabelecida pelo Protocolo Relativo à Prevenção, Repressão e Punição do Tráfico de Pessoas, em Especial Mulheres e Crianças, que complementa a Convenção das Nações Unidas contra o Crime Organizado Transnacional, amplamente conhecida como Convenção de Palermo.²⁰⁸

O tráfico de pessoas permeia temas plurais, de diversos ramos, sejam questões relacionadas aos direitos humanos, questões de gênero, aplicação legislativa, defasagens primárias na estrutura governamental, o comércio que cerceia toda essa indústria, entre outros cenários, de modo que se faz pungente uma atuação plurissubsistente, compreendendo um campo abrangente de pesquisas que exponham diversas ópticas, sendo posto, inclusive, pela Corte Internacional de Justiça, como uma obrigação *erga omnes*,²⁰⁹ tendo em vista tratar-se de proteção aos direitos humanos, devendo ser um compromisso adotados por todo e qualquer indivíduo²¹⁰.

Dentre os principais problemas apresentados no combate ao tráfico internacional de crianças se constata a insuficiência de dados e a carência de intercomunicação das

207 KELLY, Emma. ECPAT UK. **Bordering on Concern: Child Trafficking in Wales**. Copyright ECPAT. Londres, 2009.

208 UNODC - UNITED NATIONS OFFICE ON DRUGS AND CRIME. Tráfico de Pessoas e Contrabando de Migrantes. Disponível em: <https://www.unodc.org/lpo-brazil/pt/trafico-de-pessoas/index.html#:~:text=O%20tr%C3%A1fico%20de%20pessoas%20%C3%A9,receber%20pagamentos%20ou%20benef%C3%ADcios%20para>. Acesso em: 17 maio 2023.

209 Expressão em latim que significa “contra todos”, “frente a todos” ou “relativamente a”. Costuma ser usada no âmbito jurídico para se referir a uma lei ou norma que vale para todos os indivíduos (efeito vinculante).

210 FERREIRA, Micaela Amorim; BORGES, Paulo Cesar Corrêa. **TRÁFICO DE PESSOAS COMO PROBLEMA INTERNACIONAL E PANORAMA LEGISLATIVO DE COMBATE**. In: Brasil. Ministério Público Federal. 2º Câmara de Coordenação e Revisão. Tráfico de pessoas / 2ª Câmara de Coordenação e Revisão, Criminal; organização: Stella Fátima Scampini. – Brasília : MPF, 2017. p. 20-39. Coletânea de artigos Tráfico de Pessoas; v. 2. Disponível em: <https://www.mpf.mp.br/atuacao-tematica/ccr2/publicacoes/coletaneas-de-artigos>. Acesso em: 14 jan. 2023.

informações. Esses dois pontos podem ser traduzidos em: por ser uma prática criminal que se perfaz nos caminhos escuros da traficância e que se concretiza na baixa taxa de recuperação dos desaparecidos, conflui em um resultado ínfimo de respostas aos questionamentos que circundam essa prática, como também na deficiência em identificar vítimas, a qual está plenamente associada à total ineficiência em prevenir que a condição de vítima se torne realidade.

Ainda que com acesso a dados limitado, pode-se constatar o estado crítico brasileiro em que cerca de 50.000 crianças desaparecem por ano, sendo que 15 a 20% delas permanecem desaparecidas por muito tempo. Embora não se possa afirmar que todas sejam vítimas de tráfico humano, há uma falha evidente na proteção de crianças e adolescentes²¹¹.

Dessa forma, o Estado deve aplicar políticas de combate ao tráfico de pessoas divididas em três tipos: políticas econômicas, políticas de migração e políticas de enfrentamento ao tráfico, destacando-se que a última só terá efetividade a partir do momento em que as outras duas estiverem em consonância, dando suporte a um gerenciamento forte à supressão desse ato criminoso²¹².

O caso Amber²¹³, nesse momento, se faz interessante a título de verificação da vulnerabilidade real identificada na rotina de diversas crianças e adolescentes, e a importância que, aparentemente, pequenas atitudes podem ter no âmbito protetivo e de resolução consistente dessas demandas.

No dia 13 de janeiro de 1996, Amber Rene Hagerman, uma criança de nove anos, foi subitamente sequestrada em Arlington, Texas, EUA, desencadeando uma cadeia de eventos que mudaria a forma como o mundo lida com casos de crianças desaparecidas. Enquanto brincava em um estacionamento abandonado, Amber foi levada por um sequestrador desconhecido, deixando sua família e comunidade devastadas. Quatro dias depois, seu corpo foi descoberto em um riacho próximo, revelando sinais perturbadores de violência sexual e múltiplos ferimentos.

A tragédia do sequestro e assassinato de Amber chamou a atenção nacional nos Estados Unidos, mas foi a ação específica de Diana Simone, uma mãe que acompanhava as informações e se compadeceu com a dor vivida por Donna Whitson, mãe de Amber, que

211 AGÊNCIA SENADO. **Trata-se de estimativa do Conselho Federal de Medicina. Desaparecimento de crianças: Conselho de Medicina faz alerta na Comissão de Direitos Humanos.** Disponível em: <https://www12.senado.leg.br/noticias/materias/2018/04/06/desaparecimento-de-criancas-conselho-de-medicina-faz-alerta-na-comissao-de-direitos-humanos>. Acesso em: 10 jan. 2023.

212 RIBEIRO, Renata. **O TRÁFICO INTERNACIONAL DE CRIANÇAS E ADOLESCENTES: POLÍTICAS PÚBLICAS DE PREVENÇÃO E ENFRENTAMENTO NO BRASIL.** 2019, 61f. Monografia. (Curso Bacharelado em Direito). Universidade de Santa Cruz do Sul, 2019.

213 BORGES, B. S. **Alerta Amber: a trágica história que deu origem ao sinal de emergência.** [S. l.], 29 ago. 2023. Disponível em: <https://canalcienciascriminais.com.br/alerta-amber-a-tragica-historia/>. Acesso em: 10 out. 2023.

desencadeou uma mudança significativa na resposta a casos semelhantes. Simone expressou sua indignação e compaixão ao questionar por que o governo priorizava alertas meteorológicos em vez de notificar a população sobre crianças desaparecidas. Essa pergunta importante levou à criação do programa America's Missing: Broadcast Emergency Response, ou Alerta Amber, que visa mobilizar a comunidade para auxiliar na localização rápida de crianças desaparecidas.

Desde sua implementação em 1996, o Alerta Amber tem sido amplamente adotado. O programa está em operação em 50 estados dos EUA e em 22 países em todo o mundo. Os alertas, Amber, são emitidos por meio de diversos canais de comunicação, incluindo rádio, televisão, mensagens de texto e telões em rodovias, a fim de alcançar o maior número possível de pessoas e obter informações relevantes para a busca imediata.

Os resultados do Alerta Amber informados pelo Departamento de Justiça dos EUA, no período compreendido entre 2009 e 2019, indicam que mais de 900 crianças foram resgatadas graças à ativação do programa²¹⁴. O Alerta Amber demonstrou-se uma ferramenta crucial na recuperação de crianças desaparecidas, proporcionando esperança e apoio às famílias em momentos de desespero, originando-se de forma despretensiosa e em meio à utilização de mecanismos de compartilhamento de informação dispostos à época.

O caso de Amber Hagerman, embora trágico e sem resolução do crime, desencadeou uma resposta global que impactou positivamente a forma como as autoridades e a comunidade em geral lidam com o desaparecimento de crianças²¹⁵.

Nesse mesmo sentido, cabe citar o projeto DNA-PROKIDS, criado em 2009 no Laboratório de Identificação Genética da Universidade de Granada, que tem como objetivo combater o tráfico ilegal de seres humanos, especialmente crianças e adolescentes, por meio da identificação genética das vítimas e suas famílias. O projeto estabelece parcerias com países da Europa, Ásia, Américas e, incluindo o Brasil, tendo como principal objetivos a redução de casos de adoção ilegal, uma vez que muitas dessas adoções podem envolver subtração de crianças. No DNA-PROKIDS, as mães que desejam entregar seus filhos para adoção devem fornecer provas genéticas de que são as mães biológicas. Os países participantes do projeto têm um banco de dados integrado para compartilhar informações

214 FIGARO-GARCIA, Claudia. **Uma proposta de prática psicológica para casos de desaparecimento de crianças e adolescentes**. Orientador: Prof. Dra. Eliana Herzberg. 2010. 249 p. Dissertação (Doutorado em Psicologia) - Universidade de São Paulo, São Paulo, 2010. DOI 10.11606/T.47.2010.tde-26072010-123243. Disponível em: <https://www.teses.usp.br/teses/disponiveis/47/47133/tde-26072010-123243/pt-br.php>. Acesso em: 23 maio 2023.

215 RAMOS, Alana Figueiredo. **Inovações nos métodos de busca de crianças e adolescentes desaparecidos: adaptação do alerta amber na legislação brasileira implantada pelo adendo da lei 13.812 de 2019**. Orientador: Camila Damasceno de Andrade. 2022. 56 p. Trabalho de Conclusão de Curso (Bacharelado em Direito) - Universidade do Sul de Santa Catarina, Tubarão, 2022. Disponível em: <https://repositorio.animaeducacao.com.br/handle/ANIMA/26868>. Acesso em: 23 maio 2023.

e, por conseguinte, conectar cenários protetivos e novas práticas a esse respeito²¹⁶.

A partir disso, dentre os três tipos de políticas de combate ao tráfico de pessoas apresentadas anteriormente, gize-se apontar a possibilidade de um quarto plano de enfrentamento que envolve o uso de IA para proteção e cruzamento de dados. A atividade é clandestina, com poucas atualizações, baixa taxa de recuperação de vítimas e difícil acesso à informação, dificultando a criação de políticas e mecanismos repressivos eficazes. Essa dificuldade de acesso à informação constante e atualizada impossibilita a produção de políticas públicas, bem como a utilização de mecanismos repressivos eficazes²¹⁷.

Nesta esteira, a curva de estudo permite mostrar o caminho da vítima desde sua captura até sua entrega, percorrendo barreiras, locais, seguranças e situações que facilitaram essa passagem, desde seu centro familiar, até a parte protetiva do Estado. Para interromper esse caminho, é necessário usar ferramentas modernas para criar novas possibilidades de proteção, tal qual foi possibilitado através do Alerta Amber e dos demais identificados nesta pesquisa.

Dados do “Relatório sobre o Tráfico de Pessoas 2022 – Brasil” disponibilizado pela Embaixada e Consulado dos EUA no Brasil indicam que no que se refere ao combate ao tráfico sexual e à identificação das vítimas entre populações extremamente vulneráveis, como crianças e pessoas LGBTQI+, o governo apresentou esforços limitados, verificando-se que algumas autoridades possuem um conhecimento deficiente sobre o crime de tráfico de pessoas, resultando na penalização de vítimas vulneráveis por atos ilícitos que foram coagidas a cometer por seus traficantes. Além disso, os mecanismos de proteção às vítimas, incluindo abrigos, continuaram sendo inadequados e variaram consideravelmente de estado para estado²¹⁸.

2.1. A nova ordem jurídica – direito digital

Nesse momento, aposta-se no uso da inteligência artificial, mecanismo esse que surgiu frente às necessidades de desenvolvimento presentes em diversos segmentos, sendo um deles o uso de IA como uma ferramenta do direito, uma ferramenta do sistema que

216 FIGARO-GARCIA, Claudia. **Uma proposta de prática psicológica para casos de desaparecimento de crianças e adolescentes**. Orientador: Prof. Dra. Eliana Herzberg. 2010. 249 p. Dissertação (Doutorado em Psicologia) - Universidade de São Paulo, São Paulo, 2010. p.33. DOI 10.11606/T.47.2010.tde-26072010-123243. Disponível em: <https://www.teses.usp.br/teses/disponiveis/47/47133/tde-26072010-123243/pt-br.php>. Acesso em: 23 maio 2023.

217 GIRONI, Marcela Caroline Vaz. OS MECANISMOS DE REPRESSÃO AO TRÁFICO DE PESSOAS. In: Brasil. **Ministério Público Federal. 2º Câmara de Coordenação e Revisão. Tráfico de pessoas / 2ª Câmara de Coordenação e Revisão, Criminal**; organização: Stella Fátima Scampini. – Brasília: MPF, 2017. p. 52-100. Coletânea de artigos Tráfico de Pessoas; v. 2. Disponível em: <https://www.mpf.mp.br/atuacao-tematica/ccr2/publicacoes/coletaneas-de-artigos>. Acesso em: 14 jan. 2023.

218 BRASIL. Governo Federal. **Relatório sobre o Tráfico de Pessoas 2022 – Brasil**. Brasília, DF, 2022. Disponível em: <https://br.usembassy.gov/pt/relatorio-sobre-o-trafico-de-pessoas-2022-brasil/>. Acesso em: 10 abr. 2023.

contribua de modo positivo e contemporâneo a questões que transcendem gerações²¹⁹.

Essa nova possibilidade vem de encontro à uma grande preocupação do CNJ, a qual se materializa na questão de os sistemas dos diversos tribunais do país conseguirem “conversar” entre si, ou seja, tornar os sistemas integrados, com intercâmbio de dados e capacidade de realização de trocas mútuas. Essa funcionalidade é necessária para que se possa produzir métricas, diagnósticos, como também, na seara da presente pesquisa, definir novas formas de atuação protetiva²²⁰. Assim, estar-se-ia unindo diversos prismas de ação com um único objetivo, impedir essa prática de violência, fazendo com que, a exemplo do que ocorreu nos EUA com o Alerta Amber, onde 900 crianças, no período de 2009 a 2019, pudessem ser resgatadas através de uma das tantas mudanças que podem ocorrer num sistema de proteção à criança e adolescentes²²¹.

Entretanto, essa transformação precisa fundar-se no respeito e na preservação dos direitos humanos fundamentais dos indivíduos, sendo necessária a observação de princípios que consigam abranger a implementação dessas novas tecnologias, trazendo uma situação confortável e segura na sua aplicação e a garantia de um uso confiável.

Nessa sequência, apresenta-se a Carta Européia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu ambiente, a qual foi adotada pela Comissão Europeia para Eficácia da Justiça. A Carta elencou cinco princípios essenciais para que se possa garantir o uso da Inteligência Artificial com segurança e bases sólidas, sendo esses:

- a. Princípio do Respeito pelos Direitos Fundamentais: garantir que a elaboração e a implementação de Inteligências Artificiais sejam coadunáveis com os direitos e garantias fundamentais;
- b. Princípio da Não-Discriminação: certificar que essas tecnologias não desenvolvam, reproduzam ou agravem as discriminações existentes, através da análise de dados algorítmicos recebidos em sua fase de implantação e no uso de aprendizado de máquina;

219 BRYSON, Joanna J. **The Artificial Intelligence of the Ethics of Artificial Intelligence: An Introductory Overview for Law and Regulation**. In: DUBBER, Markus D.; PASQUALE, Frank; DAS, Sunit (Ed.). *The Oxford Handbook of Ethics of Artificial Intelligence*. Oxford: Oxford University Press, 2020.

220 BRAGANÇA, Fernanda; BRAGANÇA, Laurinda Fátima da F. P. G. **Revolução 4.0 no Poder Judiciário: LEVANTAMENTO DO USO DE INTELIGÊNCIA ARTIFICIAL NOS TRIBUNAIS BRASILEIROS**. Revista da Seção Judiciária do Rio de Janeiro. Rio de Janeiro, v. 23, n. 46, p. 65-76, jul./out. 2019.

221 RAMOS, Alana Figueiredo. **Inovações nos métodos de busca de crianças e adolescentes desaparecidos: adaptação do alerta amber na legislação brasileira implantada pelo adendo da lei 13.812 de 2019**. Orientador: Camila Damasceno de Andrade. 2022. 56 p. Trabalho de Conclusão de Curso (Bacharelado em Direito) - Universidade do Sul de Santa Catarina, Tubarão, 2022. p.45. Disponível em: <https://repositorio.animaeducacao.com.br/handle/ANIMA/26868>. Acesso em: 23 maio 2023.

- c. Princípio da Qualidade e Segurança: atestar que, no processamento de decisões e dados judiciais, as informações inseridos em um software – que arquiteta um algoritmo de aprendizado de máquina – venha através de fontes seguras e de dados intangíveis, para garantia de um ambiente tecnológico seguro;
- d. Princípios da Transparência, Imparcialidade e Justiça: garantir que o acesso aos métodos de processamento seja acessível e de fácil compreensão, explicado de forma clara e com linguagem familiar sobre a natureza dos serviços oferecidos, as ferramentas desenvolvidas, os riscos de erro e o desempenho;
- e. Princípio “Sob Controle do Usuário”: certificar que os usuários sejam informados e possuam o controle das suas escolhas, dando aos profissionais do sistema judicial a possibilidade de que, a qualquer momento, sejam revistas as decisões judiciais e os dados fornecidos para a produção de resultados²²².

Ato contínuo, deve-se buscar a garantia de uma aplicação coerente e segura da IA, para que novos mecanismos e questões que a utilizarem possam servir-se de um direcionamento confiável. Reforça-se os cinco desafios éticos dos dados (5V’s) que deverão ser seguidos durante todo esse caminho, sendo eles:

- volume, com evidente tendência ao crescimento em progressão geométrica;
- velocidade, tanto de produção como de desatualização é vertiginosa;
- variedade, pois não seguem um único padrão e nem fornecem o mesmo tipo de informação;
- veracidade, sendo necessário estabelecer quais são os dados verídicos e que ainda correspondem ao momento atual;
- valor, ou seja, qual a abordagem que será feita desses dados²²³.

À vista disso, a intenção do estudo abarcado por essa pesquisa condiciona-se a um claro

222 CEPEJ, Comissão Europeia para a eficácia da justiça. **Carta Europeia de Ética, sobre o Uso da Inteligência Artificial em Sistemas Judiciais e o seu ambiente**. Estrasburgo, 2018. Disponível em: <https://rm.coe.int/carta-etica-traduzida-paraportugues-revista/168093b7e0>. Acesso em: 22 fev. 2023.

223 JÁCOMO, António. **Saúde e inteligência artificial: uma perspectiva bioética**. Lex Medicinæ, Ano 16, n. 31-32, p. 3-11, 2019.

questionamento: de que modo a Inteligência Artificial poderia auxiliar no combate ao tráfico internacional de crianças? O motivo desse questionamento se dá em razão do argumento, de que se os sistemas inteligentes atuais são projetados para serem aplicados em cenários específicos pré-ditados e escolhidos pelo ser humano em razão de uma necessidade prévia, eles também poderiam ser usados no contexto de mazelas sociais que ainda se colocam sem efetividade em relação ao cunho protetivo e de controle.

Esses sistemas inteligentes podem, muitas vezes, superar especialidades humanas do domínio de aplicação, ou seja, ultrapassar padrões anteriores e que eram tidos como máximos. Exemplo desse ponto se dá com o sistema AlphaZero desenvolvido pelo Google DeepMind (DEEPMIND), inicialmente desenvolvido para aprender o jogo Go (denominado primeiramente como AlphaGo), mas que atualmente é capaz de vencer enxadristas profissionais e times profissionais de e-sports como o StarCraft II (Sistema AlphaStar) demonstrando que esses sistemas inteligentes conseguem alcançar patamares desconhecidos anteriormente²²⁴.

Trazendo essa perspectiva para o foco presente, parte-se a apresentação de quais ferramentas poderão ser utilizadas com o intuito de aumentar a efetividade da segurança e proteção de crianças e adolescentes em relação a uma espécie em singular, o tráfico internacional, compreendendo tratar-se de apenas uma singela parte que merece proteção e cuidado dentre a imensidão de questões existentes.

O intuito primário e desafiador desse estudo se faz por tratar-se de um tema ainda pouco discutido, em relação ao uso da Inteligência Artificial, e improvável, à primeira vista, no que tange a associar seu uso ao combate do tráfico internacional de crianças e adolescentes. Todavia, com o direcionamento apontado pelos caminhos aqui percorridos, começam a tornar-se claros os roteiros a serem percorridos.

Num primeiro momento, impende tracejar quais atitudes o Brasil vinha tomando recentemente, com intuito protetivo, nessa seara. Importante ponto foi a criação de lei 13.812 de 16 de março de 2019, que estabelece a instituição da Política Nacional de Busca de Pessoas Desaparecidas e ainda cria o Cadastro Nacional de Pessoas Desaparecidas, alterando a Lei 8.069, de 13 de julho de 1990 – Estatuto da Criança e do Adolescente, passando a adotar de forma adaptada o método Alerta Amber, em conformidade às realidades enfrentadas em cada estado²²⁵.

224 DA CRUZ, Fabricio Bittencourt et al. ROBÔS SUBSTITUEM JUÍZES? O ESTADO DA ARTE DA INTELIGÊNCIA ARTIFICIAL NO JUDICIÁRIO BRASILEIRO. **Revista Antinomias**, [s. l.], ano 2022, v. 3, n. 1, p. 8-41, jan/jun 2022. Disponível em: <https://antinomias.com.br/index.php/revista/article/view/36>. Acesso em: 15 fev. 2022.

225 BRASIL. Lei nº 13.812, de 16 de março de 2019. **Institui a Política Nacional de Busca de Pessoas Desaparecidas, cria o Cadastro Nacional de Pessoas Desaparecidas e altera a Lei nº 8.069, de 13 de julho de 1990 (Estatuto da Criança e do Adolescente)**. Brasília, 2019. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/lei/l13812.htm#:~:text=Art.,Distrito%20Federal%20e%20aos%20Territ%C3%B3rios. Acesso em: 10 maio. 2023.

Todavia, devido em parte significativa à pandemia de COVID-19, a implantação de medidas acerca do método ficaram suspensas e pouco se falou a respeito no período de 2020 a 2022, de modo que apenas o estado do Rio de Janeiro deu continuidade à aplicação do método renomeado de Alerta PRI²²⁶, buscando o combate a números impressionantes de casos de crianças desaparecidas, tal qual as 172 notificações de casos identificados em 2021 apenas no estado²²⁷. À vista disso, mesmo que de forma prematura e vagarosa, o Brasil tem buscado atrelar novas ferramentas no combate a essa prática, de modo a verificar uma nova relação entre proteção e tecnologia, havendo uma aproximação de novas práticas com questões que já existem há muito tempo.

Nesse sentido, a atuação protetiva estatal, tal qual apresentado pela Delegada da DEPCA (Delegacia de Proteção à Criança e ao Adolescente) de Campo Grande, Mato Grosso do Sul, Anne Karine Trevizan, que dispôs que enquanto criminosos usam inteligência artificial e a internet para cometer delitos diversos, os investigadores que atuam no NIP (Núcleo de Inteligência Policial) da DEPCA, conseguem identificar e prender os envolvidos, com o uso de recursos semelhantes, sendo um trabalho em parceria com forças policiais de todo o Brasil e também do exterior²²⁸.

Observa-se que no contexto jurídico brasileiro, as soluções de IA atuais basicamente são utilizadas na identificação de padrões nos dados buscando, na maioria das vezes, a redução de esforço criado por tarefas repetitivas, caracterizando um uso ingênuo de tamanha força disruptiva. Em outros termos, a força da IA estaria sendo utilizada para “ganhar tempo” e não para evitar questões que ainda assolam a sociedade e que não apresentam soluções consistentes.

2.2. IA cumulada aos mecanismos protetivos de combate ao tráfico de crianças

Partindo aos termos práticos de uso da IA, reforça-se outro ponto importante, relacionado ao Poder Judiciário brasileiro. Esse tem em seu íterim a presença de um grande aliado desse processo, as Varas de Infância e Juventude, varas especializadas em fomentar a proteção de crianças e adolescentes em situação de risco e oferecer atenção especial a eles.

226 O nome faz referência a Priscila Belfort. Ela que desapareceu em 2004, sendo vista pela última vez por uma colega de trabalho, quando saiu para o intervalo de almoço e não retornou mais, deixando a dúvida em acessão sobre seu paradeiro que até hoje não foi identificado, embora já tivesse 28 anos na época do caso, este nome foi escolhido não somente para homenagear a jovem, mas também para atrair a atenção, visto que ela é irmã do ex lutador de MMA renomado Vitor Belfort.

227 RAMOS, Alana Figueiredo. **Inovações nos métodos de busca de crianças e adolescentes desaparecidos: adaptação do alerta amber na legislação brasileira implantada pelo adendo da lei 13.812 de 2019**. Orientador: Camila Damasceno de Andrade. 2022. 56 p. Trabalho de Conclusão de Curso (Bacharelado em Direito) - Universidade do Sul de Santa Catarina, Tubarão, 2022. p.41. Disponível em: <https://repositorio.animaeducacao.com.br/handle/ANIMA/26868>. Acesso em: 23 maio 2023.

228 **Tecnologia auxilia investigações para proteção de crianças e adolescentes em MS**. A Crítica. 2023. Disponível em: <https://www.acritica.net/editorias/geral/tecnologia-auxilia-investigacoes-para-protecao-de-criancas-e/662328/>. Acesso em: 15 jun. 2023.

Pois bem, a partir de agora, essa atenção se voltará a analisar possíveis situações de risco que possam evidenciar alguns problemas que estejam associados ao tráfico internacional, buscando, ao mesmo tempo, direcionar a aplicabilidade da inteligência artificial.

Nessa seara, destaca-se a ocorrência de casos em que crianças e adolescentes encontram-se em paradeiro incerto, não sabido, desconhecido, por motivos múltiplos, podendo permanecer, inclusive, por tempo indefinido. Essa situação pode caracterizar uma realidade em que crianças e adolescentes passem a estar em risco, podendo demonstrar, de modo camuflado, a ocorrência de um rapto não reportado, ou de uma ausência intencional que desencadeou em um desaparecimento, ou em qualquer outra resposta à situação de risco que provavelmente já estaria sendo acompanhada anteriormente e que não teve o alerta necessário.

Isto é, tem-se nas Varas de Infância e Juventude um acompanhamento de casos de risco que muitas vezes pode vir a ser arquivado por ausência de informações ou pelo tempo transcorrido, sendo que o simples transcurso do tempo sem informação já seria um dado que pudesse ser computado.

Visando essa questão, levanta-se a possibilidade de gerar um intercâmbio informacional propiciado pela Inteligência Artificial capaz de conectar informações, dados sigilosos e situações num campo informacional de combate ao tráfico internacional. Explicando, à IA comportaria a análise de dados que complementaríamos as informações atinentes ao combate do tráfico internacional de crianças e adolescentes, desde a apresentação de situação de risco, o paradeiro desconhecido, informações registrais de nome, parentesco, identificação, e tantos outros dados aos quais a IA teria acesso tendo contato com as informações existentes em Varas de Infância e Juventude distribuídas pelo País.

Dessa forma, havendo respeito aos princípios éticos de uso da IA já demonstrados anteriormente, ela estaria apta a consultar informações abarcadas por todos os graus de sigilo, permitindo a análise dessas informações e gerando uma base de dados que poderá ser acessada quando necessário for e pelas autoridades apropriadas com vistas a evitar a ocorrência dessa prática. Salienta-se que a necessidade do uso desses dados estaria norteadas pelo próprio artigo 14, § 3º²²⁹, da LGPD²³⁰, em que há autorização para a coleta e tratamento dos dados pessoais de crianças, mesmo que sem o consentimento, se a finalidade for para a sua proteção.

229 Art. 14. O tratamento de dados pessoais de crianças e de adolescentes deverá ser realizado em seu melhor interesse, nos termos deste artigo e da legislação pertinente.

§ 3º Poderão ser coletados dados pessoais de crianças sem o consentimento a que se refere o § 1º deste artigo quando a coleta for necessária para contatar os pais ou o responsável legal, utilizados uma única vez e sem armazenamento, ou para sua proteção, e em nenhum caso poderão ser repassados a terceiro sem o consentimento de que trata o § 1º deste artigo.

230 BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 10 fev. 2023.

Uma situação hipotética para esclarecer sua aplicação: criança acompanhada em determinada Vara da Infância e Juventude por apresentar demandas protetivas e situação de risco no âmbito familiar; as informações familiares divergem acerca do paradeiro da criança; não há qualquer informação consistente acerca dos motivos que se deram seu paradeiro incerto, apenas relatos vagos de onde estaria; a partir desse momento a IA já pode acessar todas as informações relevantes sobre os dados dessa criança, sobre a situação em que estava inserida e pode emitir alertas sobre o desconhecimento de seu paradeiro, com vistas a contatar uma rede informacional que não dependeria da atuação manual e mecânica do ser humano, mas que passaria automaticamente por uma inteligência pré-disposta a gerenciar unicamente essa questão; a partir disso, sistemas de localização estariam correlacionados e conectados, desde outros órgãos do judiciário, aeroportos, redes de proteção, ONG's, centros de transporte e passagem, e demais meios pelos quais o tráfico se efetiva; assim, a transmissão de dados permaneceria resguardada, tendo em vista o cuidado que necessita e os agentes a que ela fará voz, bem como propiciaria uma convergência informacional em níveis ainda desconhecidos.

Tratando-se de modelo apenas idealizado até o momento, faz-se importante demonstrá-lo em etapas de funcionamento para melhor visualização:

1. **Acesso legal aos processos:** É necessário garantir que a leitura dos processos seja realizada de acordo com as leis de proteção de dados e privacidade, bem como com as regulamentações específicas relacionadas aos processos em sigilo. É importante obter autorização e colaboração das autoridades competentes.
2. **Coleta de dados:** Os processos em sigilo contêm informações sensíveis sobre as crianças e suas famílias. É fundamental garantir que os dados sejam coletados e armazenados de forma segura e em conformidade com as regulamentações de proteção de dados, preservando a privacidade das partes envolvidas. Ato contínuo, cumpre salientar a presença da LGPD em meio a esse fornecimento de dados, a qual já estabelece dispositivos para a proteção de dados pessoais de crianças e adolescentes. O artigo 14 da LGPD²³¹ formaliza que o processamento de dados pessoais de crianças e adolescentes deve ser realizado no melhor interesse deles, garantindo seus direitos à privacidade, proteção de dados pessoais e imagem, exigindo, também, que o processamento de dados pessoais de crianças e adolescentes deva ser transparente²³².
3. **Desenvolvimento do modelo de IA:** É necessário treinar um modelo

231 Art. 14. O tratamento de dados pessoais de crianças e de adolescentes deverá ser realizado em seu melhor interesse, nos termos deste artigo e da legislação pertinente.

232 BOTELHO, Marcos. **A LGPD e a proteção ao tratamento de dados pessoais de crianças e adolescentes**. *Indep. Políticas Públicas*, v. 9, n. 1, p. 1-16, 2020. Disponível em: <http://www.unifafibe.com.br/revista/index.php/indp/article/view/1026>. Acesso em: 12 jun 2023.

de inteligência artificial capaz de analisar os documentos e identificar possíveis indicadores de risco ou desaparecimento de crianças. Para isso, você precisará de uma grande quantidade de dados de treinamento anotados corretamente por profissionais especializados, como assistentes sociais e advogados especializados em direitos das crianças. Nessa etapa, requer-se uma avaliação do modelo, de modo a usar o conjunto de teste para avaliar o desempenho final do modelo a ser criado, envolvendo o cálculo de métricas a serem selecionadas conforme a relevância para a tarefa em questão, desde precisão, *recall*, F1-score, acurácia e curva ROC²³³.

4. **Identificação de padrões:** O modelo de IA deve ser capaz de identificar padrões relevantes nos processos, como informações sobre negligência, abuso, violência doméstica ou qualquer outra circunstância que possa indicar um risco para a criança. Além disso, ele pode buscar informações relacionadas a possíveis destinos de tráfico internacional de crianças.
5. **Alertas e notificações:** Quando o modelo identificar uma possível situação de risco ou desaparecimento, ele deve gerar alertas para os profissionais responsáveis, como assistentes sociais, promotores de justiça ou autoridades policiais. Essas notificações permitirão que as medidas apropriadas sejam tomadas para proteger a criança e investigar a situação.
6. **Monitoramento contínuo:** A inteligência artificial pode ser usada para monitorar continuamente os processos e fornecer atualizações em tempo real sobre possíveis casos de risco. Isso ajudará as autoridades a agir prontamente para proteger as crianças envolvidas.

Gize-se ressaltar a necessidade de assistência mútua internacional de cada país desde o fornecimento de dados e informações, desenvolvimento de políticas combatíveis, disseminação de conhecimento sobre o tema, e com recorte especial na presente pesquisa, a utilização de mecanismos modernos que se desenvolveram consoante aos avanços da humanidade e que viabilizaram uma nova vertente de atuação²³⁴.

Destarte, o combate ao tráfico de pessoas tenderá a destruir impérios econômicos que se favorecem dessa prática, isto é, destruir estatísticas que apontam quase um terço do total

233 MARIANO, Diego. **Métricas de avaliação em machine learning:** aurária, sensibilidade, precisão, especificidade, f-score e curva ROC. 2021. Disponível em: <https://diegomariano.com/metricas-de-avaliacao-em-machine-learning/>. Acesso em: 12 abr. 2023.

234 GIRONI, Marcela Caroline Vaz. OS MECANISMOS DE REPRESSÃO AO TRÁFICO DE PESSOAS. In: Brasil. **Ministério Público Federal. 2º Câmara de Coordenação e Revisão. Tráfico de pessoas / 2ª Câmara de Coordenação e Revisão, Criminal;** organização: Stella Fátima Scampini. – Brasília: MPF, 2017. p. 52-100. Coletânea de artigos Tráfico de Pessoas; v. 2. Disponível em: <https://www.mpf.mp.br/atuacao-tematica/ccr2/publicacoes/coletaneas-de-artigos>. Acesso em: 14 jan. 2023.

das vítimas de tráfico de pessoas no mundo sendo meninos e meninas, de acordo com o Relatório Global sobre o Tráfico de Pessoas 2016, lançado pelo Escritório das Nações Unidas sobre Drogas e Crime (UNODC)²³⁵. Isso caracteriza a dificuldade que ainda está por vir, tendo em vista tratar-se de giro monetário de grandes proporções, todavia, isso apenas demonstra a real insignificância que se deu ao combate do tráfico de pessoas por tanto tempo, permitindo que tomasse essas proporções, tratando o ser humano como mercadoria e garantindo que sua comercialização fosse algo comercialmente atrativo.

Logo, deve-se frisar a importância em atrelar proteção, segurança e direitos, ao passo que diante da interconexão entre essas tecnologias de inteligência artificial e os diversos direitos abarcados nessa demanda, propõe-se a atenção às normas de direitos humanos como uma forma de evitar o fenômeno que a literatura de regulação da inteligência artificial designa como *ethics washing*: o compromisso pró-forma das empresas e governos com conjuntos de princípios de inteligência artificial que sejam vagos a ponto de não imporem restrições substanciais à sua conduta, devendo ser direcionados por todas aquelas bases éticas principiadas anteriormente²³⁶.

Outrossim, o Marco Civil da Internet também se mostra relevante no presente contexto, porém a regulação dos dados pessoais foi abordada de maneira mais aprofundada pela Lei Geral de Proteção de Dados. No tocante ao tema, é importante ressaltar que o projeto inicial do Marco da Inteligência Artificial, designado como PL 240/2020²³⁷, aponta diretamente a necessidade de regulação, principalmente, considerando que o aprendizado de máquina perpetuará uma cultura estatística preconceituosa enraizada ao conglomerado de informações dos quais a IA se utilizará, representando uma sociedade enviesada que direcionará seus vieses a uma máquina, de modo a ser pungente a regulamentação quanto ao uso da IA, a coleta de dados que será feita, e a disponibilização desses.

Nesse sentido, diante do vácuo legislativo ainda muito presente em relação ao assunto, tanto a Resolução 332/2020 do Conselho Nacional de Justiça²³⁸, que dispõe sobre a ética, transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário, quanto os marcos regulatórios e leis apresentadas são importantes para uma nova realidade do Poder Judiciário. Porém, não se elimina a necessidade de

235 UNODC. **United Nations Convention against Organized Crime celebrates 10 years**. Disponível em: <https://www.unodc.org/unodc/en/frontpage/2013/September/un-convention-against-organized-crime-celebrates-10-years.html>. Acesso em: 20 maio. 2023.

236 ALMADA, Marco Antonio Lasmar. **Considerações sobre inteligência artificial e direito internacional**. 20 jul 2020. Disponível em: https://www.researchgate.net/publication/343083348_Consideracoes_sobre_inteligencia_artificial_e_direito_internacional. Acesso em: 10 jan 2022.

237 BRASIL. Câmara dos Deputados. Projeto de Lei nº 240, de 11 de fevereiro de 2020. Cria a Lei da Inteligência Artificial, e dá outras providências. Brasília: Câmara dos Deputados, 2020. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2236943>. Acesso em: 15 maio. 2023.

238 CONSELHO NACIONAL DE JUSTIÇA. **Resolução nº 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 Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3429>. Acesso em: 10 maio 2023.

um diálogo institucional com o CNJ, que perpassa os fóruns acadêmicos adequados e até o papel do Congresso, no que diz respeito a eventuais futuras reformas das legislações vigentes.

E a partir disso, buscar, nas palavras de Isabela Ferrari na obra *Justiça Digital*, uma “oportunidade de escapar às limitações trazidas pela estrutura do Poder Judiciário”²³⁹.

3. CONSIDERAÇÕES FINAIS

Pode-se confirmar que atualmente os desafios diários que os seres humanos enfrentam superam às 24 horas disponíveis em um dia, até mesmo porque trata-se de seres dotados de necessidades básicas que vão além do trabalho e da prática de atividades, acabando por limitar a quantidade de afazeres suportáveis. Essa realidade humana não se aplica às máquinas, as quais não necessitam parar, não interrompem sua atuação e podem multiplicar-se em todos os locais, garantindo uma funcionalidade mais ampla à necessidade apresentada pelo combate ao tráfico internacional de crianças, suportando uma intercomunicabilidade de dados constante e não apenas temporária.

Como demonstrado, as questões envolvendo o tráfico internacional de crianças e adolescentes ainda apresenta grandes hiatos que carecem de apoio, suporte e combate. Nesse mesmo sentido foram pontuados instrumentos de normatização e legislação que rodeiam o tema, mas que ainda requerem aperfeiçoamento quanto à sua aplicação prática e realística.

Ademais, pode-se perceber a quantidade de recentes mecanismos e meios que começaram a surgir com o advento da IA, confirmando que outras possibilidades poderiam ser cogitadas. Assim, foi verificado que a aplicação da IA cumulada ao Direito acaba desencadeando muitos questionamentos e divergência de opiniões quanto aos seus riscos, entretanto, através de todo o exposto pode-se conferir que as vantagens se mostram superiores, sendo que as desvantagens e riscos acabam sendo mitigadas com as adaptações possíveis frente a essa nova realidade.

O uso da IA como uma ferramenta associada ao direito protetivo de crianças e adolescentes é capaz de movimentar-se no intento de combater esse crime, de modo a associar-se aos direitos basilares dessas vítimas, bem como nortear-se de modo ético e principiológico para atuar de forma segura.

O principal ponto que se observou diz respeito à carência por respostas, por pesquisa e por fomento a esse novo ramo. O uso da IA ainda se mostra muito singelo, deficiente quanto às diversas abordagens que pode ter, caracterizando um desperdício de oportunidades, transparecendo os recortes que demonstram os motivos de crimes como esse ainda se

perpetrarem por tanto tempo.

Portanto, o que foi buscado como objetivo máximo do presente estudo se relaciona intimamente com o interesse em fomentar a pesquisa, o debate, novos questionamentos, o alcance de respostas e, mais que isso, dar abertura ao novo, permitir que não haja medo do diferente, mas que este influencie pensamentos distintos, aplicações modernas para solucionar o que já é antigo.

Ademais, no contexto de uma nova perspectiva, à medida que a temática ganhe maior abrangência, haverá uma transformação na mentalidade dos profissionais do Direito e da comunidade jurídica como um todo. Nesse sentido, é possível que as instituições de ensino superior incluam em seus currículos disciplinas que abordem essa evolução do Direito, a fim de capacitar seus estudantes e prepará-los para uma aplicação eficaz desse novo arcabouço jurídico, priorizando, sobretudo, a ética.

Assim, a proteção integral desses sujeitos passaria pela garantia de que todos os meios que possam ser usados a fim de protegê-los estão sendo empregados. Logo, tratando-se de proteção à criança e ao adolescente, como elucidado pela própria Corte Internacional de Justiça, insta compreender que se trata de um valor universal, que confere responsabilidade erga omnes, mantendo as expectativas de que haverá a união de forças nesse combate.

Por fim, salienta-se que não foi o intuito durante o estudo o exaurimento da temática ou a apresentação de uma solução infalível e inequívoca, mas ao contrário, foi apresentado um modelo que pode desencadear diversos questionamentos, e a partir desses construir algo que efetive o interesse primário, combater uma mazela social que caminha junto ao desenvolvimento da humanidade.

4. Referências

AGÊNCIA SENADO. **Trata-se de estimativa do Conselho Federal de Medicina. Desaparecimento de crianças: Conselho de Medicina faz alerta na Comissão de Direitos Humanos.** Disponível em: <https://www12.senado.leg.br/noticias/materias/2018/04/06/desaparecimento-de-criancas-conselho-de-medicina-faz-alerta-na-comissao-de-direitos-humanos>. Acesso em: 10 jan. 2023.

ALMADA, Marco Antonio Lasmar. **Considerações sobre inteligência artificial e direito internacional.** 20 jul 2020. Disponível em: https://www.researchgate.net/publication/343083348_Consideracoes_sobre_inteligencia_artificial_e_direito_internacional. Acesso em: 10 jan 2022.

BOTELHO, Marcos. **A LGPD e a proteção ao tratamento de dados pessoais de crianças e adolescentes.** Indep. Políticas Públicas, v. 9, n. 1, p. 1-16, 2020. Disponível em: <http://>

www.unifafibe.com.br/revista/index.php/indp/article/view/1026. Acesso em: 12 jun 2023.

BRAGANÇA, Fernanda; BRAGANÇA, Laurinda Fátima da F. P. G. **Revolução 4.0 no Poder Judiciário: LEVANTAMENTO DO USO DE INTELIGÊNCIA ARTIFICIAL NOS TRIBUNAIS BRASILEIROS**. Revista da Seção Judiciária do Rio de Janeiro. Rio de Janeiro, v. 23, n. 46, p. 65-76, jul./out. 2019.

BRASIL. Câmara dos Deputados. Projeto de Lei nº 240, de 11 de fevereiro de 2020. Cria a Lei da Inteligência Artificial, e dá outras providências. Brasília: Câmara dos Deputados, 2020. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2236943>. Acesso em: 15 maio. 2023.

BRASIL. Governo Federal. **Relatório sobre o Tráfico de Pessoas 2022 – Brasil**. Brasília, DF, 2022. Disponível em: <https://br.usembassy.gov/pt/relatorio-sobre-o-trafico-de-pessoas-2022-brasil/>. Acesso em: 10 abr. 2023.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. **Lei Geral de Proteção de Dados Pessoais (LGPD)**. Brasília, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 10 fev. 2023.

BRASIL. Lei nº 13.812, de 16 de março de 2019. **Institui a Política Nacional de Busca de Pessoas Desaparecidas, cria o Cadastro Nacional de Pessoas Desaparecidas e altera a Lei nº 8.069, de 13 de julho de 1990 (Estatuto da Criança e do Adolescente)**. Brasília, 2019. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/lei/l13812.htm#:~:text=Art.,Distrito%20Federal%20e%20aos%20Territ%C3%B3rios. Acesso em: 10 maio. 2023.

BRYSON, Joanna J. **The Artificial Intelligence of the Ethics of Artificial Intelligence: An Introductory Overview for Law and Regulation**. In: DUBBER, Markus D.; PASQUALE, Frank; DAS, Sunit (Ed.). *The Oxford Handbook of Ethics of Artificial Intelligence*. Oxford: Oxford University Press, 2020.

CEPEJ, Comissão Europeia para a eficácia da justiça. **Carta Europeia de Ética, sobre o Uso da Inteligência Artificial em Sistemas Judiciais e o seu ambiente**. Estrasburgo, 2018. Disponível em: <https://rm.coe.int/carta-etica-traduzida-paraportugues-revista/168093b7e0>. Acesso em: 22 fev. 2023.

CONSELHO NACIONAL DE JUSTIÇA. Resolução nº 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 Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3429>. Acesso em: 10 maio 2023.

DA CRUZ, Fabricio Bittencourt et al. **ROBÔS SUBSTITUEM JUÍZES? O ESTADO DA ARTE**

DA INTELIGÊNCIA ARTIFICIAL NO JUDICIÁRIO BRASILEIRO. **Revista Antinomias**, [s. l.], ano 2022, v. 3, n. 1, p. 8-41, jan/jun 2022. Disponível em: <https://antinomias.com.br/index.php/revista/article/view/36>. Acesso em: 15 fev. 2022.

FERRARI, Isabela (coord.). et al. *Justiça Digital*. 1. Ed. São Paulo: Thomson Reuters Brasil, 2020.

FERREIRA, Micaela Amorim; BORGES, Paulo Cesar Corrêa. **TRÁFICO DE PESSOAS COMO PROBLEMA INTERNACIONAL E PANORAMA LEGISLATIVO DE COMBATE**. In: Brasil. Ministério Público Federal. 2º Câmara de Coordenação e Revisão. *Tráfico de pessoas / 2ª Câmara de Coordenação e Revisão, Criminal*; organização: Stella Fátima Scampini. – Brasília: MPF, 2017. p. 20-39. Coletânea de artigos Tráfico de Pessoas; v. 2. Disponível em: <https://www.mpf.mp.br/atuacao-tematica/ccr2/publicacoes/coletaneas-de-artigos>. Acesso em: 14 jan. 2023.

FIGARO-GARCIA, Claudia. **Uma proposta de prática psicológica para casos de desaparecimento de crianças e adolescentes**. Orientador: Prof. Dra. Eliana Herzberg. 2010. 249 p. Dissertação (Doutorado em Psicologia) - Universidade de São Paulo, São Paulo, 2010. DOI 10.11606/T.47.2010.tde-26072010-123243. Disponível em: <https://www.teses.usp.br/teses/disponiveis/47/47133/tde-26072010-123243/pt-br.php>. Acesso em: 23 maio 2023.

GIRONI, Marcela Caroline Vaz. OS MECANISMOS DE REPRESSÃO AO TRÁFICO DE PESSOAS. In: Brasil. **Ministério Público Federal. 2º Câmara de Coordenação e Revisão. Tráfico de pessoas / 2ª Câmara de Coordenação e Revisão, Criminal**; organização: Stella Fátima Scampini. – Brasília: MPF, 2017. p. 52-100. Coletânea de artigos Tráfico de Pessoas; v. 2. Disponível em: <https://www.mpf.mp.br/atuacao-tematica/ccr2/publicacoes/coletaneas-de-artigos>. Acesso em: 14 jan. 2023.

JÁCOMO, António. **Saúde e inteligência artificial: uma perspectiva bioética**. *Lex Medicinæ*, Ano 16, n. 31-32, p. 3-11, 2019.

KELLY, Emma. ECPAT UK. **Bordering on Concern: Child Trafficking in Wales**. Copyright ECPAT. Londres, 2009.

MARIANO, Diego. **Métricas de avaliação em machine learning**: aurária, sensibilidade, precisão, especificidade, f-score e curva ROC. 2021. Disponível em: <https://diegomariano.com/metricas-de-avaliacao-em-machine-learning/>. Acesso em: 12 abr. 2023.

RAMOS, Alana Figueiredo. **Inovações nos métodos de busca de crianças e adolescentes desaparecidos: adaptação do alerta amber na legislação brasileira implantada pelo adendo da lei 13.812 de 2019**. Orientador: Camila Damasceno de Andrade. 2022. 56 p. Trabalho de Conclusão de Curso (Bacharelado em Direito) - Universidade do Sul de Santa Catarina, Tubarão, 2022. Disponível em: <https://repositorio.animaeducacao.com>.

[br/handle/ANIMA/26868](https://br.handle/ANIMA/26868). Acesso em: 23 maio 2023.

RIBEIRO, Renata. **O TRÁFICO INTERNACIONAL DE CRIANÇAS E ADOLESCENTES: POLÍTICAS PÚBLICAS DE PREVENÇÃO E ENFRENTAMENTO NO BRASIL**. 2019, 61f. Monografia. (Curso Bacharelado em Direito). Universidade de Santa Cruz do Sul, 2019.

Simpson, J. (ed.). (2017). **Oxford English Dictionary** (3^a ed.). Oxford: Oxford University Press.

Tecnologia auxilia investigações para proteção de crianças e adolescentes em MS. A Crítica. 2023. Disponível em: <https://www.acritica.net/editorias/geral/tecnologia-auxilia-investigacoes-para-protecao-de-criancas-e/662328/>. Acesso em: 15 jun. 2023.

UNODC - UNITED NATIONS OFFICE ON DRUGS AND CRIME. Tráfico de Pessoas e Contrabando de Migrantes. Disponível em: <https://www.unodc.org/lpo-brazil/pt/trafico-de-pessoas/index.html#:~:text=O%20tr%C3%A1fico%20de%20pessoas%20%C3%A9,receber%20pagamentos%20ou%20benef%C3%ADcios%20para>. Acesso em: 17 maio 2023.

UNODC. **United Nations Convention against Organized Crime celebrates 10 years**. Disponível em: <https://www.unodc.org/unodc/en/frontpage/2013/September/un-convention-against-organized-crime-celebrates-10-years.html>. Acesso em: 20 maio. 2023.

PARÂMETROS REGULATÓRIOS DO USO DA INTELIGÊNCIA ARTIFICIAL NA SEGURANÇA PÚBLICA: UMA ANÁLISE ENTRE BRASIL E UNIÃO EUROPEIA

REGULATORY PARAMETERS FOR THE USE OF ARTIFICIAL INTELLIGENCE IN PUBLIC SECURITY: AN ANALYSIS BETWEEN BRAZIL AND THE EUROPEAN UNION

Juliana Roman²⁴⁰

Rafaela Ferreira²⁴¹

Victor Vieira²⁴²

Resumo

O desenvolvimento de tecnologias de Inteligência Artificial (IA) para segurança pública tem gerado discussões regulatórias a nível internacional e nacional, constando na agenda de diversos países. Diante desse cenário, objetiva-se investigar no presente trabalho quais os parâmetros regulatórios do uso da IA no Brasil e na União Europeia (UE). A fim de verificar se existem regulações voltadas ao uso das tecnologias que envolvem IA na segurança pública, propõe-se análise micro-comparativa sobre o estado da arte do tema. A hipótese principal foi que encontrar-se-ia pontos de convergência e divergência diante do tratamento da matéria em ambos sistemas jurídicos. A presente pesquisa utilizou-se do método indutivo e, como fonte de pesquisa, investigou o tema em materiais bibliográficos, jurisprudência e legislações, nacionais e estrangeiras. Após análise dos diferentes cenários regulatórios, foi possível construir um panorama da regulação de IA em ambas as regiões, viabilizando a constatação de uma diferença significativa na

240 É pesquisadora e líder de projetos relacionados à proteção de dados pessoais no IRIS. Bacharel em Direito pela PUCRS. Mestre em Direito pela UFRGS. Especialista em Direito Digital pela FMP. Especialista em Direito do Consumidor pela Universidade de Coimbra (UC).

241 É graduada em Direito pela Universidade Federal da Bahia (UFBA). Pesquisadora no Instituto de Referência em Internet e Sociedade (IRIS), líder de projeto sobre moderação de conteúdo.

242 É bacharel em Direito pela UFMG e pós-graduando em Direito da Proteção e Uso de Dados na Puc Minas. Pesquisador e encarregado de proteção de dados pessoais no Instituto de Referência em Internet e Sociedade (IRIS). Certified Information Privacy Professional - Europe (CIPP/E) pela IAPP.

robustez dos direitos e garantias reconhecidas aos cidadãos no Brasil e na UE, ainda que não haja legislação específica vigente sobre a temática em ambos.

Palavras-chave: Regulação; Segurança Pública; Inteligência Artificial; Viés Algorítmico.

Abstract

The development of Artificial Intelligence (AI) technologies for public safety has generated regulatory discussions at an international and national level, appearing on the regulatory agenda of several countries. Given this scenario, the objective of this work is to investigate the regulatory parameters for the use of AI in Brazil and in the European Union (EU). In order to verify if there are regulations aimed at the use of technologies involving AI in public safety, a micro-comparative analysis is proposed regarding the state of the art of the theme. The main hypothesis was that points of convergence and divergence will be found in the treatment of the matter in both legal systems. The present research used the deductive method and, as a source of research, investigated the theme in national and foreign bibliographical materials, jurisprudence, and legislation. After analyzing the different regulatory scenarios, it was possible to construct an overview of AI regulation in both regions, enabling the verification of a significant difference in the robustness of the rights and guarantees recognized to citizens in Brazil and in the European Union, even though there is no specific current legislation on the subject in any aforementioned regions.

Keywords: Regulation; Public Security; Artificial intelligence; Algorithmic Bias.

1. INTRODUÇÃO

1.1. Contexto geral

Recentemente, o desenvolvimento de tecnologias baseadas em algoritmos de Inteligência Artificial (IA) tem repercutido no surgimento de uma vasta gama de aplicações práticas para ferramentas dessa natureza. Alguns exemplos variam desde a criação de programas geradores de imagens, como o *OpenArt* e o *HotPot*, até o uso dessas técnicas para desenvolvimento de redes neurais que viabilizam modelos de linguagem, como o *ChatGPT* desenvolvido pela OpenAI.

A inteligência artificial pode ser utilizada para possibilitar diversas funcionalidades em aplicações digitais e de internet, tais como a implementação de modelos que permitem a interpretação e o processamento de linguagem natural, a fim de possibilitar uma interação com a máquina que se assemelha à comunicação com uma pessoa. Além disso, a IA pode ser implementada conjuntamente a bancos de dados para possibilitar

o rápido processamento de informações e a geração de inferências e descobertas com base nos dados analisados.

Esta última implementação, em especial, traz, por sua vez, repercussões que devem ser estudadas cautelosamente, dados os impactos que podem gerar. Tecnologias baseadas em IA têm também sido cogitadas e implementadas pelo Poder Público para finalidades ligadas a uma suposta melhora nos sistemas de manutenção da segurança pública. Soluções dessa natureza, contudo, trazem repercussões sociais adicionais, relacionadas ao fortalecimento exacerbado do aparato estatal de vigilância e persecução criminal. Aplicações de IA com esse objetivo – o de fomentar o aparato estatal em prol de um alegado ganho para a segurança pública – representam algumas das soluções que mais têm angariado críticas e preocupação por parte de diversos setores da sociedade. Pode-se citar, a título de exemplo dessas tecnologias, o uso de aplicações de IA para cálculo da probabilidade de reincidência criminal por parte de indivíduos já previamente condenados por atividades ilícitas²⁴³, para técnicas de policiamento preditivo e direcionamento de tropas policiais²⁴⁴, entre diversos outros.

O uso indiscriminado de tecnologias voltadas à IA gera questionamentos quanto à possibilidade de perpetuar e fortalecer práticas discriminatórias, bem como outras formas de injustiça social – dinâmica esta que se intensifica em um país como o Brasil, fortemente afetado por uma realidade de racismo estrutural e de forte segregação socioeconômica. Nesse sentido, o desenvolvimento de tecnologias de IA para segurança pública tem gerado discussões regulatórias a nível internacional e nacional, constando na agenda de diversos países. A UE, por exemplo, tem adotado uma série de medidas para regular o uso da IA em seu território. O Brasil, por sua vez, recentemente presenciou o surgimento de debates em âmbito legislativo para regular o uso de ferramentas baseadas em IA.

Ressalta-se, de antemão, que a regulação de aplicações tecnológicas dessa natureza mostra-se imperativa para que haja a interposição de limites e salvaguardas legais para o uso da inteligência artificial. A positivação da matéria, contudo, deve ser realizada de maneira abrangente e contar com aportes técnicos de setores da sociedade que compreendem o funcionamento dessa tecnologia e suas possíveis implicações para a garantia de direitos fundamentais. Nesse sentido, importa ressaltar a importância de que parâmetros regulatórios para o uso ético de inteligência artificial não podem restringir-se ao emprego dessa tecnologia para fins privados, mas devem aplicar-se também à atividade estatal, a fim de evitar o emprego da inteligência artificial para a perpetuação de práticas repressivas pelo Poder Público.

243 VIEIRA, Leonardo Marques. **A problemática da inteligência artificial e dos vieses algorítmicos: caso COMPAS**. In: 2019 Brazilian Technology Symposium. [S.l.], 2019. Disponível em: <https://lcv.fee.unicamp.br/images/BTSym-19/Papers/090.pdf>. Acesso em: 13 jul. 2023.

244 PREFEITURA DE SÃO PAULO. **Nova Plataforma de Videomonitoramento Smart Sampa**. São Paulo, 2023. Disponível em: <https://participemais.prefeitura.sp.gov.br/legislation/processes/209>. Acesso em: 13 jul. 2023.

1.2. Definições conceituais relevantes

Primeiramente, entende-se como necessário definir o que aqui se entende como “segurança pública” e “inteligência artificial”. Trata-se dos temas centrais do presente estudo, motivo pelo qual se mostra essencial o apontamento de uma conceituação precisa como ponto de partida para a análise.

Entende-se “segurança pública” a partir da definição apresentada pela Carta de Adesão aos Princípios do Fórum Brasileiro de Segurança Pública (FBSP)²⁴⁵. Segundo o documento, a segurança pública constitui “um serviço público, baseado na prevenção e na repressão qualificada, com respeito à equidade, à dignidade humana e guiado pelo primado dos Direitos Humanos e do Estado democrático de Direito”. Importa ressaltar o caráter de respeito e observância de critérios de equidade, dignidade humana e respeito aos Direitos Humanos e ao Estado democrático de Direito na definição fornecida pela organização.

“Inteligência artificial”, por sua vez, é compreendida à luz da definição adotada no documento de apresentação da Estratégia Brasileira de Inteligência Artificial (EBIA)²⁴⁶. Este, por sua vez, reitera a conceituação proposta pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE) em suas Recomendações do Conselho sobre Inteligência Artificial. No documento, apresenta-se a inteligência artificial como “um sistema baseado em máquina que pode, para um determinado conjunto de objetivos definidos pelo homem, fazer previsões, recomendações ou tomar decisões que influenciam ambientes reais ou virtuais”. Em resumo, trata-se de um sistema computacional que pode ter um grau variável de autonomia (desde totalmente autônomo até consideravelmente dependente da tomada de decisões humana) e que busca, por meios algorítmicos, assimilar informações constantes em bases de dados e tomar decisões a partir de inferências providenciadas por esse processo prévio de assimilação de dados.

Um outro fator importante nesta definição diz respeito ao reconhecimento da participação e influência humana nos diversos estágios de criação e operação de tecnologias baseadas em IA. Trata-se esta de uma característica fundamental para refutar as alegações relativas a uma suposta “neutralidade” das tecnologias baseadas em inteligência artificial: a geração de inferências e a tomada de decisões não se tornam neutras e objetivas por serem processadas por um sistema computacional, haja vista que a cadeia de comandos que movimenta o algoritmo (o código) e mesmo as informações que alimentam o sistema (os *inputs* submetidos à IA) apresentam fatores humanos que perpetuam vieses, imperfeições e visões de mundo que, inerentemente, afetam o funcionamento de todo o algoritmo.

245 FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA. **Carta de adesão aos princípios do Fórum Brasileiro de Segurança Pública**. Disponível em: https://forumseguranca.org.br/storage/Carta_de_principios_-_FBSP.pdf. Acesso em: 13 jul. 2023.

246 MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES. **Estratégia Brasileira de Inteligência Artificial**. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-documento_referencia_4-979_2021.pdf. Acesso em: 13 jul. 2023. P. 9.

1.3. Metodologia

O presente estudo buscou verificar se existem regulações voltadas ao uso de tecnologias baseadas em IA na segurança pública – e, se sim, quais essas regulações. A pergunta orientadora do trabalho foi: quais os parâmetros regulatórios do uso da IA na Segurança Pública no Brasil e na União Europeia?

Para tanto, utilizou-se a técnica da análise micro-comparativa^{247 248} entre o estado da arte do tema no Brasil e na União Europeia (UE) a nível regulatório. Por análise micro-comparativa, especifica-se que a comparação realizada aprofundou-se em uma análise de como sistemas jurídicos distintos abordam um problema em comum a nível de comandos normativos propriamente ditos – propostos, em tramitação legislativa ou já positivados. Uma análise macro-comparativa, alternativamente, centraria-se na comparação entre os “estilos jurídicos” implementados e nos valores sociais priorizados em cada um desses sistemas jurídicos.

A escolha de pesquisar sobre a temática da IA justifica-se uma vez que o tema contempla repercussões jurídicas, econômicas e políticas que aparecem na pauta de diversas agendas internacionais. Utilizou-se do método indutivo e, como fonte de pesquisa, foram consultados qualitativamente artigos científicos, livros e publicações de organizações nacionais e internacionais cuja atuação se mostra relevante ao estudo proposto, bem como fontes jurisprudenciais e legislativas, tanto nacionais quanto estrangeiras.

1.4. Adendo sobre a comparação entre Brasil e UE

Importa salientar que a escolha por uma comparação entre os avanços brasileiros e europeus na regulação do tema não se deu sob o pretexto de uma relação colonial de admiração incontestada ao modelo regulatório da União Europeia. Ainda, ressalta-se que o objetivo do método comparativo aqui utilizado não propõe uma “importação” de um sistema jurídico estrangeiro e suas normativas para o Direito brasileiro. Contrariamente, a comparação realizada no presente estudo justifica-se por dois fatores centrais.

O primeiro deles diz respeito ao conceito de Transconstitucionalismo proposto por Marcelo Neves²⁴⁹, que consiste na ideia de que ordenamentos jurídicos distintos e autônomos tratam – concomitantemente – de problemas constitucionais comuns entre si. Essas questões em comum, nesse sentido, superam as fronteiras de uma jurisdição específica para tornarem-se problemas “transconstitucionais”. Trata-se de um fenômeno vivenciado com crescente frequência, tendo-se em vista principalmente as repercussões jurídico-

247 VICENTE, Dário Moura. **Direito Comparado**. v. 1: Introdução e Parte Geral. Almedina: Coimbra, 2011.

248 MARCONI, Marina; LAKATOS, Eva. **Metodologia científica**. São Paulo: Editora Atlas, 2004.

249 NEVES, Marcelo. **Transconstitucionalismo**. 2009. São Paulo: WMF Martins Fontes.

legislativas trazidas pelo avanço tecnológico e, em especial, pelas implicações globais e transfronteiriças causadas pela internet.

O segundo fator diz respeito à escolha da União Europeia como parâmetro comparativo sobre o uso de IA na segurança pública. Essa decisão metodológica foi feita, primeiramente, devido ao fato de que uma análise comparativa com o modelo da União Europeia diverge de uma comparação restrita a uma única jurisdição nacional – em decorrência da própria pluralidade de países sob a qual o bloco se organiza. A isto, soma-se o fato da UE representar uma região que já conta com um acúmulo considerável sobre temas relacionados à implementação e regulação de novas tecnologias.

Além disso, cabe apontar uma última constatação, relacionada ao poderio jurídico-político e o consequente poder de *lobby* do qual o bloco europeu é dotado. Conforme elabora Abraham L. Newman²⁵⁰, a UE usa de sua influência e relevância política internacional para impor ao restante do mundo seus padrões regulatórios e salvaguardas legais, sob o risco de interposição de sanções diplomáticas e barreiras comerciais insuperáveis a qualquer Estado que se opuser aos padrões de segurança mínima que são considerados adequados sob a ótica europeia. Nesse sentido, mostra-se não somente necessário ter em vista os avanços regulatórios do bloco europeu, como também fútil buscar um afastamento das discussões lá vigentes sobre o tema – com o adendo de que não se pretende aqui uma importação acrítica das discussões europeias sobre o tema, mas sim uma comparação dos avanços das duas regiões analisadas no tópico analisado, considerando os pontos positivos e negativos de cada uma dessas abordagens.

2. O ESTADO DA REGULAÇÃO DE IA NA UNIÃO EUROPEIA

A União Europeia tem investido significativamente em pesquisa e desenvolvimento no campo da inteligência artificial nos últimos anos. Em abril de 2018, a Comissão Europeia lançou a “Estratégia Europeia para a Inteligência Artificial”²⁵¹, que estabeleceu um plano de ação para a UE se tornar líder mundial em IA ética, confiável e segura. Diversas normativas tratam sobre aspectos da implementação e uso de tecnologias baseadas em IA no bloco.

Notoriamente e a priori, pode-se citar o Regulamento Geral sobre a Proteção de Dados

250 NEWMAN, Abraham. **Protectors of privacy: Regulating personal data in the global economy.** Cornell University Press, 2008.

251 COMISSÃO EUROPEIA. **A European approach to artificial intelligence.** Disponível em: <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>. Acesso em: 13 jul. 2023.

(RGPD)²⁵², vigente desde 2018. A normativa, de caráter vinculante a todos os países que integram a UE, trata de maneira geral sobre atividades de tratamento de dados pessoais e estabelece uma série de direitos e garantias para titulares de dados localizados no território europeu, bem como um rol de deveres e obrigações que devem ser observados por agentes de tratamento – localizados ou não na Europa – que utilizarem dados pessoais desses cidadãos europeus.

O RGPD versa brevemente sobre atividades de tratamento automatizado de dados pessoais em seu artigo 22 – tecnologias baseadas em inteligência artificial, em regra, são englobadas por essa definição. Contudo, importa apontar que as garantias apresentadas pelo RGPD nesse sentido contemplam diversas ressalvas.

Primeiramente, as previsões do artigo 22 dizem respeito ao direito dos cidadãos europeus a se oporem a atividades de tratamento de dados totalmente automatizadas – atividades de tratamento desempenhadas por algoritmos de IA, mas que incluam qualquer elemento de ação humana podem ser consideradas como não enquadradas nessa definição. Na ausência de uma definição mais precisa sobre os limites da atuação humana necessária para exaurir uma atividade de tratamento da tutela desse dispositivo legal, a lacuna observada quanto ao enunciado representa um risco potencial para que agentes de tratamento apoiem-se em “tecnicidades legais” para evitar a aplicabilidade do enunciado às suas atividades.

Além disso, destaca-se o fato de que, conforme enuncia o artigo 2(2) da normativa, o RGPD como um todo não se aplica a atividades voltadas à manutenção da segurança pública, persecução penal, manutenção da justiça, investigação de crimes e correlatas. Nesse sentido, o Regulamento não versa em momento algum sobre o uso de tecnologias de IA pelo Poder Público nas atividades que compõem o objeto de estudo do presente trabalho.

Por sua vez, o Regulamento dos Serviços Digitais (*Digital Services Act – DSA*)²⁵³ foi aprovada juntamente com a Regulamento dos Mercados Digitais (*Digital Markets Act – DMA*)²⁵⁴, em uma tentativa do bloco europeu de melhor regular o fornecimento de serviços por plataformas de internet, bem como proteger direitos das pessoas que

252 UNIÃO EUROPEIA. **Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho** de 27 de abril de 2016 relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados). Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32016R0679&qid=1689301134978>. Acesso em: 13 jul. 2023.

253 UNIÃO EUROPEIA. **Regulamento (UE) 2022/2065 do Parlamento Europeu e do Conselho** de 19 de outubro de 2022 relativo a um mercado único para os serviços digitais e que altera a Diretiva 2000/31/CE (Regulamento dos Serviços Digitais). Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32022R2065>. Acesso em: 13 jul. 2023.

254 UNIÃO EUROPEIA. **Regulamento (UE) 2022/1925 do Parlamento Europeu e do Conselho** de 14 de setembro de 2022 relativo à disputabilidade e equidade dos mercados no setor digital e que altera as Diretivas (UE) 2019/1937 e (UE) 2020/1828 (Regulamento dos Mercados Digitais). Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32022R1925>. Acesso em: 13 jul. 2023.

utilizam essas plataformas dentro da UE²⁵⁵. As leis foram aprovadas em 2022, mas ainda não eram aplicáveis em sua totalidade quando da escrita deste relatório.

A Lei de Serviços Digitais traz diversos mecanismos que objetivam a garantia de direitos fundamentais da população europeia durante o uso de produtos e serviços na internet. Esses direitos incluem, entre outros, a proteção da liberdade de expressão, a transparência sobre os algoritmos que norteiam o funcionamento dessas plataformas e o estabelecimento de regras relativas à remoção de conteúdo ilegal.

No que tange ao uso de tecnologias de inteligência artificial, o DSA traz previsões que autorizam, embora de forma limitada, o uso de meios automatizados para identificação, moderação, denúncia e contingenciamento de conteúdo ilegal em plataformas digitais. Plataformas devem, por exemplo, prestar contas quanto aos índices de falha dos modelos algorítmicos empregados em atividades conduzidas por IA. Os limites ao uso dessas tecnologias, contudo, apresentam uma lacuna similar àquela que foi apontada quanto ao artigo 22 do RGPD: desautoriza-se o uso de meios totalmente automatizados para administrar a moderação de conteúdo, a comunicação desses procedimentos a usuários, entre outros – contudo, a inserção de um único elemento humano nessa cadeia de custódia pode ser suficiente para afastar de provedores de aplicações de internet as repercussões previstas quanto

Adicionalmente, outra similaridade entre o DSA e o RGPD diz respeito à inaplicabilidade de seus enunciados às atividades relacionadas à manutenção da segurança pública por entidades estatais – por se tratar de uma normativa voltada para a regulação do mercado digital no âmbito europeu. Por esse motivo, não há na lei quaisquer previsões pertinentes ao objeto central de análise do presente estudo.

Já a Diretiva 2016/680²⁵⁶ foi elaborada para tratar sobre as atividades de tratamento de dados pessoais que fogem ao escopo de aplicabilidade do RGPD – mais precisamente, aquelas previstas no artigo 2 do Regulamento, como as atividades pertinentes à manutenção da segurança pública pelo Estado, a segurança nacional, a investigação e punição de crimes e delitos, entre outras. Nesse sentido, a Diretiva 680 representa o primeiro exemplo de normativa europeia que versa especificamente sobre temas relacionados ao uso de dados pessoais na segurança pública – seja mediante o tratamento de dados pessoais de forma manual, automatizada ou parcialmente automatizada.

255 PARLAMENTO EUROPEU. **A Lei dos Mercados Digitais e a Lei dos Serviços Digitais da UE em detalhe**. Publicado em 14 dez. 2021. Disponível em: <https://www.europarl.europa.eu/news/pt/headlines/society/20211209STO19124/a-lei-dos-mercados-digitais-e-da-lei-dos-servicos-digitais-da-ue-explicadas>. Acesso em: 13 jul. 2023.

256 UNIÃO EUROPEIA. **Diretiva (UE) 2016/680 do Parlamento Europeu e do Conselho** de 27 de abril de 2016 relativa à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais pelas autoridades competentes para efeitos de prevenção, investigação, deteção ou repressão de infrações penais ou execução de sanções penais, e à livre circulação desses dados, e que revoga a Decisão-Quadro 2008/977/JAI do Conselho. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32016L0680>. Acesso em: 13 jul. 2023.

Cabe apontar, contudo, que, entre as exceções de aplicabilidade previstas no artigo 2 da Diretiva, pode-se constatar que a legislação em análise não se aplica às entidades administrativas que compõem a União Europeia propriamente dita – apenas as autoridades dos Estados-membros do bloco que são competentes para conduzir atividades relacionadas ao escopo material da Diretiva. Observa-se nesse ponto, portanto, o primeiro limite de aplicação cuja menção é relevante quanto a essa normativa.

Quanto às suas previsões procedimentais e de garantia de direitos, a Diretiva elenca uma série de direitos de titulares de dados que devem ser observados durante o curso de investigações criminais. Dados pessoais devem, por exemplo, ser utilizados em observância de critérios de proporcionalidade, necessidade e adequação, de maneira similar aos direitos que foram inicialmente previstos no RGPD. Além disso, autoridades públicas dos Estados-membros da UE devem cumprir requisitos de transparência sobre o uso de dados pessoais e, para isso, manter registros detalhados sobre quais informações são armazenadas sobre uma pessoa, e quais as atividades de tratamento às quais cada uma dessas informações é submetida. Titulares têm também direito a exigir que dados desatualizados ou incorretos sejam retificados ou excluídos, bem como exigir reparações em casos de tratamentos ilegais de suas informações pessoais. Ainda, relatórios de impacto devem ser realizados para averiguar se o nível de risco ao qual os titulares de dados pessoais envolvidos nessas atividades é inferior ao grau de interesse público derivado destas.

Finalmente, quanto ao uso totalmente automatizado de dados pessoais, a Diretiva busca – sem grande sucesso – uma evolução quanto aos pontos controversos trazidos pelo RGPD e o DSA e já abordados anteriormente. Mais especificamente, a Diretiva 680 busca em um primeiro momento proibir processos de tomada de decisões totalmente automatizadas pelas autoridades investigativas estatais, em seu artigo 11(1). Mais adiante no mesmo enunciado legal, contudo, esse comando é relativizado para permitir a tomada de decisões totalmente automatizada quando a legislação nacional do país em questão diretamente autorizar esse procedimento – o que esvazia a proteção legal inicialmente pretendida.

O artigo 11(2), por sua vez, visa ao menos evitar que tomadas de decisões totalmente automatizadas sejam feitas mediante o uso de categorias especiais de dados pessoais – ou seja, dados sensíveis, como aqueles relativos a origem étnico-racial, opiniões políticas, religiosas ou filosóficas, filiação sindical, dados genéticos, biométricos, entre outros. Novamente, contudo, essa garantia é flexibilizada logo na sequência, com a única expectativa de que dados pessoais sensíveis sejam utilizados em decisões totalmente automatizadas em observância de critérios não específicos de proteção aos direitos e liberdades fundamentais dos titulares envolvidos, bem como para fins não-discriminatórios.

Por fim, cabe abordar também o Proposta de Regulamento sobre Inteligência Artificial

da UE²⁵⁷. O texto – ainda em tramitação quando da redação deste estudo – busca ser a primeira normativa especificamente voltada para a regulação do desenvolvimento, uso e comercialização de tecnologias de inteligência artificial.

Embora o texto vigente da proposta quando da redação deste estudo excluísse do escopo de aplicação da lei o uso de tecnologias de IA para fins exclusivamente militares, o uso desses algoritmos por autoridades estatais para fins de segurança pública ainda estava contemplado pela proposta. Em sua atual versão, o texto contém previsões que representam avanços significativos quanto à proteção de direitos fundamentais no contexto do uso de tecnologias baseadas em IA para prevenção e combate ao crime.

Pode-se citar, por exemplo, o artigo 5 da Proposta, que prevê a proibição do uso de tecnologias de perfilamento e atribuição de “notas” para medir a “confiabilidade” de pessoas com base em características físicas ou hábitos. Essa previsão se dá especificamente quanto a contextos em que essas notas podem ser posteriormente utilizadas para perpetuar um tratamento desfavorável para uma pessoa ou grupo, ou de forma desproporcional ou inadequada com relação aos propósitos originais do tratamento.

Ainda no artigo 5, observa-se também a proibição do uso de tecnologias que possibilitam o monitoramento e a identificação de pessoas em tempo real mediante o uso de dados biométricos em lugares de acesso público. Há, contudo, exceções para essa proibição a fim de possibilitar atividades como a busca por vítimas de crimes ou crianças perdidas; a prevenção de crimes como terrorismo e ameaças à saúde ou à vida de pessoas, quando existem informações substanciais sobre sua iminência; e a identificação e persecução de suspeitos ou responsáveis por condutas gravemente lesivas e puníveis com detenção por ao menos três anos, como ataques terroristas, exploração de menores, entre diversos outros, constantes no artigo 2(2) da Decisão-quadro do Conselho da UE 2002/584/JAI²⁵⁸.

Apesar de essas exceções ensejarem possíveis aberturas da regulação em comento para abusos por parte de autoridades estatais – em especial, esta última exceção, dada a quantidade de condutas previstas no artigo 2(2) da Decisão-quadro do Conselho 2002/584/JAI –, observa-se na Proposta de Regulamento sobre IA da UE um avanço significativo e desejável contra o uso indiscriminado da IA pelo Estado. Câmeras de vigilância com tecnologia de reconhecimento facial e tecnologias de policiamento preditivo baseado em IA, por exemplo, parecem perder espaço mediante as novas regras propostas pela proposta – e medidas como essas representam parcela considerável

257 COMISSÃO EUROPEIA. **Proposta de Regulamento do Parlamento Europeu e do Conselho** que estabelece regras harmonizadas em matéria de inteligência artificial (Regulamento Inteligência Artificial) e altera determinados atos legislativos da União. Publicado em 21 abr. 2021. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:52021PC0206>. Acesso em: 13 jul. 2023.

258 CONSELHO EUROPEU. **2002/584/JAI**: Decisão-quadro do Conselho, de 13 de junho de 2002, relativa ao mandado de detenção europeu e aos processos de entrega entre os Estados-Membros - Declarações de alguns Estados-Membros aquando da aprovação da decisão-quadro. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32002F0584>. Acesso em: 13 jul. 2023.

dos investimentos de autoridades públicas em tecnologias de inteligência artificial, não apenas na UE como também no Brasil.

A referida proposta estabelece, por fim, uma série de requisitos para o uso de sistemas de IA considerados de alto risco, incluindo avaliações de risco obrigatórias, mecanismos de mitigação desses riscos, testes de proporcionalidade, registros de uso, entre outros. Em termos gerais, apesar de ainda passível de críticas quanto a algumas de suas previsões, a Proposta de Regulamento sobre Inteligência Artificial da UE representa um avanço significativo no uso de tecnologias de IA tanto para fins privados quanto – e especialmente – para a manutenção da segurança pública.

Observa-se, assim, que o atual ambiente regulatório europeu sobre o uso de tecnologias de IA na segurança pública ainda carece de melhorias e desenvolvimentos. A Diretiva 680, apesar de trazer algumas salvaguardas sobre essas práticas, ainda necessita de complementação para que exista um cenário abrangente de defesa de direitos civis perante o uso irrestrito de novas tecnologias pelo Estado. O avanço da Proposta de Regulamento sobre IA da UE, contudo, apresenta perspectivas de melhoras na forma como o bloco europeu endereça legislativamente essa questão – embora não represente ainda um modelo ideal de governança dessas tecnologias.

3. O ESTADO DA REGULAÇÃO DE IA NO BRASIL

O uso cada vez mais disseminado da Inteligência Artificial (IA) levanta questões urgentes sobre a necessidade de regulamentação para garantir sua aplicação ética e responsável. No Brasil, o país ainda carece de uma legislação específica para regulamentar a IA.

Esse cenário torna-se particularmente preocupante à medida que tecnologias baseadas em IA já estão sendo implementadas em diversas áreas, inclusive na segurança pública. Neste capítulo, examinaremos o estado atual da regulação de IA no Brasil, destacando os desafios e as lacunas que precisam ser superados para promover um ambiente regulatório adequado que concilie a inovação tecnológica com a proteção dos direitos e valores fundamentais da sociedade.

Apesar da ausência de regulamentação específica para a IA no Brasil, existem propostas legislativas e movimentações estatais relevantes. Assim, o quadro regulatório para o desenvolvimento e uso responsável dessa tecnologia está em desenvolvimento e é composto por normas vinculantes de outras temáticas tangenciais, além de iniciativas e diretrizes. Na sequência, serão elencadas as principais experiências relacionadas ao tema no território brasileiro.

No âmbito do Poder Executivo, já existiam articulações nesse sentido, o que culminou na

Portaria nº 1.122/2020 do Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC)²⁵⁹, em que fixou-se que as ações e iniciativas sobre IA seriam prioritárias entre 2020 a 2023.

Nesse cenário, em 2021, o Ministério da Ciência, Tecnologia e Inovação (MCTI) criou a Estratégia Brasileira de Inteligência Artificial (EBIA), cujo propósito é “potencializar o desenvolvimento e a utilização da tecnologia com vistas a promover o avanço científico e solucionar problemas concretos do País, identificando áreas prioritárias nas quais há maior potencial de obtenção de benefícios”²⁶⁰. São elencados nove eixos temáticos, dentre eles a segurança pública, além de estabelecidos um total de 74 ações estratégicas.

Ao tratar sobre o uso de IA para fins de segurança pública, a EBIA identifica como principais aplicações: “(i) plataformas de cidades inteligentes e cidades seguras; (ii) sistemas de reconhecimento facial; e (iii) policiamento inteligente.”²⁶¹ Na mesma linha, pontua alguns casos concretos dessa aplicação para fins de vigilância e segurança pública: instrumentos de análise de áudio com envio automático de alerta para autoridades, por exemplo, em caso de identificação de tiros ou o que poderia classificar como “tumulto”; mecanismos de identificação de pessoas através de imagens e vídeos, inclusive os sistemas de reconhecimento facial (SRF), que servem tanto para identificar situações suspeitas (como a invasão de um domicílio) quanto para identificar potenciais comportamentos criminosos em vias públicas.

Contudo, o documento referência da EBIA registra as implicações danosas e as controvérsias atinentes a tais usos, que vêm se popularizando nos últimos anos no Brasil. Destacam-se os vieses e práticas discriminatórias decorrentes do funcionamento dessas tecnologias, sobretudo em recortes de gênero e raça²⁶². No caso da aplicação de tecnologias de reconhecimento facial, ressalta-se que existem um receio significativo em relação às “falsas identificações positivas”, já que esses erros podem gerar “constrangimentos, prisões arbitrárias e violações de direitos fundamentais”²⁶³.

Nesse ponto, a título ilustrativo, pode-se citar o relatório do Centro de Estudos de

259 BRASIL. Ministério da Ciência, Tecnologia, Inovações e Comunicações. **Portaria nº 1.122, de 19 de março de 2020**. Define as prioridades, no âmbito do Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC), no que se refere a projetos de pesquisa, de desenvolvimento de tecnologias e inovações, para o período 2020 a 2023. Disponível em: <https://www.in.gov.br/en/web/dou/-/portaria-n-1.122-de-19-de-marco-de-2020-249437397>. Acesso em: 05 jul. 2023.

260 MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO (MCTI). **Estratégia Brasileira para Inteligência Artificial (EBIA)**. Brasília, 2021, p. 4. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ia_estrategia_diagramacao_4-979_2021.pdf. Acesso em: 10 mar. 2022.

261 MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO (MCTI), *op. cit.*, p. 45.

262 MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO (MCTI). **Estratégia Brasileira para Inteligência Artificial (EBIA)**. Brasília, 2021, p. 45. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ia_estrategia_diagramacao_4-979_2021.pdf. Acesso em: 10 mar. 2022.

263 *Ibidem*.

Segurança e Cidadania (CESeC), expressamente mencionado no documento referência da EBIA, construído a partir do monitoramento de prisões realizado de março a outubro de 2019 com dados dos estados da Bahia, Rio de Janeiro, Santa Catarina e Paraíba. Uma das conclusões foi: “Em relação aos casos em que havia informações sobre raça e cor, ou quando havia imagens dos abordados (42 casos), 90,5% das pessoas eram negras e 9,5% eram brancas.”²⁶⁴

Outra pesquisa relevante citada foi promovida pelo National Institute of Standards and Technology, instituição vinculada ao Departamento de Comércio do Governo dos Estados Unidos: nela, constata-se que a ocorrência das “falsas identificações positivas” supracitadas se concentram em pessoas com fenótipo afrodescendente ou asiático quando comparadas a outras com traços caucasianos.²⁶⁵ Ressalte-se que tais tecnologias não são construídas de maneira neutra e estão permeadas por condicionantes sociais, desde o banco de dados que a alimenta, perpassando pela equipe que a desenvolve, até os vetores políticos que decidem quando, onde, como e para quem elas serão aplicadas com fins de vigilância e repressão.

A EBIA²⁶⁶ reconhece, então, algumas medidas para evitar essas situações danosas e orientar o desenvolvimento responsável e ético dessa tecnologia, assim como de sua aplicação na seara da proteção de dados pessoais, dentre os quais estão os guias e recomendações de boas práticas (*soft law*) e a adoção de padrões técnicos estabelecidos pela comunidade técnico-científica.

Tendo em vista esse quadro, a EBIA²⁶⁷ elenca um conjunto de ações estratégicas, que envolvem a fixação de mecanismos de supervisão e fiscalização do uso da IA na segurança pública, além do estímulo à apresentação de relatórios de impacto e de resultados do serviço implementado pelos órgãos públicos que utilizarem tal recurso. Ainda, apontam-se como estratégicas a disponibilização de mecanismo de resposta das pessoas monitoradas, o preenchimento da lacuna legislativa na aplicação da LGPD a fins de segurança pública e, por fim, um incentivo a criação de ambientes regulatórios experimentais de inovação em temas de privacidade e proteção de dados voltados à segurança pública (*sandboxes* regulatórios).

Já no Poder Judiciário, cuja atuação é fundamental no ecossistema da segurança pública

264 NUNES, Pablo. Novas ferramentas, velhas práticas: reconhecimento facial e policiamento no Brasil. In: RAMOS, Silvia (coord.). **Retratos da Violência: Cinco meses de monitoramento, análises e descobertas.** Rio de Janeiro: Rede de Observatórios da Segurança/CESeC, novembro de 2019, p. 69. Disponível em: <https://cesecseguranca.com.br/textodownload/retratos-da-violencia-cinco-meses-de-monitoramento-analises-e-descobertas/>. Acesso em: 05 jul. 2023.

265 GROTH, Patrick; NGAN, Mei; HANAOKA, Kayee. **Face Recognition Vendor Test (FRVT) – Part 3: Demographic Effects.** National Institute of Standards and Technology (NIST) – U. S. Department of Commerce. Dez. 2019, p. 2. Disponível em: <https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf>. Acesso em: 05 jul. 2023.

266 MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO (MCTI), op. cit., p. 45 - 48.

267 *Ibidem*, p. 45 a 48.

brasileira, destaca-se a Resolução nº 332 de 21/08/2020²⁶⁸, do Conselho Nacional de Justiça (CNJ), pois estabelece diretrizes éticas, transparência e governança para a produção e utilização de Inteligência Artificial (IA) no âmbito do Poder Judiciário. Em particular, a resolução enfatiza que a aplicação da IA nos processos de tomada de decisão deve cumprir critérios éticos fundamentais, como transparência, previsibilidade, possibilidade de auditoria e garantia de imparcialidade e justiça substancial.

Além disso, ressalta-se a importância de que as decisões judiciais apoiadas pela IA preservem valores essenciais, como igualdade, não discriminação, pluralidade, solidariedade e julgamento justo. Nesse sentido, fixa-se como fundamental a adoção de medidas que eliminem ou minimizem a opressão, a marginalização do ser humano e os erros de julgamento decorrentes de preconceitos, a fim de garantir a aplicação justa e imparcial da IA no contexto da segurança pública e do sistema de justiça.

No âmbito do Poder Legislativo, diversos projetos de lei propostos versam sobre a IA no país. Dentre eles, merece destaque o Projeto de Lei (PL) nº 21 de 2020,²⁶⁹ através do qual buscou-se apresentar um marco regulatório para essa temática. Em seu art. 4º, elenca como fundamentos a proteção aos direitos humanos, inclusive à igualdade, à pluralidade, à não-discriminação, à livre iniciativa e à privacidade. Pontos relevantes como a transparência e explicabilidade, além da segurança aparecem no art. 6º, dispositivo em que o PL lista os princípios que orientam o uso responsável de IA no território nacional.

Contudo, o projeto foi recebido com um grande volume de críticas – em especial, por parte da sociedade civil e da comunidade científica e tecnológica –, em decorrência de uma evidente falta de amadurecimento técnico, conceitual e normativo do texto original.²⁷⁰ Isto porque o PL, em um primeiro momento, limitou-se a apresentar uma série de princípios norteadores de cunho geral sobre a disciplina no território nacional, sem enunciar sobre parâmetros e exigências éticas, de transparência e auditabilidade dos sistemas de IA utilizados, bem como as hipóteses de viabilidade de uso dessas tecnologias, entre outros.

Nesse contexto, em dezembro de 2022, uma comissão de juristas, presidida pelo Ministro do do Superior Tribunal de Justiça (STJ) Ricardo Villas Bôas Cueva, foi encarregada de

268 BRASIL. Conselho Nacional de Justiça (CNJ). **Resolução nº 332 de 21/08/2020**. Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3429>. Acesso em 05 jul. 2023.

269 BRASIL. **Projeto de Lei 21/2020**. Estabelece fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil; e dá outras providências. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2236340>. Acesso em: 05 jul. 2023.

270 NOGUEIRA, Pablo. Projeto de marco legal da IA no Brasil é pouco consistente e pode ser inútil, dizem especialistas. **Jornal da UNESP**, 29 jul. 2021. Disponível em: <https://jornal.unesp.br/2021/07/29/projeto-de-marco-legal-da-ia-no-brasil-e-pouco-consistente-e-pode-ser-inutil-dizem-especialistas/>. Acesso em: 05 jul. 2023.

apresentar um relatório²⁷¹ contendo um substitutivo de teor mais robusto e detalhado sobre as diretrizes nacionais relativas ao desenvolvimento e uso de tecnologias de IA. O substitutivo foi apresentado ao Senado Federal em 03 de maio de 2023 e tramita atualmente como PL nº 2338/2023²⁷².

Com normas mais robustas, ainda que em alinhamento com as balizas citadas quanto ao PL nº 21/2020, o PL nº 2338/2023 está distribuído em cinco eixos fundamentais: os princípios; os direitos das pessoas afetadas; a classificação de riscos; as obrigações e requisitos de governança dos sistemas de IA; e a supervisão e responsabilização²⁷³. Assim, ele deve ser palco para diversos debates interessantes sobre as escolhas democráticas a serem feitas para determinar os rumos da regulação da IA no Brasil nos próximos meses. Até o momento da confecção do presente trabalho, contudo, não há lei específica aprovada para regular a matéria no contexto brasileiro.

Nesse cenário, importa lembrar que a Lei Geral de Proteção de Dados Pessoais (LGPD)²⁷⁴ estabelece regras para o uso e proteção de dados pessoais no país, incluindo aqueles que são processados por sistemas de IA. Porém, essa legislação prevê expressamente – nas alíneas de seu art. 4º, inciso III – que não está direcionada a regular o tratamento de dados para fins exclusivos de segurança pública, defesa nacional, segurança do Estado e persecução penal. Já em seu art. 4º, §1º, esse diploma legal prevê que esse rol de exceções será disciplinado por legislação específica.

Em vista disso, existem relevantes propostas normativas em curso para sanar essa lacuna legislativa, destacam-se: primeiro, o Anteprojeto de Lei de Proteção de Dados Pessoais para Segurança Pública e Persecução Penal²⁷⁵ – conhecida popularmente como “LGPD Penal” –, que foi debatido multissetorialmente por cerca de um ano e apresentado em novembro de 2020; após quase um ano e meio, o PL nº 1515/2022²⁷⁶

271 BRASIL. **Relatório Final: COMISSÃO DE JURISTAS RESPONSÁVEL POR SUBSIDIAR ELABORAÇÃO DE SUBSTITUTIVO SOBRE INTELIGÊNCIA ARTIFICIAL NO BRASIL.** Brasília, DF, Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento/download/777129a2-e659-4053-bf2e-e4b53edc3a04>. Acesso em: 10 mar. 2023.

272 BRASIL. **Projeto de Lei nº 2338, de 2023.** Dispõe sobre o uso da Inteligência Artificial. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>. Acesso em: 05 jul. 2023.

273 SUPERIOR TRIBUNAL DE JUSTIÇA. **Projeto que regula IA é apresentado ao Senado após trabalho da comissão liderada pelo ministro Cueva.** 04 maio 2023. Disponível em: <https://www.stj.jus.br/sites/portals/Paginas/Comunicacao/Noticias/2023/04052023-Projeto-que-regula-IA-e-apresentado-ao-Senado-apos-trabalho-da-comissao-liderada-pelo-ministro-Cueva.aspx>. Acesso em 05 jul. 2023.

274 BRASIL. **Lei nº 13.709, de 14 de agosto de 2018.** Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2020. Disponível em: https://www.planalto.gov.br/ccivil_03/ato2019-2022/2020/lei/l14020.htm. Acesso em: 14 abr. 2023.

275 GT - COMISSÃO DE JURISTAS - SEGURANÇA PÚBLICA. **Anteprojeto de Lei de Proteção de Dados para segurança pública e persecução penal.** Disponível em: <https://www2.camara.leg.br/atividade-legislativa/comissoes/grupos-de-trabalho/56a-legislatura/comissao-de-juristas-dados-pessoais-seguranca-publica/outros-documentos>. Acesso em 05 jul. 2023.

276 BRASIL. **Projeto de Lei 1515/2022.** Lei de Proteção de Dados Pessoais para fins exclusivos de segurança do Estado, de defesa nacional, de segurança pública, e de investigação e repressão de infrações penais. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2326300>. Acesso em: 05 jul. 2023.

foi proposto, sob a justificativa de ter como objetivo “resguardar segurança pública e defesa nacional”²⁷⁷. Apesar de seguir a mesma estrutura do anteprojeto citado, o PL propõe modificações significativas em seu conteúdo, o que culminou em críticas incisivas sobre uma debilitação do texto do anteprojeto, no sentido de reduzir direitos, mecanismos e conceitos voltados à garantir o controle do tratamento de dados na área penal.

Nesse ponto, há uma discussão sobre as decisões automatizadas, que costumam remeter ao funcionamento de sistemas baseados em IA.²⁷⁸ Azevedo e outros²⁷⁹ apontam que foi possível identificar, em uma análise comparativa do anteprojeto e do projeto lei mencionados, um:

Enfraquecimento de direitos e proteções referentes a decisões automatizadas, como exigência de autorização prévia e de relatórios de impacto adequadamente procedimentalizados, em favor de disposições genéricas e de aplicabilidade limitada.

Sob os debates descritos, além de outros possíveis – os quais não se pretende esgotar aqui, diante de sua complexidade e a limitação do escopo deste texto –, tais propostas legislativas seguem e, atualmente, subsiste a lacuna jurídica no que se refere a como lidar com o tratamento de dados na aplicação de IA e outras tecnologias na seara da segurança pública.

Por outro lado, o PL nº 8.045/2010 e seus apensados²⁸⁰ – que buscam reformar e atualizar o Código de Processo Penal (CPP)²⁸¹, texto normativo vigente desde 1941, ainda que tenha sofrido alterações esparsas – também trazem normas que tangenciam aspectos da segurança pública. Apesar de representarem uma tentativa de aprimoramento da persecução penal, com as devidas atualizações em decorrência da era digital, especialistas têm apontado que as normas propostas podem violar direitos fundamentais e outras garantias constitucionais, inclusive o devido processo legal.

277 AGÊNCIA CAMARA DE NOTÍCIAS. **Projeto altera Lei de Proteção de Dados para resguardar segurança pública e defesa nacional**. 12 ago. 2022. Disponível em: <https://www.camara.leg.br/noticias/893704-projeto-altera-lei-de-protecao-de-dados-para-resguardar-seguranca-publica-e-defesa-nacional/>. Acesso em 05 jul. 2023.

278 AZEVEDO, Cynthia Picolo Gonzaga de; LIMA, Eliz Marina Bariviera de; SILVA, Felipe Rocha da; RODRIGUES, Gustavo Ramos; DUTRA, Luiza Corrêa de Magalhães; SANTARÉM, Paulo Rená da Silva; RODRIGUES, Victor Barbieri Vieira Rodrigues. **Nota técnica**: análise comparativa entre o anteprojeto de LGPD penal e o PL 1515/2022. Instituto de Referência em Internet e Sociedade (IRIS) e Laboratório de Políticas Públicas e Internet (LAPIN), novembro de 2022, p. 19 e 20. Disponível em: <https://lapin.org.br/wp-content/uploads/2022/11/Nota-tecnica-Analise-comparativa-entre-o-anteprojeto-de-LGPD-Penal-e-o-PL-15152022-1.pdf>. Acesso em: 06 jul. 2023.

279 *Ibidem*, p. 3.

280 BRASIL. **Projeto de Lei 8045/2010 e seus apensados**. Código de Processo Penal. Disponível em: <https://www.camara.leg.br/propostas-legislativas/490263>. Acesso em: 10 jul. 2023.

281 BRASIL. **Decreto-lei nº 3.689, de 3 de outubro de 1941**. Código de Processo Penal. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto-lei/del3689.htm. Acesso em: 10 jul. 2023.

Neste ponto, a Coalizão Direitos na Rede (CDR)²⁸² chama atenção para os seguintes aspectos, que são de interesse para aplicação de IA para fins de segurança pública no PL nº 8.045/2010: a ampliação excessiva da retenção em larga escala de dados para investigações futuras; a regressão nas garantias envolvendo a interceptação de comunicações; a imposição de requisitos aos provedores que podem resultar em fragilidades de segurança em seus sistemas e serviços; e a legitimação de atividades de hacking realizadas por agentes governamentais. Sob esses embates, o PL nº 8.045/2010 e seus apensados seguem em tramitação, sem perspectiva exata de uma possível aprovação.

Ao longo deste capítulo, exploramos o quadro regulatório da Inteligência Artificial (IA) para fins de segurança pública no Brasil, examinando as principais normas oriundas do Legislativo, Executivo e Judiciário. Foi possível constatar que, embora existam iniciativas promissoras, além dos avanços e esforços de diferentes esferas governamentais, ainda há lacunas significativas a serem abordadas. Dessa forma, faz-se necessário um esforço conjunto entre os poderes e uma maior articulação entre as esferas multissetoriais para garantir a harmonização e a efetividade das regulamentações.

Nota-se, contudo, uma tendência nítida ao reconhecimento de alguns pilares para a regulação da IA no país: a proteção dos direitos humanos fundamentais, a transparência, a imparcialidade e a responsabilidade são considerados valores primordiais no desenvolvimento e aplicação da IA, inclusive na segurança pública. Dessa forma, é essencial que o Brasil avance na construção de um arcabouço regulatório sólido e atualizado, capaz de garantir o uso ético, responsável e eficiente da IA em benefício da sociedade e da promoção de uma segurança pública eficaz e justa.

4. CONSIDERAÇÕES FINAIS

Através do presente trabalho, apresentou-se um panorama do estado da arte do campo regulatório da IA com ênfase na segurança pública no Brasil e na UE. Desse modo, é possível contribuir para a construção de vias regulatórias mais adequadas às melhores práticas internacionais, mas com respeito às especificidades regionais, ainda que exista uma influência política transnacional das experiências jurídicas europeias.

Após análise detida dos diferentes cenários regulatórios no território brasileiro e da UE, foi possível confirmar a hipótese inicial, já que foram constatadas divergências e convergências.

Nesse sentido, há uma diferença significativa na robustez das garantias sendo

282 COALIZÃO DIREITOS NA REDE. **Reforma do Código de Processo Penal pode aumentar vigilância e precisa de equilíbrio em questões de tecnologia.** 20 maio 2021. Disponível em: <https://direitosnarede.org.br/2021/05/20/reforma-do-codigo-de-processo-penal-pode-aumentar-vigilancia-e-precisa-de-equilibrio-em-questoes-de-tecnologia/>. Acesso em: 10 jul. 2023.

reconhecidas através de normas regulamentadoras vigentes e sendo gestadas pelos respectivos órgãos legiferantes, em que o Brasil aparece em desvantagem nesse aspecto. Contudo, é possível notar desafios regulatórios e a necessidade de amadurecimento das propostas normativas correlatas também no âmbito da União Europeia.

As principais diferenças encontradas entre os regulamentos dizem respeito aos limites estabelecidos para o uso de dados pessoais em atividades de tratamento derivadas de coleta via uso de tecnologias de IA. O bloco europeu (historicamente protetivo de salvaguardas e garantias relativas à privacidade e à proteção de dados de sua população) parece estabelecer limites mais concretos sobre o uso de tecnologias dessa natureza – com previsões legais mais robustas e limitantes da atividade estatal já em vigor, embora esparsas, e um projeto de lei ainda mais abrangente em tramitação. Enquanto isso, no Brasil, observa-se uma vacância considerável nesse quesito, e a propositura de projetos de lei que buscam uma consolidação das lacunas já existentes para o uso de IA pelo Poder Público – projetos estes que, ao menos, vêm sendo fortemente contestados por setores como a comunidade técnico-científica e a sociedade civil, o que, espera-se, abre portas para um aprimoramento das previsões brasileiras sobre o uso de IA na segurança pública.

Contudo, observa-se que há certo consenso no sentido de defender o uso cuidadoso e responsável na segurança pública, garantindo transparência, justiça, imparcialidade, proteção de dados e privacidade, além do controle humano sobre as decisões finais. Em prol de uma suposta otimização da segurança dos cidadãos (que carece de elementos comprobatórios), os impactos negativos à privacidade, proteção de dados e direitos humanos vêm sendo perpetuados em ambas as regiões, tais como a reprodução de vieses raciais nos casos de sistemas de reconhecimento facial.

Em face desse contexto, que prevalece tanto no Brasil quanto na UE, constatou-se alguns dos cuidados que devem ser tomados no uso de IA na segurança pública, que incluem os seguintes pontos.

Primeiro, *transparência e responsabilidade*: é importante que as decisões tomadas pela IA sejam transparentes e que haja responsabilização pelo uso danoso dessa tecnologia. Os responsáveis pela sua criação e desenvolvimento devem ser capazes de explicar como as decisões são tomadas e como os dados são usados. Ainda que essa explicabilidade se limite a certos limites técnicos, a essência dos vetores condicionantes desses retornos da máquina devem ser contemplados.

Segundo, a *limitação de vieses*: a IA pode ser influenciada por preconceitos e vieses existentes nos dados que alimentam os algoritmos. É importante garantir que os dados usados sejam representativos da população geral e que as decisões tomadas pela IA sejam justas e imparciais, sobretudo para sua aplicação à segurança pública. Além disso, a garantia da não discriminação perpassa diferentes fases do ciclo de uma IA, de modo que a mitigação desses vieses exige iniciativas amplas, frequentemente anunciadas de

forma principiológicas para que possam ser atualizadas em diferentes contextos.

Terceiro, a *proteção de dados pessoais*: a IA pode coletar e usar dados pessoais de indivíduos, como registros criminais e dados biométricos. As autoridades devem garantir que esses dados sejam coletados e usados de forma legal e ética, em conformidade com as leis de proteção de dados e, diante de eventual lacuna legislativa, em respeito às balizas constitucionais, outras normas tangenciais e às boas práticas elencadas em guias e recomendações.

Quarto, a *garantia do direito à privacidade*: a IA pode ser usada para monitorar e vigiar indivíduos, o que pode afetar sua privacidade. É importante que as autoridades estabeleçam limites claros para o uso de IA na vigilância e garantam que os indivíduos sejam informados sobre o uso da tecnologia e seus direitos de privacidade.

Quinto, a *centralidade humana* e, conseqüentemente, a viabilização do *controle humano*: é importante que as decisões finais em relação ao uso da IA na segurança pública sejam tomadas por humanos, e não apenas pela tecnologia. As autoridades devem garantir que os algoritmos sejam usados como uma ferramenta de apoio, não como uma substituição para o julgamento humano.

Dessa maneira, nota-se um esforço jurídico e político de especialistas – no caso brasileiro, em muito motivado por pressões da sociedade civil e da comunidade científica e tecnológica – para observação desses pontos nos instrumentos normativos que regulam IA, o que é refletido nos regulamentos e iniciativas regulatórias na UE e no atual estado de tramitação dos projetos de lei relativos à temática no Brasil. Contudo, há pontos a serem aprimorados e os caminhos regulatórios ainda são incertos, diante de tensões políticas em ambas as regiões.

5. Referências

AGÊNCIA C MARA DE NOTÍCIAS. **Projeto altera Lei de Proteção de Dados para resguardar segurança pública e defesa nacional**. 12 ago. 2022. Disponível em: <https://www.camara.leg.br/noticias/893704-projeto-altera-lei-de-protecao-de-dados-para-resguardar-seguranca-publica-e-defesa-nacional/>. Acesso em 05 jul. 2023.

ALMEIDA, Karen Rosa de; ZAGANELLI, Margareth Vetis; GONÇALVES, Maria Célia da Silva. Inteligência Artificial: Reflexões sobre os marcos regulatórios emergentes em suas aplicações – O caso brasileiro e tendências internacionais. **Revista Derecho y Cambio Social**, n. 62, out-dez 2020. Disponível em: <https://dialnet.unirioja.es/descarga/articulo/7626195.pdf>. Acesso em 10 mar. 2023.

AZEVEDO, Cynthia Picolo Gonzaga de; LIMA, Eliz Marina Bariviera de; SILVA, Felipe Rocha da; RODRIGUES, Gustavo Ramos; DUTRA, Luiza Corrêa de Magalhães; SANTARÉM,

Paulo Rená da Silva; RODRIGUES, Victor Barbieri Vieira Rodrigues. **Nota técnica:** análise comparativa entre o anteprojeto de LGPD penal e o PL 1515/2022. Instituto de Referência em Internet e Sociedade (IRIS) e Laboratório de Políticas Públicas e Internet (LAPIN), novembro de 2022. Disponível em: <https://lapin.org.br/wp-content/uploads/2022/11/Nota-tecnica-Analise-comparativa-entre-o-anteprojeto-de-LGPD-Penal-e-o-PL-15152022-1.pdf>. Acesso em: 06 jul. 2023.

BAR-GILL, Oren. **Algorithmic Price Discrimination When Demand Is a Function of Both Preferences and (Mis)perceptions.** University of Chicago Law Review. v. 86, n. 2, 2019, p. 217-254. Disponível em: <https://chicagounbound.uchicago.edu/uclrev/vol86/iss2/12>. Acesso em 10 mar. 2023.

BUREAU EUROPÉEN DES UNIONS DE CONSOMMATEURS AISBL (BEUC). **Automated decision making and artificial intelligence: a consumer perspective.** Tech. rep., Bureau Europeen des Unions de Consommateurs. 2018. Disponível em: https://www.beuc.eu/publications/beuc-x-2018-058_automated_decision_making_and_artificial_intelligence.pdf. Acesso em: 10 mar. 2023.

BRASIL. Conselho Nacional de Justiça (CNJ). **Resolução nº 332 de 21/08/2020.** Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3429>. Acesso em 05 jul. 2023.

BRASIL. **Decreto-lei nº 3.689, de 3 de outubro de 1941.** Código de Processo Penal. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto-lei/del3689.htm. Acesso em: 10 jul. 2023.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018.** Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2020. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/lei/l14020.htm. Acesso em: 14 abr. 2023.

BRASIL. Ministério da Ciência, Tecnologia, Inovações e Comunicações. **Portaria nº 1.122, de 19 de março de 2020.** Define as prioridades, no âmbito do Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC), no que se refere a projetos de pesquisa, de desenvolvimento de tecnologias e inovações, para o período 2020 a 2023. Disponível em: <https://www.in.gov.br/en/web/dou/-/portaria-n-1.122-de-19-de-marco-de-2020-249437397>. Acesso em: 05 jul. 2023.

BRASIL. **Projeto de Lei 8045/2010 e seus apensados.** Código de Processo Penal. Disponível em: <https://www.camara.leg.br/propostas-legislativas/490263>. Acesso em: 10 jul. 2023.

BRASIL. **Projeto de Lei 21/2020.** Estabelece fundamentos, princípios e diretrizes para o

desenvolvimento e a aplicação da inteligência artificial no Brasil; e dá outras providências. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2236340>. Acesso em: 05 jul. 2023.

BRASIL. **Projeto de Lei 1515/2022**. Lei de Proteção de Dados Pessoais para fins exclusivos de segurança do Estado, de defesa nacional, de segurança pública, e de investigação e repressão de infrações penais. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2326300>. Acesso em: 05 jul. 2023.

BRASIL. **Projeto de Lei nº 2338, de 2023**. Dispõe sobre o uso da Inteligência Artificial. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>. Acesso em: 05 jul. 2023.

BRASIL. **Relatório Final**: comissão de juristas responsável por subsidiar elaboração de substitutivo sobre Inteligência Artificial no Brasil. Brasília: Distrito Federal. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento/download/777129a2-e659-4053-bf2e-e4b53edc3a04>. Acesso em: 10 mar. 2023.

BRANDÃO, Luiza; RODRIGUES, Fernanda; RODRIGUES, Gustavo; ROMAN, Juliana; RENÁ, Paulo; VIEIRA, Victor. **Marco Legal da Inteligência Artificial Contribuições do IRIS à Comissão de Juristas do Senado**. Belo Horizonte: Instituto de Referência em Internet e Sociedade. 2022. Disponível em: <https://irisbh.com.br/wp-content/uploads/2022/06/Marco-Legal-da-Inteligencia-Artificial-Contribuicoes-do-IRIS-a-Comissao-de-Juristas-do-Senado-1.pdf>. Acesso em: 10 mar. 2023.

COALIZÃO DIREITOS NA REDE. **Reforma do Código de Processo Penal pode aumentar vigilância e precisa de equilíbrio em questões de tecnologia**. 20 maio 2021. Disponível em: <https://direitosnarede.org.br/2021/05/20/reforma-do-codigo-de-processo-penal-pode-aumentar-vigilancia-e-precisa-de-equilibrio-em-questoes-de-tecnologia/>. Acesso em: 10 jul. 2023.

COMISSÃO EUROPEIA. **A European approach to artificial intelligence**. Disponível em: <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>. Acesso em: 13 jul. 2023.

COMISSÃO EUROPEIA. **Proposta de Regulamento do Parlamento Europeu e do Conselho** que estabelece regras harmonizadas em matéria de inteligência artificial (Regulamento Inteligência Artificial) e altera determinados atos legislativos da União. Publicado em 21 abr. 2021. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:52021PC0206>. Acesso em: 13 jul. 2023.

CONSELHO EUROPEU. **2002/584/JAI**: Decisão-quadro do Conselho, de 13 de junho de 2002, relativa ao mandado de detenção europeu e aos processos de entrega entre os Estados-Membros - Declarações de alguns Estados-Membros aquando da aprovação

da decisão-quadro. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32002F0584>. Acesso em: 13 jul. 2023.

FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA. **Carta de adesão aos princípios do Fórum Brasileiro de Segurança Pública**. Disponível em: https://forumseguranca.org.br/storage/Carta_de_principios_-_FBSP.pdf. Acesso em: 13 jul. 2023.

GROTHER, Patrick; NGAN, Mei; HANAOKA, Kayee. **Face Recognition Vendor Test (FRVT) – Part 3: Demographic Effects**. National Institute of Standards and Technology (NIST) – U. S. Department of Commerce. Dez. 2019. Disponível em: <https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf>. Acesso em: 05 jul. 2023.

GT - COMISSÃO DE JURISTAS - SEGURANÇA PÚBLICA. **Anteprojeto de Lei de Proteção de Dados para segurança pública e persecução penal**. Disponível em: <https://www2.camara.leg.br/atividade-legislativa/comissoes/grupos-de-trabalho/56a-legislatura/comissao-de-juristas-dados-pessoais-seguranca-publica/outros-documentos>. Acesso em 05 jul. 2023.

MARCONI, Marina; LAKATOS, Eva. **Metodologia científica**. São Paulo: Editora Atlas, 2004.

MARQUES, Claudia Lima; MUCELIN, Guilherme. **Inteligência Artificial e “opacidade” no consumo**: a necessária revalorização da transparência para a proteção do consumidor. In: TEPEDINO, G.; SILVA, R. G. O Direito Civil na Era da Inteligência Artificial. 2020. São Paulo: Revista dos Tribunais, 411-439 p.

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO (MCTI). **Estratégia Brasileira para Inteligência Artificial (EBIA)**. Brasília, 2021. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ia_estrategia_diagramacao_4-979_2021.pdf. Acesso em: 10 mar. 2022.

NEVES, Marcelo. **Transconstitucionalismo**. 2009. São Paulo: WMF Martins Fontes.

NEWMAN, Abraham. **Protectors of privacy: Regulating personal data in the global economy**. Cornell University Press, 2008.

NOGUEIRA, Pablo. Projeto de marco legal da IA no Brasil é pouco consistente e pode ser inútil, dizem especialistas. **Jornal da UNESP**, 29 jul. 2021. Disponível em: <https://jornal.unesp.br/2021/07/29/projeto-de-marco-legal-da-ia-no-brasil-e-pouco-consistente-e-pode-ser-inutil-dizem-especialistas/>. Acesso em: 05 jul. 2023.

NUNES, Pablo. Novas ferramentas, velhas práticas: reconhecimento facial e policiamento no Brasil. In: RAMOS, Silvia (coord.). **Retratos da Violência**: Cinco meses de monitoramento, análises e descobertas. Rio de Janeiro: Rede de Observatórios da

Segurança/CESeC, novembro de 2019. Disponível em: <https://cesecseguranca.com.br/textodownload/retratos-da-violencia-cinco-meses-de-monitoramento-analises-e-descobertas/>. Acesso em: 05 jul. 2023.

PARLAMENTO EUROPEU. **A Lei dos Mercados Digitais e a Lei dos Serviços Digitais da UE em detalhe**. Publicado em 14 dez. 2021. Disponível em: <https://www.europarl.europa.eu/news/pt/headlines/society/20211209STO19124/a-lei-dos-mercados-digitais-e-da-lei-dos-servicos-digitais-da-ue-explicadas>. Acesso em: 13 jul. 2023.

PREFEITURA DE SÃO PAULO. **Nova Plataforma de Videomonitoramento Smart Sampa**. São Paulo, 2023. Disponível em: <https://participemais.prefeitura.sp.gov.br/legislation/processes/209>. Acesso em: 13 jul. 2023.

RODOTÀ, Stefano. **A vida na sociedade da vigilância: a privacidade hoje**. Rio de Janeiro: Renovar, 2008.

SICHMAN, Jaime. Inteligência Artificial e sociedade: avanços e riscos. **Estudos Avançados**, São Paulo, v. 35, n. 101, p. 37-50, 2021. ISSN: 0103-4014. Disponível em: <https://www.revistas.usp.br/eav/article/view/185024>. Acesso em: 10 mar. 2023.

SUPERIOR TRIBUNAL DE JUSTIÇA. **Projeto que regula IA é apresentado ao Senado após trabalho da comissão liderada pelo ministro Cueva**. 04 maio 2023. Disponível em: <https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/2023/04052023-Projeto-que-regula-IA-e-apresentado-ao-Senado-apos-trabalho-da-comissao-liderada-pelo-ministro-Cueva.aspx>. Acesso em 05 jul. 2023.

UNIÃO EUROPEIA. **Diretiva (UE) 2016/680 do Parlamento Europeu e do Conselho** de 27 de abril de 2016 relativa à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais pelas autoridades competentes para efeitos de prevenção, investigação, deteção ou repressão de infrações penais ou execução de sanções penais, e à livre circulação desses dados, e que revoga a Decisão-Quadro 2008/977/JAI do Conselho. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32016L0680>. Acesso em: 13 jul. 2023.

UNIÃO EUROPEIA. **Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho** de 27 de abril de 2016 relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados). Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32016R0679&qid=1689301134978>. Acesso em: 13 jul. 2023.

UNIÃO EUROPEIA. Regulamento (UE) 2022/1925 do Parlamento Europeu e do Conselho de 14 de setembro de 2022 relativo à disputabilidade e equidade dos mercados no setor digital e que altera as Diretivas (UE) 2019/1937 e (UE) 2020/1828 (Regulamento dos

Mercados Digitais). Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32022R1925>. Acesso em: 13 jul. 2023.

UNIÃO EUROPEIA. **Regulamento (UE) 2022/2065 do Parlamento Europeu e do Conselho** de 19 de outubro de 2022 relativo a um mercado único para os serviços digitais e que altera a Diretiva 2000/31/CE (Regulamento dos Serviços Digitais). Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32022R2065>. Acesso em: 13 jul. 2023.

VICENTE, Dário Moura. **Direito Comparado**. v. 1: Introdução e Parte Geral. Almedina: Coimbra, 2011.

VIEIRA, Leonardo Marques. **A problemática da inteligência artificial e dos vieses algorítmicos**: caso COMPAS. In: 2019 Brazilian Technology Symposium. [S.l.], 2019. Disponível em: <https://lcv.fee.unicamp.br/images/BTSym-19/Papers/090.pdf>. Acesso em: 13 jul. 2023.

WEBER, Rolf. **Shaping internet governance**: regulatory challenges. Berlim: Springer. 2010.

Grupo de Trabalho

ÉTICA E REGULAÇÃO DA IA

PESSOAS COORDENADORAS

Ana Bárbara Gomes Pereira
e Fernanda dos Santos
Rodrigues Silva

DA EXPLICABILIDADE À INTERPRETABILIDADE: DEVIDO PROCESSO LEGAL E MODELOS DE INTELIGÊNCIA ARTIFICIAL PARA TOMADA DE DECISÃO JUDICIAL

FROM EXPLICABILITY TO INTERPRETABILITY: DUE PROCESS OF LAW AND ARTIFICIAL INTELLIGENCE MODELS FOR JUDICIAL DECISION-MAKING

Rodrigo L. Canalli²⁸³

Resumo

Sistemas de tomada de decisão por inteligência artificial (IA) vem sendo ostensivamente empregados para aplicar regras jurídicas na determinação de direitos e obrigações. A opacidade de modelos eventualmente utilizados para produzir decisões judiciais tem levantado questões relacionadas à observância do devido processo legal, na medida em que em um Estado de direito considera-se que todos têm direito de saber os fundamentos de uma decisão tomada contra si. A emergência de modelos opacos de aprendizado de máquina, cujas previsões não podem ser explicadas de modo compreensível para humanos, tem impulsionado debates em torno da ideia de explicabilidade dos modelos de IA para tomada de decisão. O artigo argumenta que (i) pelo menos em relação a modelos de IA para tomada de decisão judicial, o critério da explicabilidade, ao mesmo tempo em que se mostra insuficiente para satisfazer as exigências constitucionais de publicidade e fundamentação das decisões judiciais, impõe uma forma de desnudamento que não é exigida dos juízes humanos; e (ii) um critério mais adequado seria o de modelos de tomada de decisão judicial interpretáveis, entendidas, como tais, decisões referidas ao direito vigente (legalidade), interna e externamente coerentes (consistência) e compatíveis com a decisão de um juiz humano em caso análogo.

Palavras-chave: algoritmos; devido processo legal; explicabilidade; opacidade; racionalidade jurídica.

283 Mestre em Direito, Estado e Constituição pela Universidade de Brasília, LL.M em Competition, Innovation and Information Law pela New York University School of Law, pesquisador do Ethics4AI; professor. Foi Assessor-Chefe de Inteligência Artificial do Supremo Tribunal Federal.

Abstract

Artificial Intelligence (AI) decision-making systems are already being ostensibly employed to apply legal rules in determining rights and obligations. The opacity of algorithmic models used to produce judicial decisions has raised issues related to the observance of due process of law, which commands that everyone has the right to know the grounds of a decision against themselves. The emergence of opaque machine learning models, whose predictions cannot be explained in a way understandable for humans, has driven debates about the idea of explicability of AI models for decision making. The article argues that (i) at least in relation to AI models for judicial decision, the standard of explicability, while being insufficient to satisfy the constitutional requirements of publicity and reasoning of court decisions, imposes a form of nakedness that is not required of human judges; and (ii) a more appropriate standard would be that of interpretable judicial decision-making models characterized by due reference to applicable law, internal and external coherence and compatibility with the decision of a human judge in analogous case.

Keywords: algorithms; due process of law; explicability; opacity; legal reasoning.

1. INTRODUÇÃO: TOMADA DE DECISÃO JUDICIAL E INTELIGÊNCIA ARTIFICIAL

Sistemas de tomada de decisão por inteligência artificial (IA) vem sendo ostensivamente empregados para aplicar regras jurídicas na determinação de direitos e obrigações. Nos Estados Unidos, por exemplo, algoritmos estão sendo usados para determinar os direitos de indivíduos a benefícios sociais, avaliar a performance de empregados, selecionar quem será demitido e auxiliar juízes a concederem ou negarem fiança e progressão de regime.²⁸⁴ No Quênia, pesquisa jurídica assistida por IA e análise preditiva têm sido implementadas para reduzir o tempo de tramitação de processos no Tribunal de Meio-Ambiente e Terras.²⁸⁵

284 CRAWFORD, Kate; SCHULTZ, Jason. AI systems as state actors. **Columbia Law Review**, New York, v. 119, n. 7, 2019, p. 1941.

285 OGONJO, Florence; GITONGA, Joseph Theuri; WAIREGI, Angeline; RUTENBERG, Isaac. Utilizing AI to Improve Efficiency of the Environment and Land Court in the Kenyan Judiciary: Leveraging AI Capabilities in Land Dispute Cases in the Kenyan Environmental and Land Court System. In: INTERNATIONAL WORKSHOP ON AI AND INTELLIGENT ASSISTANCE FOR LEGAL PROFESSIONALS IN THE DIGITAL WORKPLACE (LEGALAIIA 2021), 2, 2021, São Paulo. **Joint Proceedings of the Workshops on Automated Semantic Analysis of Information in Legal Text (ASAIL 2021) & AI and Intelligent Assistance for Legal Professionals in the Digital Workplace (LegalAIIA 2021)**, São Paulo: CEUR-WS, 2021, p 59-68.

Na Índia, onde a enorme variedade idiomática é um desafio para a prestação jurisdicional, o Poder Judiciário tem feito uso de ferramentas de IA para tradução de documentos a atos processuais, além da gestão de acervos processuais.²⁸⁶ Provavelmente uma das experiências mais ousadas de automação judicial, o chamado sistema de *smart courts*, em implementação na China, combina *big data*, *blockchain* e inteligência artificial determinativa com o objetivo alegado de promover acesso à justiça e celeridade processual.²⁸⁷ E, no Brasil, tribunais têm utilizado ferramentas que incorporam métodos de inteligência artificial para gerir acervos de processos judiciais, identificar casos similares e aplicar temas de repercussão geral.²⁸⁸

Tão notáveis quanto controversos, esses são apenas alguns exemplos de uma tendência crescente. A maioria deles, no entanto, pode honestamente ser reduzida a sistemas sofisticados de seleção entre conjuntos de opções predefinidas – conjuntos de regras coordenadas para a resolução de uma dada operação, refinadas por modelos probabilísticos.²⁸⁹

Embora seja intenso o debate atual sobre o potencial prático do uso da inteligência artificial na prática judicial, bem como na tomada de decisões assistida em geral, as suas implicações para teoria da decisão judicial ainda carecem de desenvolvimento. A discussão acerca da possibilidade teórica de uma entidade não humana (no caso, um algoritmo de inteligência artificial) ser capaz de desempenhar o mesmo tipo de raciocínio jurídico realizado por juízes humanos quando decidem os casos que lhes são submetidos apresenta a perspectiva de sistemas automatizados inteligentes de tomada de decisão judicial virem a ser desenvolvidos como progressões dos atuais sistemas de auxílio à tomada de decisão e modelos preditivos de decisões.²⁹⁰

Nesse contexto, entre os aspectos éticos e regulatórios relacionados ao uso de inteligência artificial para tomada de decisões automatizadas, o conceito de explicabilidade algorítmica tem sido frequentemente apresentado como parâmetro regulatório adequado ou suficiente para assegurar a transparência dos modelos de tomada de decisão em face dos desafios apresentados pela opacidade dos modelos, apelidados, por este motivo,

286 JAUHAR, Ameen; MISRA, Vaidehi; SENGUPTA, Arghya; CHAKRABARTI, Partha P.; GHOSH, Saptarishi; GHOSH, Kripabandhu. **Responsible AI for the Indian Justice System: A Strategy Paper**. Vidhi Centre for Legal Policy, TCG-Crest, 2021.

287 SHI, Changqing; SOURDIN, Tania; LI, Bin. The Smart Court: A New Pathway to Justice in China? **International Journal for Court Administration**, [s. l.], v. 12, n. 1, p. 4, Mar. 2021. Disponível em: <https://iacajournal.org/articles/10.36745/ijca.367>. Acesso em: 10 out. 2023.

288 NASCIMENTO, André; COELHO, José Leovigildo; MIRANDA, Péricles; MELLO, Rafael F. Aplicações da Inteligência Computacional no Judiciário. In: SALOMÃO, Luis Felipe (Coord.). **Inteligência Artificial: tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário brasileiro**. 2ª ed. Rio de Janeiro: FGV, 2022, p. 27.

289 REICH, Rob; SAHAMI, Mehran; WEINSTEIN, Jeremy M. **System Error: Where Big Tech Went Wrong and How We Can Reboot**. Londres: Hodder and Stoughton, 2021, p. 82.

290 ASH, Elliot. **Judge, Jury, and EXECute file: the brave new world of legal automation**. Londres: Social Market Foundation, CAGE Research Centre, 12 jun. 2018. Disponível em: https://warwick.ac.uk/fac/soc/economics/research/centres/cage/cage-final-elliott_ash.pdf. Acesso em: 10 out. 2023.

de *black boxes* (caixas-pretas):²⁹¹ como acessar, auditar, ou entender o funcionamento de algoritmos inescrutáveis e muitas vezes não intuitivos.²⁹²

Além disso, quando falamos de modelos de IA para tomada de decisão sobre direitos e deveres, seja na esfera judicial ou administrativa, a opacidade, em si um problema, é agravada pela necessidade, de ordem normativa, de se respeitar o devido processo legal (*o due process of law*), o qual inclui a garantia, a todos assegurada em um Estado democrático de direito, de conhecer os motivos de uma decisão estatal afetando seus direitos.

Discute-se, a seguir, uma suposta insuficiência do conceito de explicabilidade, tal como concebido e desenvolvido, para assegurar, com efetividade, a transparência, o acesso à informação relevante e o devido processo legal, quando se trata de modelos algorítmicos para tomada de decisão – judicial ou administrativa – sobre direitos e deveres. As exigências do devido processo legal, argumenta-se, somente podem ser endereçadas por uma abordagem regulatória compatível com a natureza hermenêutica da prática jurídica, o que pode ser alcançado com algoritmos que sejam, mais do que explicáveis, interpretáveis.

2. ALGORITMOS DE COMPUTADOR, COGNIÇÃO E RACIONALIDADE JURÍDICA

Como se sabe, um algoritmo de computador consiste em uma classe específica de aplicações pré-programadas que, abastecidas com um conjunto de dados, oferece uma resposta precisa. Modelos algorítmicos de tomada de decisão, em geral, são treinados para encontrar padrões em dados através de um processo chamado de aprendizado de máquina.²⁹³ Tarefas que vão desde pesquisa e planejamento a reconhecimento e interpretação de imagens, passando por reconhecimento de fala e tomada de decisão em situação de incerteza, são todas redutíveis a reconhecimento de padrões e classificação de informação. Em particular, algoritmos de aprendizado de máquina são treinados, validados e testados com elementos extraídos de bases de dados a partir dos quais o algoritmo “aprende”. No caso de algoritmos de tomada de decisão judicial, esses elementos corresponderiam a precedentes, leis, procedimentos etc. À medida em que mais dados são fornecidos a um algoritmo de aprendizado de máquina, amplia-se

291 PASQUALE, Frank. **The Black Box Society: The Secret Algorithms That Control Money and Information.** Harvard University Press, 2015, p. 4.

292 SELBST, Andrew D.; BAROCAS, Solon. The Intuitive Appeal of Explainable Machines. **Fordham Law Review**, New York, v. 87, n. 3, Nov. 2018, p. 1085.

293 REICH, Rob; SAHAMI, Mehran; WEINSTEIN, Jeremy M. **System Error: Where Big Tech Went Wrong and How We Can Reboot.** Londres: Hodder and Stoughton, 2021, p. 82.

o universo de erros e acertos a serem referenciados como parâmetros para ulterior refinamento. Desse modo, o algoritmo é capaz de oferecer respostas progressivamente mais acuradas para a classe de problemas que ele foi treinado para resolver, melhorando o seu desempenho a partir da experiência.

Uma categoria especialmente complexa de algoritmos que atualmente vem passando por um período acelerado de desenvolvimento compreende os modelos cognitivos: modelos computacionais que se assemelham a processos cognitivos humanos, simulando comportamentos sofisticados. Uma vez que a arquitetura desses algoritmos emula a estrutura de um cérebro biológico, eles também são chamados de redes neurais.²⁹⁴ Essa estrutura permite que representações matemáticas de possibilidades e probabilidades espelhem as capacidades humanas de raciocínio e inferência²⁹⁵ e, como tal, desempenhem processos que podem ser descritos como análogos à ponderação e balanceamento de regras e princípios.²⁹⁶

Por outro lado, a linguagem e a cognição humanas têm sido efetivamente descritas como implementações de algoritmos particularmente complexos para conjugar símbolos hierárquicos.²⁹⁷ Como artefato linguístico, o direito também é um sistema simbólico, ou modo de comunicação²⁹⁸ e, como tal, também implementa algoritmos.²⁹⁹ O direito processual, por exemplo, é um algoritmo que permite a advogados e juízes julgarem casos de maneira padronizada e previsível, assim como são os critérios elaborados pelos tribunais para reconhecer se determinado quadro fático se enquadra no escopo de um precedente. Se as leis – e o direito em geral – são algoritmos, elas são independentes da linguagem específica em que são codificadas³⁰⁰ e, portanto, podem ser traduzidas para uma linguagem legível por máquina.

Para desempenhar o mesmo tipo de raciocínio de um juiz ao aplicar uma regra a um fato, uma entidade julgadora qualquer (como uma IA) deve ser capaz de assimilar, com elevado grau de precisão, a descrição particular de um fato à hipótese abstrata e geral veiculada na regra, seja uma regra exigindo ou proibindo uma conduta (impondo um dever ou obrigação) ou uma regra conferindo certos direitos aos indivíduos.³⁰¹ A própria estrutura da argumentação jurídica é teórica, ou seja, é abstração. Nesse sentido, aplicar uma regra a um caso é um exercício de abstração, não exigindo nenhuma

294 *Ibidem*, p. 161.

295 *Ibidem*.

296 GARCEZ, Artur S. d'Avila, GABBAY, Dov M., LAMB, Luis C. A neural cognitive model of argumentation with application to legal inference and decision making. **Journal of Applied Logic**, vol. 12, n. 2, Jun. 2014, p. 109.

297 BERWICK, Robert C. & CHOMSKY, Noam. **Why Only Us: Language and Evolution**. Cambridge: MIT Press, 2016, p. 132.

298 *Ibidem*, p. 53.

299 ABITEBOUL, Serge; DOWEK, Gilles. **The Age of Algorithms**. Cambridge University Press, 2020, p. 82.

300 HART, H.L.A. **The Concept of Law**. Oxford University Press, Oxford, 2012, p. 27.

301 BERWICK; CHOMSKY, op. cit., p. 53.

habilidade imaginativa ou criativa particular além do processo cognitivo pelo qual é implementada uma hierarquia entre símbolos.³⁰² O juiz aprende o algoritmo quando frequenta a faculdade de direito e quando ele estuda a legislação e a jurisprudência dos tribunais. Uma vez que ele aprende o algoritmo do direito, o padrão abstrato, ele o aplica aos casos particulares que lhe são submetidos.

Como já argumentei previamente,³⁰³ em tese, uma IA será capaz de tomada de decisão judicial se ela produzir decisões que, ao aplicarem regras jurídicas a conjuntos identificados de fatos, se mostrarem qualificáveis como racionais, informadas e imparciais. Em apertada síntese, uma decisão racional é aquela em que a regra aplicável é identificada de acordo com parâmetros consistentes e compreensíveis.³⁰⁴ Uma decisão informada leva em consideração e pondera adequadamente todos os fatos relevantes e interesses afetados. Em algoritmos de aprendizado de máquina, essa habilidade repousaria em uma análise granular da jurisprudência relevante.³⁰⁵ E uma decisão imparcial é aquela na qual nenhum interesse subjetivo, consciente ou inconsciente, desempenha qualquer papel. Parâmetros provisionando todas essas características podem ser integrados a modelos algorítmicos. Tais modelos seriam efetivamente aptos a produzir decisões judiciais válidas, ainda que não atendessem ao que tem sido chamado de explicabilidade. Nessas circunstâncias, questiona-se a insistência em modelos propondo a sua superação em favor de modelos que chamo de interpretáveis. Embora não seja infrequente o uso indistinto das expressões explicabilidade e interpretabilidade, adotamos, na esteira de Cynthia Rudin,³⁰⁶ uma distinção rígida entre ambos.

3. OPACIDADE, EXPLICABILIDADE E INTERPRETABILIDADE

Opacidade nada mais é do que falta de transparência. Em parte, é resultado da sonegação de acesso ao código-fonte, dentro de uma estratégia de proteção da propriedade intelectual do desenvolvedor da ferramenta. Sem acesso ao código-fonte, não é possível saber como as decisões são tomadas pela ferramenta. Por outro lado, a opacidade também pode decorrer de uma característica estrutural de modelos de linguagem complexos, porque muitas vezes os modelos matemáticos gerados após o treinamento do algoritmo não são passíveis de serem compreendidos por seres humanos, aí incluídos os próprios engenheiros que desenvolveram o algoritmo. Nesses casos, ainda que se

302 CANALLI, Rodrigo L. Artificial intelligence and the model of rules: better than us? **AI and Ethics**, Vol. 1, n. 3, Ago. 2023, p. 869.

303 HILDEBRANDT, Mireille. Law as computation in the era of artificial intelligence: speaking law to the power of statistics. **University of Toronto Law Journal**, vol. 68, n. 1, 2018, p. 12.

304 *Ibidem*, p. 23.

305 RUDIN, Cynthia. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. **Nature Machine Intelligence**, vol. 1, n. 5, mai. 2019, p. 206.

306

tenha acesso ao código-fonte utilizado para implementar o modelo, continua sendo tecnicamente impossível identificar o critério, ou conjunto de critérios, utilizado pelo algoritmo para a tomada de decisões.

A opacidade de modelos eventualmente utilizados para produzir decisões judiciais levanta questões relacionadas à observância do devido processo legal, na medida em que em um Estado de direito todos têm direito de saber os fundamentos de uma decisão tomada contra si. De fato, a própria legitimidade da decisão judicial, no Estado de direito, está tipicamente associada à sua fundamentação, à expressão de uma certa racionalidade jurídica,³⁰⁷ e a ausência ou a deficiência da fundamentação leva à nulidade da decisão judicial.³⁰⁸ No Brasil, o artigo 93, IX, da Constituição Federal³⁰⁹ exige expressamente que todos os julgamentos sejam públicos e que as decisões judiciais sejam fundamentadas, circunstância que sugere fortemente a inconstitucionalidade de sistemas de tomada de decisão judicial por IA que não atendam a critérios de transparência.

A emergência de modelos opacos de aprendizado de máquina, cujas previsões não podem ser explicadas de modo compreensível para humanos, tem impulsionado debates em torno da ideia de explicabilidade dos modelos de IA para tomada de decisão. Planos corporativos e governamentais de regulação de IA têm adotado a explicabilidade como parâmetro normativo. É o caso, entre outros, como tem sido apontado, do Artigo 22º do Regulamento Geral sobre a Proteção de Dados, na União Europeia,³¹⁰ e, de modo mais incisivo, do Artigo 20 da Lei Geral de Proteção de Dados (Lei nº 13.709/2018),³¹¹ no Brasil, que asseguram, no caso de decisão tomada com base em tratamento automatizado

307 WALDRON, Jeremy. **A dignidade da legislação**. São Paulo: Martins Fontes, 2003, p. 23.

308 CARDOSO, Evorah Lusci; FANTI, Fabíola. Movimentos Sociais e Direito: o Poder Judiciário em Disputa. In: SILVA, Felipe Gonçalves (Org.). **Manual de Sociologia Jurídica**. 2ª ed. São Paulo: Saraiva, 2017.

309 “Art. 93 (...) IX - todos os julgamentos dos órgãos do Poder Judiciário serão públicos, e fundamentadas todas as decisões, sob pena de nulidade, podendo a lei limitar a presença, em determinados atos, às próprias partes e a seus advogados, ou somente a estes, em casos nos quais a preservação do direito à intimidade do interessado no sigilo não prejudique o interesse público à informação.” (BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília: Presidência da República, 2023.)

310 O Artigo 22º, § 3 do Regulamento Geral sobre Proteção de Dados (RGDP) da União Europeia impõe ao responsável, nos casos de decisões tomadas exclusivamente com base no tratamento automatizado de dados pessoais, que produzam efeitos na esfera jurídica do seu titular, a aplicação de “medidas adequadas para salvaguardar os direitos e liberdades e legítimos interesses do titular dos dados, designadamente o direito de, pelo menos, obter intervenção humana por parte do responsável, manifestar o seu ponto de vista e contestar a decisão.” Embora frequente na literatura, a alegação de que o referido dispositivo consagra um direito à explicação de decisões automatizadas é controversa. Nesse sentido, cfr. WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. **International Data Privacy Law**, vol. 7, n. 2, 2017, p. 76.

311 “Art. 20. O titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, incluídas as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade. (Redação dada pela Lei nº 13.853, de 2019)

§ 1º O controlador deverá fornecer, sempre que solicitadas, informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada, observados os segredos comercial e industrial.

§ 2º Em caso de não oferecimento de informações de que trata o § 1º deste artigo baseado na observância de segredo comercial e industrial, a autoridade nacional poderá realizar auditoria para verificação de aspectos discriminatórios em tratamento automatizado de dados pessoais.” (BRASIL. **Lei nº 13.709/2018**. Brasília: Presidência da República, 2023.)

de dados pessoais, o fornecimento, ao titular dos dados, de uma explicação sobre os critérios e procedimentos utilizados para a tomada de decisão automatizada.

A explicabilidade também é adotada como parâmetro na proposta da Comissão Europeia para a introdução de um marco normativo e regulatório para a inteligência artificial,³¹² e no Projeto de Lei nº 2338, de 2023, em tramitação no Senado Federal brasileiro. Este último, igualmente, visa a estabelecer normas gerais de caráter nacional para o desenvolvimento, implementação e uso responsável de sistemas de inteligência artificial no Brasil, com o objetivo de proteger os direitos fundamentais e garantir a implementação de sistemas seguros e confiáveis, em benefício da pessoa humana, do regime democrático e do desenvolvimento científico e tecnológico.

Modelos de aprendizado de máquina explicáveis podem ser descritos como passíveis de uma explicação *post hoc* mais ou menos detalhada dos seus processos internos, o que pode ser alcançado, no caso de modelos gerados por algoritmos proprietários fechados, inclusive por um segundo modelo criado para explicar o primeiro.³¹³ A explicabilidade é problemática porque frequentemente a informação oferecida sobre o funcionamento da ferramenta não é confiável, e pode até mesmo ser enganosa.³¹⁴ Além disso, o conceito parece sofrer de uma imprecisão crônica, uma vez que não há clareza sobre o que conta como explicação satisfatória de um modelo de tomada de decisão.³¹⁵

No presente artigo, argumenta-se que: (i) pelo menos em relação a modelos de IA para tomada de decisão judicial, o critério da explicabilidade, ao mesmo tempo em que se mostra insuficiente para satisfazer as exigências constitucionais de publicidade e fundamentação das decisões judiciais, impõe uma forma de desnudamento que não é exigida dos juízes humanos; e (ii) um critério mais adequado seria o de modelos de tomada de decisão judicial interpretáveis.

Por se tratar de uma noção cuja definição é dependente do domínio,³¹⁶ a interpretabilidade, na construção de um modelo algorítmico, é delineada de acordo com o seu propósito. Isso tem como consequência que uma noção de interpretabilidade para algoritmo de tomada de decisão sobre direitos não vai servir para modelos de diagnóstico de saúde, ou modelos para veículos autônomos. Por essa razão, a definição de modelo interpretável aqui proposta é específica para modelos de tomada de decisão judicial (decisão sobre direitos subjetivos e obrigações).

312 EUROPEAN UNION. **Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain Union legislative acts.** European Parliament, 2023.

313 RUDIN, Cynthia. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. **Nature Machine Intelligence**, vol. 1, n. 5, mai. 2019.

314 *Ibidem*, p. 207.

315 *Ibidem*, p. 214.

316 *Ibidem*, p. 206.

4. MODELOS DE INTELIGÊNCIA ARTIFICIAL INTERPRETÁVEIS E DEVIDO PROCESSO LEGAL

Não temos acesso aos processos cognitivos que entram em ação quando um juiz aplica a lei a um conjunto de fatos para decidir um caso, e que, não obstante, identificamos como raciocínio jurídico. Tais processos não são, por assim dizer, explicáveis. O que temos são justificativas, frequentemente produzidas posteriormente ao processo cognitivo de tomada de decisão, pelas quais os mecanismos mentais envolvidos são significados retrospectivamente – interpretados. Da mesma forma, os processos internos de um hipotético ente algorítmico capaz de processos cognitivos análogos a raciocínio jurídico podem, como visto, ser igualmente inacessíveis. De toda sorte, a fundamentação de uma decisão judicial – a exposição argumentativa dos fundamentos jurídicos que levaram o juiz a determinada conclusão – traduz uma interpretação dos seus processos cognitivos, e não a sua explicação.

Tomada a observância do devido processo legal como norte conceitual do uso de ferramentas de inteligência artificial no âmbito do Poder Judiciário, a explicabilidade mostra-se insuficiente para satisfazer o parâmetro constitucional, na dimensão da justificação das decisões, porque, com já visto, modelos de linguagem complexos nem sempre são efetivamente explicáveis.

Diferentemente, a interpretabilidade da decisão judicial produzida por um modelo algorítmico significa que, ainda que não seja possível explicar tecnicamente como um modelo de linguagem apontou para determinado resultado, essa conclusão deve, pelo menos, ser passível de ser reconhecida como uma decisão consistente com a disciplina legal relativa à matéria, interna e externamente coerente, e assimilável a uma decisão judicial uma tomada por um juiz humano.

Sem prejuízo de desdobramentos e desenvolvimentos futuros, e lembrando que a noção de interpretabilidade é específica para cada domínio,³¹⁷ sugiro que modelos interpretáveis de tomada de decisão judicial devam ser construídos contemplando, como características já identificadas, pelo menos: legalidade, consistência e compatibilidade. Por legalidade, entende-se a produção de decisões que tenham como referência o direito vigente, observada a hierarquia das espécies normativas e a jurisprudência dos tribunais. Uma decisão consistente apresenta coerência interna (ausência de contradição argumentativa) e externa (devida consideração dos fatos sobre os quais se decide). Por fim, a decisão deve ser reconhecida como sendo compatível com uma decisão que seria tomada por um juiz humano julgando um caso análogo.

317 RUDIN, Cynthia. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. **Nature Machine Intelligence**, vol. 1, n. 5, mai. 2019.

Essas características podem ser decompostas em aspectos procedimentais e materiais mais específicos. Em um modelo de jurisdição adversarial, a consistência externa da decisão pode ser vinculada, por exemplo, à efetiva consideração, ou avaliação, pelo modelo de IA, dos argumentos normativos veiculados pelas partes.³¹⁸

Mais do que a tão propalada explicabilidade, e tanto quanto qualquer juiz humano, para satisfazer o devido processo legal, um modelo de IA para tomada de decisão judicial deve observar as condições que assegurem a sua interpretabilidade.

5. CONCLUSÃO

Conforme exposto acima, modelos algorítmicos explicáveis podem ser descritos como sendo passíveis de uma explicação *post hoc* mais ou menos detalhada dos seus processos internos. A literatura explorada autoriza a conclusão de que o conceito de explicabilidade, além de impreciso, é limitado, conduzindo ao oferecimento de informações frequentemente pouco confiáveis, irrelevantes e até mesmo enganosas sobre o processo pelo qual um algoritmo chega a determinada decisão.

No presente artigo, argumentei no sentido da insuficiência da explicabilidade como critério para assegurar, em relação a modelos de IA para tomada de decisão judicial, a observância do devido processo legal, propondo que modelos interpretáveis seriam mais adequados. Entendida como conceito cujo conteúdo depende do domínio em que aplicado, a interpretabilidade precisa ser delineada, na construção de um modelo algorítmico, de acordo com o seu propósito. No caso de modelos interpretáveis de tomada de decisão judicial, a aderência ao devido processo legal supõe a produção de decisões nas quais identificados, como critérios legitimadores, legalidade, consistência interna e externa, e compatibilidade, nos moldes descritos acima.

Se o uso de ferramentas de inteligência artificial para tomada de decisão judicial parece inescapável, seus efeitos dependem de como moldaremos o ecossistema normativo no qual essas ferramentas podem operar. Nessa seara, o melhor que se pode esperar são decisões mais justas e robustas sobre direitos e deveres. Para que isto se torne realidade, é imprescindível a incorporação de parâmetros assecuratórios do devido processo legal na própria estrutura das ferramentas de IA empregadas na prática decisória.

6. Referências

ABITEBOUL, Serge; DOWEK, Gilles. **The Age of Algorithms**. Cambridge University Press, 2020.

318 CITRON, Danielle Keats. Technological Due Process. **Washington University Law Review**, v. 85, n. 6, 2008, p. 1249.

ASH, Elliot. **Judge, Jury, and EXEcute file: the brave new world of legal automation.** Londres: Social Market Foundation, CAGE Research Centre, 12 jun. 2018. Disponível em: https://warwick.ac.uk/fac/soc/economics/research/centres/cage/cage-final-elliott_ash.pdf. Acesso em: 10 out. 2023.

BERWICK, Robert C.; CHOMSKY, Noam. **Why Only Us: Language and Evolution.** Cambridge: MIT Press, 2016.

BRASIL. **Constituição da República Federativa do Brasil de 1988.** Brasília: Presidência da República, 2023.

BRASIL. **Lei nº 13.709/2018.** Brasília: Presidência da República, 2023.

CANALLI, Rodrigo L. Artificial intelligence and the model of rules: better than us? **AI and Ethics**, Vol. 1, n. 3, Ago. 2023.

CARDOSO, Evorah Lusci; FANTI, Fabíola. **Movimentos Sociais e Direito: o Poder Judiciário em Disputa.** In SILVA, Felipe Gonçalves. Manual de Sociologia Jurídica. 2ª ed. São Paulo: Saraiva, 2017.

CITRON, Danielle Keats. Technological Due Process. **Washington University Law Review**, vol. 85, n. 6, 2008.

CRAWFORD, Kate; SCHULTZ, Jason. AI systems as state actors. **Columbia Law Review**, New York, v. 119, n. 7, 2019.

DOSHI-VELEZ, Finale; KIM, Been. **Towards a Rigorous Science of Interpretable Machine Learning.** arXiv.org. 2017.

EUROPEAN UNION. **Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain Union legislative acts.** European Parliament, 2023.

GARCEZ, Artur S. d'Avila, GABBAY, Dov M., LAMB, Luis C. A neural cognitive model of argumentation with application to legal inference and decision making. **Journal of Applied Logic**, vol. 12, n. 2, Jun. 2014.

HART, H.L.A. **The Concept of Law.** Oxford University Press, Oxford, 2012.

HILDEBRANDT, Mireille. Law as computation in the era of artificial intelligence: speaking law to the power of statistics. **University of Toronto Law Journal**, vol. 68, n. 1, 2018.

JAUHAR, Ameen; MISRA, Vaidehi; SENGUPTA, Arghya; CHAKRABARTI, Partha P.; GHOSH, Saptarishi; GHOSH, Kripabandhu. **Responsible AI for the Indian Justice System: A**

Strategy Paper. Vidhi Centre for Legal Policy, TCG-Crest, 2021.

OGONJO, Florence; GITONGA, Joseph Theuri; WAIREGI, Angeline; RUTENBERG, Isaac. Utilizing AI to Improve Efficiency of the Environment and Land Court in the Kenyan Judiciary: Leveraging AI Capabilities in Land Dispute Cases in the Kenyan Environmental and Land Court System. *In: International Workshop on AI and Intelligent Assistance for Legal Professionals in the Digital Workplace (LegalAIIA 2021)*, 2, 2021, São Paulo. **Joint Proceedings of the Workshops on Automated Semantic Analysis of Information in Legal Text (ASAIL 2021) & AI and Intelligent Assistance for Legal Professionals in the Digital Workplace (LegalAIIA 2021)**, São Paulo: CEUR-WS, 2021, p 59-68.

PASQUALE, Frank. **The Black Box Society: The Secret Algorithms That Control Money and Information**. Harvard University Press, 2015.

REICH, Rob; SAHAMI, Mehran; WEINSTEIN, Jeremy M. **System Error: Where Big Tech Went Wrong and How We Can Reboot**. Londres: Hodder and Stoughton, 2021.

RUDIN, Cynthia. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. **Nature Machine Intelligence**, vol. 1, n. 5, mai. 2019.

SALOMÃO, Luis Felipe (Coord.). **Inteligência Artificial: tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário brasileiro**. 2ª ed. Rio de Janeiro: FGV, 2022.

SELBST, Andrew D.; BAROCAS, Solon. The Intuitive Appeal of Explainable Machines. **Fordham Law Review**, New York, v. 87, n. 3, Nov. 2018.

SHI, Changqing; SOURDIN, Tania; LI, Bin. The Smart Court: A New Pathway to Justice in China? **International Journal for Court Administration**, [s. l.], v. 12, n. 1, p. 4, Mar. 2021. Disponível em: <https://iacajournal.org/articles/10.36745/ijca.367>. Acesso em: 10 out. 2023.

UNIÃO EUROPEIA. **Regulamento Geral sobre Proteção de Dados (RGDP)**. Parlamento Europeu, 2016.

WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. **International Data Privacy Law**, vol. 7, n. 2, 2017.

WALDRON, Jeremy. **A dignidade da legislação**. São Paulo: Martins Fontes, 2003.

AVALIAÇÃO DE IMPACTO ALGORÍTMICO NA REGULAÇÃO DE INTELIGÊNCIA ARTIFICIAL: ELEMENTOS ESTRUTURAIS SOBRE ESSE INSTRUMENTO DE ACCOUNTABILITY

*ALGORITHMIC IMPACT ASSESSMENT ON ARTIFICIAL
INTELLIGENCE REGULATION: STRUCTURAL
ELEMENTS ABOUT THIS ACCOUNTABILITY
INSTRUMENT*

João Araújo Monteiro Neto³¹⁹

Isabelle Brito Bezerra Mendes³²⁰

Luis Henrique de Menezes Acioly³²¹

Matheus Fernandes da Silva³²²

Resumo

O presente trabalho parte da percepção de que a regulação de Inteligência Artificial pressupõe a cooperação entre agente regulador e regulado na delimitação de práticas preventivas e responsivas, bem como que a inserção de um sistema inteligente deve ser acompanhada da previsibilidade dos riscos que dele decorrem. Nesse panorama, um instrumento de regulação de IA é a Avaliação de Impacto Algorítmico (AIA), que,

319 PhD em Direito pela Universidade de Kent no Reino Unido. Mestre em Direito Constitucional pela Universidade de Fortaleza. Professor do Centro de Ciências Jurídicas da Universidade de Fortaleza. Advogado. Presidente da Comissão de Direito Digital da OAB/CE. Contato: joaoneto@unifor.br.

320 Advogada. Tax Consultant na EY. Pós-Graduada em Proteção de Dados e Governança Digital pela UNIFOR. Mestranda em Direito Constitucional pela UFC. Pesquisadora no Grupo de Estudos em Tecnologia, Informação e Sociedade (GETIS). Contato: isabellemendes06@gmail.com.

321 Graduando em Direito (UniRuy). Curador de Pesquisa e pesquisador no Grupo de Estudos em Tecnologia, Informação e Sociedade (GETIS - Unifor). *Data Protection Officer* e Vice-Presidente do Laboratório de Inovação e Direitos Digitais da UFBA (LABID²). Contato: acioly10@gmail.com.

322 Mestrando em Direito (Ufersa). Advogado (OAB/RN). Pesquisador no Grupo de Estudos em Tecnologia, Informação e Sociedade (GETIS). Contato: mateuszenan@gmail.com.

no âmbito de uma moldura de correção, estabelece a necessidade de prestação de contas acerca das consequências desse sistema na demarcação de danos. Dessa forma, a pesquisa busca compreender a inserção desse instrumento na regulação de IA e identificar seus elementos estruturantes a partir da noção da prestação de contas como um mecanismo regulatório relacional, investigando-se, ainda, a correlação desses elementos com a visão legislativa do tema expressa no Projeto de Lei nº 2338/2023. Para tanto empreendeu-se uma breve revisão de literatura, comparativamente com o texto do Projeto de Lei. Concluiu-se que a AIA tem por efeito o direcionamento da atuação do agente regulador e do regulado na interação regulatória e responsividade de suas atividades. Concluiu-se também que, na minuta da proposta legislativa, aplicou-se uma visão institucional deste instrumento sem observância da noção de fórum público de tomadas de contas e descentralização policêntrica da regulação. Reforça-se a necessidade de atualização normativa da proposta legislativa para albergar a Governança Social de sistemas de IA.

Palavras-chave: Inteligência Artificial; Avaliação de Impacto Algorítmico; Correção; Accountability; Regulação Policêntrica Descentralizada.

Abstract

The present work starts from the perception that the regulation of Artificial Intelligence presupposes the cooperation between the regulatory agent and the regulated in the delimitation of preventive and responsive practices, as well as that the insertion of an intelligent system must be accompanied by the predictability of the risks that result from it. In this scenario, an AI regulation instrument is the Algorithmic Impact Assessment (AIA), which, within a framework of co-regulation, establishes the need for accountability regarding the consequences of this system in the demarcation of damages. In this way, the research seeks to understand the insertion of this instrument in the regulation of AI and to identify its structuring elements from the notion of accountability as a relational regulatory mechanism, also investigating the correlation of these elements with the legislative view of the theme expressed in Bill No. 2338/2023. To this end, a brief literature review was undertaken, compared with the text of the Bill. It was concluded that the AIA has the effect of directing the performance of the regulatory agent and the regulated in the regulatory interaction and responsiveness of their activities. It was also concluded that, in the draft of the legislative proposal, an institutional view of this instrument was applied without observing the notion of a public forum for accountability and polycentric decentralization of regulation. It reinforces the need for normative updating of the legislative proposal to accommodate the Social Governance of AI systems.

Keywords: Artificial Intelligence; Algorithmic Impact Assessment; Coregulation; Accountability; Decentralized Polycentric Regulation.

1. INTRODUÇÃO

No contexto da crescente utilização de sistemas de Inteligência Artificial, a discussão por seu uso ético e como instrumento de empoderamento humano tem norteadado diversas abordagens nas mais diversas áreas científicas, incluindo as dedicadas à governança e regulação. A utilização de sistemas de IA para tomada de decisões que afetem a vida humana tem o potencial de repetir vieses discriminatórios do ambiente em que se inserem³²³, com implicações em direitos fundamentais. A dificuldade de compreensão dos algoritmos aliada à malversação dos direitos de propriedade intelectual pode conduzir a uma opacidade que beneficia a sobreposição de interesses econômicos a direitos humanos³²⁴.

A atuação do Estado na regulação de Inteligência Artificial, contudo, encontra barreiras na elevada versatilidade de sistemas inteligentes, bem como a partir da constante atualização e inovação tecnológica³²⁵. Buscando estabelecer um modelo cooperativo, alguns ordenamentos jurídicos têm previsto a obrigação de elaboração de uma Avaliação de Impacto Algorítmico, deferindo ao agente regulado a condução de um processo verificação das consequências desse sistema no complexo de direitos titularizados pelos possíveis afetados.

Dessa forma, o presente trabalho tem por objetivo compreender a inserção da Avaliação de Impacto Algorítmico no sistema de Regulação de Inteligência Artificial no que se refere aos seus elementos estruturantes, a partir da noção da prestação de contas como uma eminentemente dialógica. Busca-se, ainda, compreender a abordagem regulatória deferida às tecnologias contemporâneas e os elementos estruturantes da relação entre agente regulador e agente regulado, bem como entender como esses contornos são aplicáveis a esse instrumento de *accountability* em outros sistemas e na propositura legislativa brasileira de regulação de Inteligência Artificial, que tramita atualmente no Senado Federal.

Para a consecução desse objetivo, metodologicamente, empreendeu-se uma breve revisão de literatura, materializada por uma pesquisa bibliográfica de cunho descritivo e natureza qualitativa, em que se preconizou a verificação com o texto normativo e diálogo entre autores das diferentes áreas das ciências sociais. O presente trabalho se estrutura em três seções e as considerações finais. A primeira seção investiga a regulação de tecnologia em correlação com *accountability*. A segunda seção aborda os elementos estruturais da Avaliação de Impacto Algorítmico a partir da noção de prestação de contas. A terceira, por sua vez, explicita a visão institucional desse instrumento na Minuta de Substitutivo aos Projetos de Lei de IA.

323 DUARTE, ALAN. **A Antidiscriminação no contexto da inteligência artificial**: possibilidades de governança mediante a normatização de algoritmos. Fortaleza: Mucuripe, 2021.

324 PASQUALE, Frank. **The Black Box Society**. Cambridge: Harvard Unity Press, 2015.

325 GABRIEL, Martha. **Inteligência Artificial**: Do zero ao metaverso. Rio de Janeiro: Forense, 2022.

2. AVALIAÇÃO DE IMPACTO ALGORÍTMICO E SEU PAPEL PARA A PRESTAÇÃO DE CONTAS EM SISTEMA INTELIGENTES

A regulação de Inteligência Artificial no ordenamento brasileiro ganhou novos contornos a partir do aporte de conhecimento da Comissão de Juristas Responsável por Subsidiar a Elaboração de Substitutivo sobre Inteligência Artificial no Brasil (CJUSBIA). O Relatório Final apresentado pela Comissão de Jurista teve seu ingresso no processo legislativo a partir do Projeto de Lei nº 2338/2023,³²⁶ que o incorporou. O PL 2338 prevê em seu art. 22 e seguintes a obrigação de determinados agentes que desenvolvem ou usam artefatos de inteligência artificial em elaborar uma Avaliação de Impacto Algorítmico, quando o sistema for considerado de alto risco em avaliação preliminar.

A Avaliação de Impacto Algorítmico (AIA), embora seja instrumento cuja aparição no legislativo Brasil tenha se dado no PL nº 2338, já era previsto em atos normativos da IA em outros sistemas jurídicos, notadamente no *Artificial Intelligence and Data Act* (AIDA) canadense. Ante a ausência de definição no corpo da propositura legislativa³²⁷, busca-se na literatura e, na prática internacional um sucedâneo à sua conceituação e posicionamento no horizonte das teorias regulatórias.

2.1. Linhas gerais da regulação de tecnologia sob o prisma da accountability

A esse respeito, Watkins et al.³²⁸ apontam que o termo “Algorithmic Impact Assessment” tem sido utilizado em diversos contextos para fazer alusão a processos e documentações que fazem parte do “*toolkit*” da regulação de IA, em vistas a documentar

326 BRASIL. **Relatório final da Comissão de Juristas responsável por subsidiar a elaboração de Substitutivo sobre Inteligência Artificial instituída pelo Ato do Presidente do Senado n. 4, de 2022, aprovado em 1º de dezembro de 2022.** Senado Federal. Coordenação de Comissões Especiais, Temporárias e Parlamentares de Inquérito. Brasília, 2022. Disponível em: <https://www.stj.jus.br/sites/portalp/SiteAssets/documentos/noticias/Relato%CC%81rio%20final%20CJSUBIA.pdf>. Acesso em: 05 mar. 2023.

327 Semelhantemente, o Artificial Intelligence and Data Act (AIDA) canadense não tratou da conceituação da Avaliação de Impacto Algorítmico, apenas estabelecendo que “A person who is responsible for an artificial intelligence system must, in accordance with the regulations, assess whether it is a high-impact system”. Disponível em: <https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading>. Acesso em: 05 mar. 2023.

328 WATKINS, Elizabeth Anne; MOSS, Emanuel; METCALF, Jacob; SINGH, Ranjit; ELISH, Madeleine Clare. Governing algorithmic systems with impact assessments: six observations. **Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society**, Nova York, mai 19–21, p.1010-1021, 2021, Virtual Event. Disponível em: <https://dl.acm.org/doi/10.1145/3461702.3462580>. Acesso em: 05 mar. 2023.

o desenvolvimento e impacto do sistema algorítmico e prevenir e mitigar potenciais danos.

Semelhantemente, Selbst³²⁹ indica que a Avaliação de Impacto Algorítmico constitui estratégia regulatória para abordar e corrigir potenciais danos algorítmicos, tendo-se noção que o criador do sistema de IA é quem melhor tem expertise e acesso a informações do projeto. Portanto, o agente de IA³³⁰ teria melhores condições de conduzir e documentar uma avaliação dos impactos daquele sistema inteligente.

A regulação de tecnologia tem substancial relevo no ambiente regulatório, posto que se difere especialmente quanto aos desafios que ultrapassam a mera dificuldade de acompanhar o ambiente de inovação³³¹. Moses³³² elenca, ao lado do problema de ritmo, a desconexão regulatória e o dilema de Collingridge como desafios regulatórios nesse contexto. A desconexão regulatória ocorre quando há um vazio normativo para a nova tecnologia ou há a inadequação desta com as práticas de mercado já reguladas no setor³³³. Por sua vez, o dilema de Collingridge consiste na dificuldade de delimitação do momento da regulação da tecnologia sem que ocorra um sufocamento da inovação³³⁴. De todo modo, a regulação de tecnologia tem se descortinado num ambiente de incertezas, demonstrando a necessidade de uma estratégia regulatória que realize a avaliação dos riscos sociais e regulatórios de forma a legitimar a interferência estatal na atividade privada³³⁵.

A estratégia de interação regulatória em torno de delegação das funções do Estado, no todo ou em parte, ao agente regulado, em si mesmo ou coletivamente considerado, não é exclusividade das normas do século XXI, mas constituiu suporte à Teoria da Regulação Responsiva, inaugurada por Ayres e Braithwaite³³⁶, que potencializou cooperação entre

329 SELBST, Andrew D. An institutional view of algorithmic impact assessment. **Harvard Journal of Law & Technology**, Massachusetts, v. 35, n. 1, p.117-173, 2021. Disponível em: <https://jolt.law.harvard.edu/assets/articlePDFs/v35/Selbst-An-Institutional-View-of-Algorithmic-Impact-Assessments.pdf>. Acesso em: 05 mar. 2023.

330 Na forma do Projeto de Lei nº 2338/2023, seriam “agentes de inteligência artificial”, os fornecedores (desenvolvedores) e operadores de sistemas de IA.

331 KELLER, Clara Iglesias. Regulação nacional de serviços na Internet: exceção, legitimidade e o papel do Estado. 2019. 300 f. Tese (Doutorado em Direito Público) - Faculdade de Direito. Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2019.

332 MOSES, Lyria Bennett. How to Think about law, regulation and technology: problems with ‘technology’ as a regulatory target. **Innovation and Technology**, 1-20(2013). (2013) 5(1). UNSW Law Research Paper n. 2014-30. Disponível em: <http://ssrn.com/abstract=2464750>. Acesso em: 11 jun. 2023.

333 MOSES, Lyria Bennett. How to Think about law, regulation and technology: problems with ‘technology’ as a regulatory target. **Innovation and Technology**, 1-20(2013). (2013) 5(1). UNSW Law Research Paper n. 2014-30. Disponível em: <http://ssrn.com/abstract=2464750>. Acesso em: 11 jun. 2023.

334 *Ibidem*.

335 BLACK, Julia; MURRAY, Andrew. Regulation AI and machine learning: setting the regulatory agenda. **European Journal of Law and Technology**, [S.l.], v. 10, n. 3, 2019. Disponível em: <https://www.ejlt.org/index.php/ejlt/article/view/722>. Acesso em: 05 mar. 2023.

336 AYRES, Ian; BRAITHWAITE, John. **Responsive Regulation: transcending the deregulation debate**. Oxford: Oxford University Press, 1992.

entes privados e agentes públicos no delineamento de novos mercados, a partir da definição de responsividade.

Para além do protagonismo estatal na abordagem regulatória, no denominado sistema Comando e Controle, ou heterorregulação, Baldwin e Cave³³⁷ citam as estratégias de autorregulação, fundamentada exclusivamente na organização dos entes privados na delimitação de estruturas conformativas³³⁸ e correção, ou autorregulação regulada, na qual há delimitação de funções entre o Estado e o agente privado nesse contexto.

Keller³³⁹ explica que a autorregulação regulada se dá através delegação tarefas estatais aos agentes privados, mediante fixação de parâmetros pelo Estado. A regulação responsiva, nesse contexto, se descortina como uma forma de interação regulatória pendular e em camadas, pelo que a forma da resposta do Estado à conduta do agente privado é proporcional à sua desídia ou benesse, o que gerou a formulação da pirâmide de constrangimento e da pirâmide de estratégia regulatória³⁴⁰.

Em busca de tornar a teoria da regulação responsiva aplicável o suficiente para uma “*regulação ótima*”, Baldwin e Black³⁴¹ sistematizam que para se ter comportamentos “*realmente responsivos*”, é necessário que a responsividade, isto é a forma da interação regulatória e reação estatal, leve em consideração: (i) a própria estrutura operacional; (ii) o ambiente institucional do regime regulatório; (iii) as diferentes lógicas dos instrumentos e estratégias regulatórias; (iv) o próprio desempenho desse regime, e (v) a mudança em algum desses elementos.

A responsividade ganha corpo a partir de dinâmicas de prestação de contas, que na visão de Quelle³⁴² deve ser observada sob o enfoque dos desdobramentos na autonomia dos afetados³⁴³. A proceduralização do direito regulatório põe a prestação de contas como o

337 BALDWIN, Robert; CAVE, Martin. **Understanding Regulation: theory, strategy, and practice**. Oxford University Press, 1999.

338 Nesse contexto, Black (2001) aponta que a autorregulação se distingue da correção na medida em que, além da ausência de participação governamental, há equiparação de normas não legais e de soft-law. Ressalte-se que, conforme explica Keller (2019), a autorregulação ganhou força através dos sistemas de Internet, e ausência de controle estatal na delimitação dos padrões técnicos e normativos da rede das redes, que se delineou a partir de órgãos de abrangência global como ICANN, IETF e IGF.

339 KELLER, Clara Iglesias. **Regulação nacional de serviços na Internet: exceção, legitimidade e o papel do Estado**. 2019. 300 f. Tese (Doutorado em Direito Público) - Faculdade de Direito. Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2019.

340 AYRES, Ian; BRAITHWAITE, John. **Responsive Regulation: transcending the deregulation debate**. Oxford: Oxford University Press, 1992.

341 BALDWIN, Robert; BLACK, Julia. Really Responsive Regulation. **Revista de Direito Administrativo**, [S.l.], v. 281, n. 2, p. 45-90, 2022. Disponível em: <https://bibliotecadigital.fgv.br/ojs/index.php/rda/article/view/86044>. Acesso em: 05 mar. 2023.

342 QUELLE, Claudia. Privacy, proceduralism and self-regulation in data protection law. **Teoria e Critica della Regolazione Sociale / Theory and Criticism of Social Regulation**, [S.l.], v. 1, n. 14, p. 89-106, 2020. Disponível em: <https://mimesisjournals.com/ojs/index.php/tcrs/article/view/113>. Acesso em: 05 mar. 2023.

343 A autora (QUELLE, 2020) parte da discussão acerca da autonomia informacional e accountability no *General Data Protection Regulation – GDPR* para dimensionar os efeitos deste princípio, raciocínio que pode ser estendido para o contexto das decisões automatizadas, e conseqüentemente dos sistemas inteligentes.

cerne da concretização da regulação, de forma que o processo regulatório se baseia no fornecimento de instrumentos processuais de consolidação dos objetivos da regulação³⁴⁴.

Nesse sentido, Bioni³⁴⁵ adota um ponto de vista obrigacional dialógico entre os polos ativo e passivo da prestação de contas, entendendo a *accountability* como um conceito eminentemente relacional. Para o autor³⁴⁶, a *accountability* dirige-se a um fenômeno de poder, posto que a força de um agente é contingenciada pelo escrutínio público, a ponto de haver uma divisão do trabalho entre quem é o detentor do poder (*power-holder*) e quem pode sofrer as consequências de um mal comportamento deste (*account-holder*). Black³⁴⁷ assinala a força social da *accountability* na modulação de comportamentos do *power-holder* e, do *account-holder*, através de uma estrutura policêntrica da regulação, na qual constantemente os papéis se entrelaçam³⁴⁸, na alçada do escrutínio público.

A prestação de contas a um fórum público³⁴⁹ pressupõe a existência de regras específicas para aposição de instrumentos de *accountability*, definindo os elementos da relação obrigacional³⁵⁰, notadamente quanto aos sujeitos que compõem este fórum, a quem incumbirá desincumbir-se da responsabilidade de prestar contas, bem como ao objeto e circunstâncias que irão nortear a sua tomada e julgamento³⁵¹. Bovens³⁵² cita que para que esse fórum público de prestação de conta possa existir, é necessária a existência de três etapas: é necessário que o agente seja obrigado a informar ao fórum sobre sua conduta; deve haver possibilidade de o fórum questionar a adequação da informação ou legitimidade da conduta; e o fórum deve poder julgar sua conduta, resultando em consequências para o agente.

O suporte fático para a proceduralização do direito regulatório no escopo da regulação baseada em riscos é o próprio risco³⁵³. O risco, entendido como incerteza quantificável

344 BLACK, Julia. Proceduralisation and polycentric regulation. **Revista Direito GV**, São Paulo, v. 1.5, n. Especial 1, p. 99-130, 2005. Disponível em: <https://bibliotecadigital.fgv.br/ojs/index.php/revdireitogv/article/view/35248>. Acesso em: 05 mar. 2023.

345 BIONI, Bruno Ricardo. **Regulação e Proteção de Dados Pessoais: O Princípio da Accountability**. Rio de Janeiro: Forense, 2022.

346 Ibidem.

347 BLACK, Julia. Decentring Regulation: Understanding the role of regulation and self-regulation in a “Post-Regulatory” world. **Current Legal Problems**, [S.l.], v. 54, n. 1, p. 103-146, 2001. Disponível em: <https://doi.org/10.1093/clp/54.1.103>. Acesso em: 05 mar. 2023.

348 BLACK, Julia, 2005. *Op. cit.*

349 BOVENS, Mark. Analysing and assessing accountability: a conceptual framework. **European Law Journal**, [S.l.], vol. 13, n. 4, p. 447–468, 2007.

350 BLACK, 2001. *Op. cit.*

351 BIONI, Bruno Ricardo. *Op. cit.*

352 BOVENS, Mark. *Op. cit.*

353 BLACK, Julia. Proceduralisation and polycentric regulation. **Revista Direito GV**, São Paulo, v. 1.5, n. Especial 1, p. 99-130, 2005. Disponível em: <https://bibliotecadigital.fgv.br/ojs/index.php/revdireitogv/article/view/35248>. Acesso em: 05 mar. 2023.

ou probabilidade de um malefício ocorrer³⁵⁴, atribui objeto e justificativa da regulação, ao passo que também desempenha o papel de estruturar processos regulatórios³⁵⁵, de forma que a *accountability* se torna o vetor de cognição de riscos e modulação da interação regulatória na alçada do fórum público de prestação de contas³⁵⁶. A caixa de ferramentas da regulação de IA, para além de concretizar uma estrutura de coleta de informações e cognição dos riscos, estimula a governança social de algoritmos, notadamente através das avaliações de impacto³⁵⁷.

2.2. Elementos constitutivos da Avaliação de Impacto Algorítmico na perspectiva obrigacional da *accountability*

Nesse horizonte regulatório, a Avaliação de Impacto Algorítmico busca modelar os comportamentos dos agentes de Inteligência Artificial, a partir da prestação de contas quanto aos riscos que o seu sistema pode oferecer, bem como do agente regulador, que deverá ponderar suas ações com base nos resultados apresentados³⁵⁸. Watkins *et al.*³⁵⁹ apontam que para desenvolvimento de polos de responsabilização consentâneos à tecnologia de IA, o estabelecimento de uma Avaliação de Impacto Algorítmico pode definir padrões para avaliação de desempenho de tais sistemas, apontando a responsabilidade do agente estatal na sua intervenção, e do ente privado na prevenção e mitigação. A discussão regulatória sobre riscos da Inteligência Artificial não se restringe aos riscos à saúde ou segurança cientificamente calculados, mas alcança também os

354 ZANATTA, Rafael A. F. Proteção de dados pessoais como regulação do risco: uma nova moldura teórica? *In: I Encontro da Rede de Pesquisa em Governança da Internet*. Rio de Janeiro, 2017. **Anais**. Rio de Janeiro: Rede de Pesquisa em Governança da Internet, 2018. p. 175-193.

355 BLACK, Julia. The role of risk in regulatory processes. *In: BALDWIN, Robert; CAVE, Martin; LODGE, Martin (ed.). The Oxford Handbook of Regulation*. Oxford: Oxford University Press, 2010. p.302-348.

356 LODGE, Martin; STIRTON, Lindsay. Accountability in the Regulatory State. *In: BALDWIN, Robert; CAVE, Martin; LODGE, Martin (ed.). The Oxford Handbook of Regulation*. Oxford: Oxford University Press, 2010. p. 349-370.

357 KAMINSKI, Margot; MALGIERI, Gianclaudio. Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations. **U of Colorado Law Legal Studies Research Paper**, [S.l.], n. 19-28, 2020. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3456224. Acesso em 24 jun. 2023.

358 SELBST, Andrew D. An institutional view of algorithmic impact assessment. **Harvard Journal of Law & Technology**, Massachusetts, v. 35, n. 1, p.117-173, 2021. Disponível em: <https://jolt.law.harvard.edu/assets/articlePDFs/v35/Selbst-An-Institutional-View-of-Algorithmic-Impact-Assessments.pdf>. Acesso em: 05 mar. 2023.

359 WATKINS, Elizabeth Anne; MOSS, Emanuel; METCALF, Jacob; SINGH, Ranjit; ELISH, Madeleine Clare. Governing algorithmic systems with impact assessments: six observations. **Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society**, Nova York, mai 19–21, p.1010-1021, 2021, Virtual Event. Disponível em: <https://dl.acm.org/doi/10.1145/3461702.3462580>. Acesso em: 05 mar. 2023.

chamados riscos sociais, que englobam esferas de discriminação³⁶⁰.

A previsibilidade e prevenção dos danos oriundos da inserção em mercado de um produto de inteligência artificial se amolda a uma regulação ex ante, de forma que Watkins et al.³⁶¹ associam a Avaliação de Impacto Algorítmico a outros instrumentos de *accountability* em regulações baseadas no risco, como a Avaliação de Impacto Ambiental e a Avaliação de Impacto Fiscal. Seguindo na mesma linha, Selbst³⁶² aduz que, após a edição da *National Environmental Policy Act* – NEPA estadunidense, de 1969, as avaliações de impacto constaram em diversos sistemas de regulação, como a ambiental, de Direitos Humanos, Proteção de Dados e Políticas de Tecnologia.

De forma semelhante, a Avaliação de Impacto à Proteção de Dados (AIPD ou “DPIA”) como prescrito no *General Data Protection Officer* (GDPR) - ou Relatório de Impacto à Proteção de Dados Pessoais (RIPD), previsto na Lei Geral de Proteção de Dados -, é instrumento de *accountability* para tratamentos de dados pessoais que possam gerar riscos às liberdades civis e direitos fundamentais dos titulares, cuja elaboração deve ser feita prioritariamente em momento anterior à inserção de novas tecnologias e processos no mercado³⁶³, representando um instrumento da abordagem regulatória baseada no risco – *risk-based approach*³⁶⁴.

O processo de risquificação da proteção de dados, nesse contexto, se configura a partir dos elementos de: (i) participação de sociedade civil no diálogo com Autoridades de Proteção de Dados a fim de definir riscos do fluxo informacional e instrumentos de tutela coletiva; (ii) instrumentos de regulação ex ante, impondo obrigações aos controladores na identificação de riscos a direitos fundamentais; (iii) existência e utilização de metodologias de gestão de risco e proporcionalidade com imunidades jurídicas para o progresso da tecnologia³⁶⁵.

Quanto à forma de elaboração da Avaliação de Impacto Algorítmico, Selbst³⁶⁶ delimita

360 GELLERT, Raphael. The role of the risk-based approach in the General data protection Regulation and in the European Commission’s proposed Artificial Intelligence Act. Business as usual?. **Journal of Ethics and Legal Technologies**, [S.l.], v. 3, n. 2, p. 15-33, 2021. Disponível em: <https://doi.org/10.14658/pupj-jelt-2021-2-2>. Acesso em: 19 jun. 2023.

361 *Ibidem*.

362 SELBST, Andrew D. An institutional view of algorithmic impact assessment. **Harvard Journal of Law & Technology**, Massachusetts, v. 35, n. 1, p.117-173, 2021. Disponível em: <https://jolt.law.harvard.edu/assets/articlePDFs/v35/Selbst-An-Institutional-View-of-Algorithmic-Impact-Assessments.pdf>. Acesso em: 05 mar. 2023.

363 GOMES, Maria Cecília O. Relatório de impacto à proteção de dados: uma breve análise da sua definição e papel na LGPD. **Revista do Advogado**, São Paulo, v. 39, n. 144, p. 174–183, 2019.

364 ZANATTA, Rafael A. F. Proteção de dados pessoais como regulação do risco: uma nova moldura teórica? In: I Encontro da Rede de Pesquisa em Governança da Internet. Rio de Janeiro, 2017. **Anais**. Rio de Janeiro: Rede de Pesquisa em Governança da Internet, 2018. p. 175-193.

365 *Ibidem*.

366 SELBST, Andrew D. An institutional view of algorithmic impact assessment. **Harvard Journal of Law & Technology**, Massachusetts, v. 35, n. 1, p.117-173, 2021. Disponível em: <https://jolt.law.harvard.edu/assets/articlePDFs/v35/Selbst-An-Institutional-View-of-Algorithmic-Impact-Assessments.pdf>. Acesso em: 05 mar. 2023.

três modelos: (1) Modelos Baseados na Avaliação de Impacto Ambiental, oriunda do NEPA (ora denominados modelos “NEPA”); (2) Modelos baseados na Avaliação de Impacto à Proteção de Dados (ora denominados modelos “DPIA”); (3) Modelos de questionário.

Os modelos baseados na Avaliação de Impacto Ambiental implicam em uma validação altamente detalhada dos riscos e circunstâncias potencialmente danosas, pressupondo respostas completas a perguntas abertas³⁶⁷, à disposição de um fórum público de agentes, mediante um sistema de notificação e audiências³⁶⁸. Esse modelo melhor se relaciona com a teoria da governança nodal, ou governança em rede, que fornece substrato intelectual ao processo dialógico da *accountability*³⁶⁹. A governança nodal constitui forma de delimitação do fórum público de tomada de contas através de nodos de poder, isto é, conjunto de agentes dos diversos stakeholders que mobilizam recursos e conhecimento para a permanente avaliação da regulação³⁷⁰.

Os modelos baseados no DPIA se relacionam com o objetivo de formulação para abranger somente os riscos aos direitos e liberdades da pessoa física, sem necessariamente uma estrutura de audiências públicas e diálogo com os possíveis afetados pela IA³⁷¹. A seu turno, o modelo baseado em questionário é utilizado pelo governo canadense, previsto na *Directive on Automated Decision-Making*³⁷², e é construído a partir de níveis de impacto das decisões automatizadas de instituições governamentais, apontando resumidamente os aspectos principais da avaliação, na forma de 60 perguntas³⁷³, conduzida após o funcionamento do sistema.

A partir da tônica da regulação *ex ante*, os modelos de Avaliação de Impacto Algorítmico baseados na Avaliação de Impacto Ambiental, do NEPA, e na Avaliação de Impacto à Proteção de Dados, do GDPR, se mostram preponderantes quando observada a abordagem baseada em risco que permeia as regulações de tecnologias contemporâneas³⁷⁴⁻³⁷⁵.

367 REISMAN, Dillon; SCHULTZ, Jason; CRAWFORD, Kate; WHITTAKER, Meredith. **Algorithmic Impact Assessment: a practical framework for public agency accountability.** AI NOW, 2018.

368 SELBST, Andrew D. An institutional view of algorithmic impact assessment. **Harvard Journal of Law & Technology**, Massachusetts, v. 35, n. 1, p.117-173, 2021. Disponível em: <https://jolt.law.harvard.edu/assets/articlePDFs/v35/Selbst-An-Institutional-View-of-Algorithmic-Impact-Assessments.pdf>. Acesso em: 05 mar. 2023.

369 BURRIS, Scott; DRAHOS, Peter; SHEARING, Clifford. Nodal governance. **Australian Journal of Legal Philosophy**, [S.l.], vol. 30, 2005. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=760928. Acesso em: 11 jun. 2023.

370 *Ibidem*.

371 SELBST, Andrew D. *Op. cit.*

372 CANADÁ. **Directive on Automated Decision-Making.** Canadá: [s.n.], 2021. Disponível em: <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592>. Acesso em: 05 mar. 2023.

373 SELBST, Andrew D. *Op. cit.*

374 KELLER, Clara Iglesias. **Regulação nacional de serviços na Internet: exceção, legitimidade e o papel do Estado.** 2019. 300 f. Tese (Doutorado em Direito Público) - Faculdade de Direito. Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2019.

375 BIONI, Bruno Ricardo. **Regulação e Proteção de Dados Pessoais: O Princípio da Accountability.** Rio de Janeiro: Forense, 2022.

Ponho-se em foco o modelo “NEPA”, Reisman et al³⁷⁶ sistematizam que a Avaliação de Impacto Algorítmico deve conter os seguintes elementos chave: (1) Realização de autoavaliação dos sistemas de decisão automatizada existentes e que irão ser implementados; (2) Os agentes devem estabelecer procedimentos que envolvam auditores externos para descobrir, medir e rastrear impactos ao longo do tempo; (3) Os agentes devem notificar o público sobre sua definição de decisão automatizada, sistemas existentes e circunstâncias das avaliações conduzidas; (4) Os agentes devem solicitar comentários do público e esclarecer dúvidas pendentes; (5) Os governos devem disponibilizar mecanismos de devido processo para os afetados contestarem avaliações inadequadas, ou uso injustificado do sistema.

2.3. Delimitação dos elementos da Avaliação de Impacto Algorítmico na minuta do Projeto de Lei nº 2338/2023

A Minuta do Projeto de Lei nº 2.338/2023 prevê como elemento causal para elaboração da Avaliação de Impacto Algorítmico a aferição preliminar de alto risco do sistema inteligente, cuja definição segue os parâmetros do seu artigo 18, prevendo a possibilidade alargamento das hipóteses por Autoridade Competente³⁷⁷. Nesse processo de *accountability* dialógica³⁷⁸, o *power-holder* escolhido na propositura legislativa para prestar contas através da Avaliação de Impacto Algorítmico é o agente da IA de alto risco.

É relevante pontuar que a eficácia da correção quanto à cognição do impacto algorítmico pressupõe uma atuação mais ativa do próprio *power-holder*, a quem a norma atribui certo grau de discricionariedade³⁷⁹⁻³⁸⁰. Em vista a evitar possíveis desvios de utilização dessa discricionariedade, Selbst³⁸¹ advoga por um processo contínuo

376 REISMAN, Dillon; SCHULTZ, Jason; CRAWFORD, Kate; WHITTAKER, Meredith. **Algorithmic Impact Assessment: a practical framework for public agency accountability.** AI NOW, 2018.

377 Embora o projeto de lei preveja a criação de uma “Autoridade Competente”, a minuta não delimitou a sua denominação, estrutura e natureza jurídica, deixando a cargo do Poder Executivo Federal o realizar, tendo em vista que a criação de órgãos da Administração Pública Federal somente pode se dar proposta legislativa de iniciativa do Presidente da República, na forma do art. 61 da Constituição Federal de 1988.

378 BOVENS, Mark. **Analysing and assessing accountability: a conceptual framework.** European Law Journal, [S.l.], vol. 13, n. 4, p. 447–468, 2007.

379 WATKINS, Elizabeth Anne; MOSS, Emanuel; METCALF, Jacob; SINGH, Ranjit; ELISH, Madeleine Clare. Governing algorithmic systems with impact assessments: six observations. **Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society**, Nova York, mai 19–21, p.1010-1021, 2021, Virtual Event. Disponível em: <https://dl.acm.org/doi/10.1145/3461702.3462580>. Acesso em: 05 mar. 2023.

380 BALDWIN, Robert; BLACK, Julia. Really Responsive Regulation. **Revista de Direito Administrativo**, [S.l.], v. 281, n. 2, p. 45-90, 2022. Disponível em: <https://bibliotecadigital.fgv.br/ojs/index.php/rda/article/view/86044>. Acesso em: 05 mar. 2023.

381 SELBST, Andrew D. An institutional view of algorithmic impact assessment. **Harvard Journal of Law & Technology**, Massachusetts, v. 35, n. 1, p.117-173, 2021. Disponível em: <https://jolt.law.harvard.edu/assets/articlePDFs/v35/Selbst-An-Institutional-View-of-Algorithmic-Impact-Assessments.pdf>. Acesso em: 05 mar. 2023.

de auditoria algorítmica por autoridade competente. Essa alternativa descortina-se juridicamente viável no ordenamento brasileiro ante a possibilidade de a Autoridade Nacional de Proteção de Dados promover auditorias para “verificação de aspectos discriminatórios em tratamento automatizado de dados pessoais”³⁸², nos termos do § 2º do art. 20 da LGPD, que é parâmetro para delimitação de alto risco em sistemas de inteligência artificial, conforme art. 18, alínea “c” do Projeto de Lei 2.338/2023.

Questão mais sensível é a delimitação da forma da prestação de contas e a ocupação do polo de *account-holder* nesse contexto. A minuta da propositura legislativa (art. 24, § 1º) prevê que a Avaliação de Impacto Algorítmico deve registrar, ao menos: (i) os riscos conhecidos e previsíveis à época da elaboração do sistema inteligente; (ii) os benefícios associados ao sistema; (iii) a probabilidade de efeitos adversos e a quantidade de pessoas possivelmente impactadas; (iv) a gravidade das consequências e as medidas de mitigação; (v) a lógica do funcionamento do sistema inteligente; (vi) o processo e resultado dos testes e avaliações para verificação de possíveis impactos a direitos; (vii) o treinamento e ações de conscientização dos riscos associados ao sistema; (viii) as medidas de mitigação e justificação do risco residual; e (ix) as medidas de transparência pública, especialmente quanto aos possíveis usuários desse sistema.

Contudo, é importante ressaltar que o Projeto de Lei não prevê a obrigatoriedade de elaboração da AIA de forma anterior à sua inserção no mercado, como o faz em relação à avaliação preliminar³⁸³, que serve à delimitação do nível de risco do sistema. A exigibilidade *ex ante* da Avaliação de Impacto Algorítmico é categórica no contexto da aplicação do princípio da prevenção, notadamente em face de alto risco. Ressalta-se que a *accountability* na regulação de IA tem o condão de amoldar o próprio princípio da precaução, impondo práticas responsáveis na cognição e mitigação dos riscos sociais e de danos³⁸⁴.

Essencialmente, o modelo “NEPA” não se amolda com perfeição às diretrizes apontadas na minuta do Projeto de Lei de IA, por ausência de expressa previsão da participação pública na elaboração da Avaliação de Impacto Algorítmico, que ficará a critério da autoridade competente:

Art. 24: (...) § 3º A autoridade competente poderá estabelecer outros critérios e elementos para a elaboração de avaliação de impacto, incluindo a participação dos diferentes segmentos sociais afetados,

382 BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados (LGPD). Brasília, DF, 2018.

383 “Art. 13. **Previamente** a sua colocação no mercado ou utilização em serviço, todo sistema de inteligência artificial passará por avaliação preliminar realizada pelo fornecedor para classificação de seu grau de risco, cujo registro considerará os critérios previstos neste capítulo” (grifos dos autores).

384 BIONI, Bruno; LUCIANO, Maria. O Princípio da Precaução na Regulação de Inteligência Artificial: Seriam as Leis de Proteção de Dados o seu Portal de Entrada?. In: BIONI, Bruno. (org.). **Proteção de dados: contexto, narrativas e elementos fundantes**. São Paulo: B. R. Bioni Sociedade Individual de Advocacia, 2021, p. 281-313.

conforme risco e porte econômico da organização.

Art. 26. Garantidos os segredos industrial e comercial, as conclusões da avaliação de impacto serão públicas, contendo ao menos as seguintes informações: I – descrição da finalidade pretendida para a qual o sistema será utilizado, assim como de seu contexto de uso e escopo territorial e temporal; II – medidas de mitigação dos riscos, bem como o seu patamar residual, uma vez implementada tais medidas; III – descrição da participação de diferentes segmentos afetados, caso tenha ocorrido, nos termos do § 3º do art. 24 desta Lei.

Nesse ponto, posto que prescindível a participação dos possíveis afetados no processo de elaboração da Avaliação, resta comprometida a definição de fórum público de prestação de contas³⁸⁵ nesse instrumento de *accountability*, bem como por ausência de expressa previsão de mecanismo de devido processo de contestação de avaliações inadequadas³⁸⁶. A estruturação normativa da Avaliação de Impacto Algorítmico na minuta do Projeto de Lei denota o papel de *account-holder* à Autoridade Competente, restringindo a tomada de contas ao Estado, em descompasso com a noção de regulação descentralizada policêntrica³⁸⁷.

3. CONSIDERAÇÕES FINAIS

Ante o exposto, é possível concluir que a Avaliação de Impacto Algorítmico prevista na Minuta do Projeto de Lei nº 2338/2023 apresenta-se como instrumento de delegação de espectro da atividade regulatória estatal ao agente regulado, na medida em que se lhe exige a adoção de práticas responsivas e sua respectiva prestação de contas. A *accountability* em sistemas corregulação tem a função de apontar uma relação obrigacional de prestação e tomada de contas, descentralizando os polos de *power-holder* e *account-holder* para o manejo de um fórum público de escrutínio recíproco. Nesse cenário, a proceduralização regulatória se concretiza em torno do risco, que define mandatos regulatórios e ampara processos de regulação. A cognição dos riscos no bojo da prestação de contas é matéria de especial relevo na regulação de tecnologia, encontrando esteio nas diversas ferramentas de regulação de IA, notadamente no contexto das avaliações de impacto.

385 BOVENS, Mark. **Analysing and assessing accountability**: a conceptual framework. *European Law Journal*, [S.l.], vol. 13, n. 4, p. 447–468, 2007.

386 REISMAN, Dillon; SCHULTZ, Jason; CRAWFORD, Kate; WHITTAKER, Meredith. **Algorithmic Impact Assessment**: a practical framework for public agency accountability. AI NOW, 2018.

387 BLACK, Julia. Proceduralisation and polycentric regulation. *Revista Direito GV*, São Paulo, v. 1.5, n. Especial 1, p. 99-130, 2005. Disponível em: <https://bibliotecadigital.fgv.br/ojs/index.php/revdireitogv/article/view/35248>. Acesso em: 05 mar. 2023.

A Avaliação de Impacto Algorítmico nesse processo dialógico tem o condão de modular a atuação do agente regulador e do agente regulado a partir do resultado e elementos documentados. Os modelos de AIA utilizados em outros sistemas regulatórios, de forma análoga à Avaliação de Impacto Ambiental, preconizam a participação dos afetados na estrutura dorsal desse instrumento, pondo em foco a noção de fórum público de *accountability*, no contexto de uma governança nodal ao redor do desenvolvimento da tecnologia.

Esse raciocínio não foi replicado na minuta do Projeto de Lei de IA, que além de não prever a participação dos possíveis afetados no processo de elaboração da Avaliação, não aponta mecanismos de contestação pública de avaliações inadequadas, restringindo a função de *account-holder* ao Estado, em contraposição à noção de descentralização policêntrica da regulação. Nesse panorama, a inserção desses mecanismos no texto do Projeto de Lei teria o condão de adequar a Avaliação de Impacto Algorítmico a necessária Governança Social dos sistemas de IA no contexto regulatório, reconhecendo os nodos de poder que envolvem a estrutura de concretização da regulação responsiva.

4. REFERÊNCIAS

AYRES, Ian; BRAITHWAITE, John. **Responsive Regulation: transcending the deregulation debate**. Oxford: Oxford University Press, 1992.

BALDWIN, Robert; BLACK, Julia. Really Responsive Regulation. **Revista de Direito Administrativo**, [S.l.], v. 281, n. 2, p. 45-90, 2022. Disponível em: <https://bibliotecadigital.fgv.br/ojs/index.php/rda/article/view/86044>. Acesso em: 05 mar. 2023.

BALDWIN, Robert; CAVE, Martin. **Understanding Regulation: theory, strategy, and practice**. Oxford: Oxford University Press, 1999.

BIONI, Bruno; LUCIANO, Maria. O Princípio da Precaução na Regulação de Inteligência Artificial: Seriam as Leis de Proteção de Dados o seu Portal de Entrada?. In: BIONI, Bruno (org.). **Proteção de dados: contexto, narrativas e elementos fundantes**. São Paulo: B. R. Bioni Sociedade Individual de Advocacia, 2021. p. 281-313.

BIONI, Bruno Ricardo. **Regulação e Proteção de Dados Pessoais: O Princípio da Accountability**. Rio de Janeiro: Forense, 2022.

BLACK, Julia. Decentring Regulation: Understading the role of regulation and self-regulation in a “Post-Regulatory” world. **Current Legal Problems**, [S.l.], v. 54, n. 1, p. 103-146, 2001. Disponível em: <https://doi.org/10.1093/clp/54.1.103>. Acesso em: 05 mar. 2023.

BLACK, Julia. Proceduralisation and polycentric regulation. **Revista Direito GV**, São

Paulo, v. 1.5, n. Especial 1, p. 99-130, 2005. Disponível em: <https://bibliotecadigital.fgv.br/ojs/index.php/revdireitogv/article/view/35248>. Acesso em: 05 mar. 2023.

BLACK, Julia. The role of risk in regulatory processes. In: BALDWIN, Robert; CAVE, Martin; LODGE, Martin (ed.). **The Oxford Handbook of Regulation**. Oxford: Oxford University Press, 2010. p.302-348.

BLACK, Julia; MURRAY, Andrew. Regulation AI and machine learning: setting the regulatory agenda. **European Journal of Law and Technology**, [S.l.], v. 10, n. 3, 2019. Disponível em: <https://www.ejlt.org/index.php/ejlt/article/view/722>. Acesso em: 05 mar. 2023.

BOVENS, Mark. Analysing and assessing accountability: a conceptual framework. **European Law Journal**, [S.l.], vol. 13, n. 4, p. 447-468, 2007.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados (LGPD). Brasília, DF, 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 05 mar. 2023.

BRASIL, Senado Federal. **Projeto de Lei nº 2.338, de 2023**. Dispõe sobre o uso da Inteligência Artificial. Brasília, DF, 2023. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>. Acesso em: 05 jun. 2023.

BRASIL. **Relatório final da Comissão de Juristas responsável por subsidiar a elaboração de Substitutivo sobre Inteligência Artificial instituída pelo Ato do Presidente do Senado n. 4, de 2022, aprovado em 1º de dezembro de 2022**. Senado Federal. Coordenação de Comissões Especiais, Temporárias e Parlamentares de Inquérito. Brasília, 2022. Disponível em: <https://www.stj.jus.br/sites/portalp/SiteAssets/documentos/noticias/Relato%CC%81rio%20final%20CJSUBIA.pdf>. Acesso em: 05 mar. 2023.

BURRIS, Scott; DRAHOS, Peter; SHEARING, Clifford. Nodal governance. **Australian Journal of Legal Philosophy**, [S.l.], vol. 30, 2005. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=760928. Acesso em: 11 jun. 2023.

CANADÁ. **Bill C-27: An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts**. Ottawa, Canadá, 2022. Disponível em: <https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading>. Acesso em: 05 mar. 2023.

CANADÁ. **Directive on Automated Decision-Making**. Canadá: [s.n.], 2021. Disponível em: <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592>. Acesso em: 05 mar. 2023.

DUARTE, ALAN. **A Antidiscriminação no contexto da inteligência artificial:** possibilidades de governança mediante a normatização de algoritmos. Fortaleza: Mucuripe, 2021.

GABRIEL, Martha. **Inteligência Artificial:** do zero ao metaverso. Rio de Janeiro: Forense, 2022.

GELLERT, Raphael. The role of the risk-based approach in the General data protection Regulation and in the European Commission's proposed Artificial Intelligence Act. Business as usual?. **Journal of Ethics and Legal Technologies**, [S.l.], v. 3, n. 2, p. 15-33, 2021. Disponível em: <https://doi.org/10.14658/pupj-jelt-2021-2-2>. Acesso em: 19 jun. 2023.

GOMES, Maria Cecília O. Relatório de impacto à proteção de dados: uma breve análise da sua definição e papel na LGPD. **Revista do Advogado**, São Paulo, v. 39, n. 144, p. 174-183, 2019.

KAMINSKI, Margot; MALGIERI, Gianclaudio. Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations. **U of Colorado Law Legal Studies Research Paper**, [S.l.], n. 19-28, 2020. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3456224. Acesso em 24 jun. 2023.

KELLER, Clara Iglesias. **Regulação nacional de serviços na Internet:** exceção, legitimidade e o papel do Estado. 2019. 300 f. Tese (Doutorado em Direito Público) - Faculdade de Direito. Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2019.

LODGE, Martin; STIRTON, Lindsay. Accountability in the Regulatory State. In: BALDWIN, Robert; CAVE, Martin; LODGE, Martin (ed.). **The Oxford Handbook of Regulation**. Oxford: Oxford University Press, 2010. p. 349-370.

MOSES, Lyria Bennett. How to Think about law, regulation and technology: problems with 'technology' as a regulatory target. **Innovation and Technology**, 1-20(2013). (2013) 5(1). UNSW Law Research Paper n. 2014-30. Disponível em: <http://ssrn.com/abstract=2464750>. Acesso em: 11 jun. 2023.

PASQUALE, Frank. **The Black Box Society**. Cambridge: Havard Unity Press, 2015.

QUELLE, Claudia. Privacy, proceduralism and self-regulation in data protection law. **Teoria e Critica della Regolazione Sociale / Theory and Criticism of Social Regulation**, [S.l.], v. 1, n. 14, p. 89-106, 2020. Disponível em: <https://mimesisjournals.com/ojs/index.php/tcrs/article/view/113>. Acesso em: 05 mar. 2023.

REISMAN, Dillon; SCHULTZ, Jason; CRAWFORD, Kate; WHITTAKER, Meredith. **Algorithmic Impact Assessment:** a practical framework for public agency accountability.

AI NOW, 2018.

SELBST, Andrew D. An institutional view of algorithmic impact assessment. *Harvard Journal of Law & Technology*, Massachusetts, v. 35, n. 1, p.117-173, 2021. Disponível em: <https://jolt.law.harvard.edu/assets/articlePDFs/v35/Selbst-An-Institutional-View-of-Algorithmic-Impact-Assessments.pdf>. Acesso em: 05 mar. 2023.

WATKINS, Elizabeth Anne; MOSS, Emanuel; METCALF, Jacob; SINGH, Ranjit; ELISH, Madeleine Clare. Governing algorithmic systems with impact assessments: six observations. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, Nova York, mai 19–21, p.1010-1021, 2021, Virtual Event. Disponível em: <https://dl.acm.org/doi/10.1145/3461702.3462580>. Acesso em: 05 mar. 2023.

ZANATTA, Rafael A. F. Proteção de dados pessoais como regulação do risco: uma nova moldura teórica? In: *I Encontro da Rede de Pesquisa em Governança da Internet*. Rio de Janeiro, 2017. **Anais**. Rio de Janeiro: Rede de Pesquisa em Governança da Internet, 2018. p. 175-193.

ALGORITHMIC IMPACT ASSESSMENT AS A TOOL TO STRENGTHEN ETHICS BY DESIGN IN AI

*A AVALIAÇÃO DE IMPACTO ALGORÍTMICO COMO
FERRAMENTA PARA O FORTALECIMENTO DO ETHICS
BY DESIGN EM IA*

Gabriel Guedes Toscano³⁸⁸

Abstract

Artificial Intelligence - AI is an increasingly present and necessary technology within the current social dynamics. The solutions that make use of this technology permeate all areas of knowledge, and influence micro and macro reality, being of broad importance for both the public and private sectors. Despite the benefits, AI also presents risks to the fundamental rights of citizens, especially due to algorithmic biases and the opacity of its solutions. In this paper, it is intended to discuss which strategies are more effective in mitigating the risks presented by AI solutions, proposing the application of the Ethics by Design approach in combination with the Algorithmic Impact Assessment - AIA. Ethics by Design proposes the addressing of ethical issues from the conception of the AI solution and throughout its use, ensuring respect for the fundamental rights of users and the appropriateness of the solution's operation. With respect to AIA, it aims to identify risks in the AI solution activity, analyze them, and propose mitigating actions, presenting the results in a transparent way to all stakeholders. With the application of the proposed approach, it was intended to present a way to ensure the suitability, safety and reliability of AI solutions, while reinforcing its aspects of transparency and accountability, in order to propose a governance arrangement capable of addressing the challenges presented today.

Keywords: Artificial Intelligence; Algorithmic Impact Assessment; Ethics By Design; Fundamental Rights.

388 BA in Law from the Federal University of Paraíba (UFPB), with academic mobility in International Business at the Hochschule Bremen, Germany; Lawyer; Policy Fellow at LACNIC; Associate Researcher at IP.rec; Civic Innovation Ambassador at OKBR.

Resumo

A Inteligência Artificial – IA é uma tecnologia cada vez mais presente e necessária dentro da dinâmica social atual. As soluções que se utilizam desta tecnologia permeiam todas as áreas do conhecimento, e influenciam a realidade micro e macro, sendo de ampla importância tanto para o setor público quanto privado. Apesar dos benefícios, a IA também apresenta riscos aos direitos fundamentais dos cidadãos, especialmente em razão dos vieses algorítmicos e da opacidade de suas soluções. No presente trabalho, pretendeu-se discutir quais estratégias são mais efetivas na mitigação dos riscos apresentados pelas soluções de IA propondo-se a aplicação da abordagem do Ethics by Design em conjunto com a realização do Algorithmic Impact Assessment – AIA. O Ethics by Design tem por preceito a preocupação com questões éticas desde a concepção da solução de IA e durante toda sua utilização, garantindo o respeito aos direitos fundamentais dos usuários e a adequação do funcionamento da solução. Com respeito ao AIA, este tem por objetivo identificar riscos na atividade da solução de IA, analisá-los, e propor ações mitigadoras, apresentando os resultados de modo transparente a todos os stakeholders. Com a aplicação da abordagem proposta, intentou-se apresentar uma maneira de garantir a adequação, segurança e confiabilidade das soluções de IA, ao mesmo tempo em que reforça seus aspectos de transparência e accountability, com fins de propor um arranjo de governança capaz de endereçar os desafios apresentados na atualidade.

Palavras-chave: Inteligência Artificial; Avaliação de Impacto Algorítmico; Ethics By Design; Direitos Fundamentais.

1. INTRODUCTION

In recent times, society has been confronted with an ever-increasing level of technological development and the digitalisation of countless processes and applications. In such a scenario, it is essential that the ability of new solutions to create a better reality goes hand in hand with the principles and social values that are the cornerstones of our society.

Artificial Intelligence (AI) and algorithmic systems have been a leading technology trend for several years³⁸⁹. It has had a major impact in a wide range of areas, helping to analyze and identify patterns in databases, enhancing decision-making capabilities and being a key tool in the development of digital solutions³⁹⁰.

389 LIKENS, Scott. Eight emerging technologies and six convergence themes you need to know about. **PwC**, Jan. 12, 2022. Available at: <https://www.pwc.com/us/en/tech-effect/emerging-tech/essential-eight-technologies.html>. Accessed on: Mar. 05, 2023.

390 BIGONHA, Carolina. Panorama setorial da Internet: Inteligência Artificial em Perspectiva. **Inteligência Artificial e Ética**, Year X, nº 2. São Paulo: NIC.br, Oct. 2018. p. 1.

However, in addition to the positive aspects of AI and its exponential capacity to change society for the better, it is important to monitor the algorithmic decision-making processes of AI in order to ensure that its operation and development are in line with ethical standards³⁹¹. The most common characteristics of AI that raise concerns about the appropriateness of solutions are 6: ubiquity, opacity, complexity, integrated values, ability to create a reality and perpetuate social conditions, and the possibility of gaming³⁹².

Each of these attributes is associated with some “red flags” of AI solutions. The ubiquity refers to the omnipresence of AI, since it influences every aspect of society; the opacity and complexity refers to the black box of AI and algorithms and the difficulty for the general public to understand the logic behind AI decisions; the integrated values refers to the fact that AI is not impartial and its models always reflect a set of ideologies and beliefs; the ability to create a reality and perpetuate social conditions refers to the risk that AI poses to society if its learning processes are filled with false and biased premises; and the possibility of gaming refers to the risk of cheating the algorithmic models if they are opened to the general public³⁹³.

With the aim of mitigating the risks and impacts caused by the above-mentioned characteristics of AI, major organizations around the world, such as the Organisation for Economic Co-operation and Development (OECD) and the European Commission, have sought to define a set of principles and recommendations that aim to “foster innovation and trust in AI by promoting the responsible stewardship of trustworthy AI while ensuring respect for human rights and democratic values.”³⁹⁴

The chosen ethical principles vary from organization to organization, but the list established by the European Commission is one that at best translates the complexity of nuances that are part of the AI and digital ecosystem, indicating the following principles as necessary for the proper development of AI solutions: respect for human agency; privacy, personal data protection and data governance; fairness; individual, social and environmental well-being; transparency; accountability and oversight³⁹⁵.

The application of such principles is usually done through an approach called Ethics by Design, which is a form of ensuring the incorporation of ethical principles throughout the development and deployment of AI solutions, as will be discussed in more detail

391 BUSUIOC, Madalina. *Accountable Artificial Intelligence: Holding Algorithms to Account*. **Public Administration Review**, Vol. 81, Iss. 5, pp. 825-836. Hoboken: Wiley Periodicals LLC, Aug. 2020, p. 1.

392 ROSA, Fernanda R. *Entendendo os Algoritmos: Propriedades e Dilemas*. In: NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO BR. **TIC Domicílios 2017: Pesquisa Sobre o Uso das Tecnologias de Informação e Comunicação nos Domicílios Brasileiros**. São Paulo: Comitê Gestor da Internet no Brasil, 2018. p. 67-73.

393 *Ibidem*.

394 OECD. **Recommendation of the Council on OECD Legal Instruments Artificial Intelligence**. Paris: OCDE, 2019. p. 3.

395 EUROPEAN COMMISSION. **Ethics By Design and Ethics of Use Approaches for Artificial Intelligence**. Brussels: European Commission, Nov. 2021. p. 3.

in the next chapter. In this sense, the present article aims to propose the application of the accountability mechanism called Algorithmic Impact Assessment - AIA, as a tool capable of strengthening the implementation of Ethics by Design in AI solutions, as a form of combating the biases presented on them and creating more transparent, secure and trustworthy technologies.

Finally, the methodology adopted for the development of this article was mainly bibliographic review, both in national and international sources. In addition, important feedback and ideas for this work were gathered by participating in the event “Mitigating Risks of AI: Responsible Research, Development, and Use of AI in Brazil”, where I exchanged experiences and thoughts with a large number of brilliant researchers and professors specialized in this field.

2. THE WAY THROUGH ETHICS BY DESIGN

It is common for the majority of AI solutions to have some level of human involvement to monitor its decision-making process. There are 3 broad approaches to such human involvement: human-in-the-loop (HITL), human-out-of-the-loop and human-over-the-loop.

In short, the HITL model is about active human involvement, where the human operator is a central part of the decision-making process. In the human-out-of-the-loop model, it is the complete opposite of the previous one, because it has no human supervision at all in the application. Finally, in the human-over-the-loop model, the human operator plays the role of supervisor of the decision-making process, acting only in unexpected or undesirable events³⁹⁶.

Although the three models provide sufficient support for optimizing AI systems with a narrow impact, when it comes to situations that the technologies may cause a broad impact on society as a whole, individual or group intervention in AI does not help much in achieving the best results. In such scenarios, the use of a different model of human oversight, called Society-in-the-Loop (SITL), may prove more appropriate³⁹⁷.

The SITL “is about embedding the judgment of society, as a whole, in the algorithmic governance of societal outcomes”³⁹⁸. It can be seen as a form of attempting to embed

396 IMDA/PDPC. **Model Artificial Intelligence Governance Framework**. Singapore: IMDA/PDPC, 2020. p. 30.

397 RAWAN, Iyad. Society-in-the-Loop: Programming the Algorithmic Social Contract. **Ethics and Information Technology**, v. 20. Berlin: Springer, 2017. p. 4.

398 RAWAN, Iyad. **Society-in-the-Loop: Programming the Algorithmic Social Contract**. Medium, 2017. Available at: <https://medium.com/mit-media-lab/society-in-the-loop-54ffd71cd802>. Accessed on: Mar. 04, 2023.

the general will in an algorithmic social contract, in which people would have to resolve ethical trade-offs between values such as security and privacy, and the concept of fairness, in order to achieve the common good³⁹⁹.

In such a model, ethical values and principles are even more important, as they become an integral part of the balancing system used to ensure the best social performance of AI. As briefly outlined above, the main ethical requirements for AI systems are as follows:

- AI systems must not negatively affect human autonomy, freedom or dignity.
- AI systems must not violate the right to privacy and to personal data protection. They MUST use data which is necessary, non-biased, representative and accurate.
- AI systems must be developed with an inclusive, fair, and non-discriminatory agenda.
- Steps must be taken to ensure that AI systems do not cause individual, social or environmental harm, rely on harmful technologies, influence others to act in ways which cause harm or lend themselves to function creeps.
- AI systems should be as transparent as possible to their stakeholders and to their end-users.
- Human oversight and accountability are required to ensure conformance to these principles and address non-compliance⁴⁰⁰.

As noted, the SITL model and the ethical requirements mentioned are fully complementary, as it is clear that society's values and principles should be taken into account from the design of AI systems and continue throughout the performance of the application to ensure its success. Such a structure is possible through the use of Ethics by Design, which is an approach that "can be used to ensure that the ethical requirements are properly addressed during the development of AI system or technique."⁴⁰¹

Ethics by Design is an essential tool for the proper development and deployment of AI solutions, as it is an optimal way to anticipate risks that may arise in the solution,

399 RAWAN, Iyad. **Society-in-the-Loop: Programming the Algorithmic Social Contract**. Medium, 2017. Available at: <https://medium.com/mit-media-lab/society-in-the-loop-54ffd71cd802>. Accessed on: Mar. 04, 2023.

400 EUROPEAN COMISSION. **Ethics By Design and Ethics of Use Approaches for Artificial Intelligence**. Brussels: European Comission, Nov. 2021. p. 12.

401 *Ibidem*, p. 11.

which could completely undermine its desired effects and even generate negative consequences for users. Ethics by Design should be present at every step of the AI development process, starting with the (I) specification of the objectives for the system, through the (II) specification of construction requirements, the (III) building of high-level design, the (IV) creation of processes for data collection and preparation, the (V) development of a detailed design, and finally the (VI) testing and evaluation of the solution⁴⁰².

However, the Ethics by Design strategy is not self-executing, and therefore a number of tools need to be applied together to achieve the goal of ensuring that the AI solution is developed as it should be. One of the most important tools that is essential for the Ethics by Design approach to work well is the ethical risk and impact assessment.

The ethical impact assessment is defined as:

A process of judging the ethical impacts of research and innovation activities, outcomes and technologies that incorporates both means for a contextual identification and evaluation of these ethical impacts and development of a set of guidelines or recommendations for remedial actions aiming at mitigating ethical risks and enhancing ethical benefits, typically in consultation with stakeholders⁴⁰³.

The main aim of conducting an ethical impact assessment is to have a clearer idea of the ethical risks associated with the AI solution, and then to analyze whether the goals to be achieved with the AI outweigh the risks and consequences that could arise as a side-effect of using the solution. Such an assessment is central to the Ethics by Design model, and typically takes place during the specification phase of the construction requirements for the new solution. It can provide the initial spark needed to guide the organization through changes to its operational structure and internal governance to demonstrate to stakeholders the responsible use of AI.

As it is always difficult to establish a general rule of what is ethical and what is not, especially when taking into account the risk appetite and reality of each organization, the most common and effective approach to help organizations apply the Ethics by Design approach is to publish guidelines and frameworks that are flexible enough to be applied in most circumstances.

In this context, it is worth mentioning the toolkit developed by the Government of Singapore, called A.I. Verify. This toolkit is still only available as a minimum viable product (MVP), but has already been tested by several companies, such as Google, Meta

402 EUROPEAN COMMISSION. **Ethics By Design and Ethics of Use Approaches for Artificial Intelligence**. Brussels: European Commission, Nov. 2021. p. 12.

403 EUROPEAN COMMITTEE FOR STANDARDIZATION. **Ethics assessment for research and innovation - Part 2: Ethical impact assessment framework**. Brussels: CEN-CENELEC, 2017. p. 8.

and Microsoft, who have provided feedback for the development of the application. The A.I. Verify toolkit also comes with a framework that is designed to enable organizations to demonstrate their use of AI in an objective and verifiable way. This is important for companies to evaluate the performance of their AI solutions and is also an interesting tool that promotes transparency and accountability, certifying to stakeholders and consumers that AI systems are fair, explainable and safe⁴⁰⁴.

All the concerns related to the application of Ethics by Design in AI are more than necessary for the development of any new solution. Nevertheless, the algorithms and algorithmic systems behind AI require some more precautions in order to guarantee the good and appropriate functioning of the applications. For this reason, the Algorithmic Impact Assessment is presented as a tool that can be used alongside the others mentioned above, as a complementary and important measure to be taken by organizations when implementing AI technologies.

2.1. Algorithmic Impact Assessment for AI

Every AI has algorithms or algorithmic systems as its engine and data as its fuel⁴⁰⁵. From now on, some issues related to the engine of AI will be addressed, starting with its definition:

The term algorithm describes a series of steps through which particular inputs can be turned into outputs. An algorithmic system is a system that uses one or more algorithms, usually as part of computer software, to produce outputs that can be used for making decisions⁴⁰⁶.

As observed, algorithms are the logic behind the functioning of AI and serve as the tools that allow the solutions to make decisions based on the knowledge collected and analyzed during the machine learning process. As a central element in the development of AI, misconceptions in the structure of algorithms are the main reason for the existence of biases in AI solutions, and the issues related to the lack of transparency and accountability of AI also concern the functioning of the algorithmic system.

Therefore, the creation of algorithmic accountability initiatives is crucial to ensure the trust and adequacy of AI tools and solutions. In this respect, one of the tools that has the strongest claim to achieve such goals is the Algorithmic Impact Assessment - AIA.

404 YU, Eileen. **Singapore touts need for AI transparency in launch of test toolkit**. ZD NET, May 26, 2022. Available at: <https://www.zdnet.com/article/singapore-touts-need-for-ai-transparency-in-launch-of-test-toolkit/>. Accessed on: Mar. 05, 2023.

405 BIGONHA, Carolina. Panorama setorial da Internet: Inteligência Artificial em Perspectiva. **Inteligência Artificial e Ética**, Year X, nº 2. São Paulo: NIC.br, Oct. 2018. p. 2.

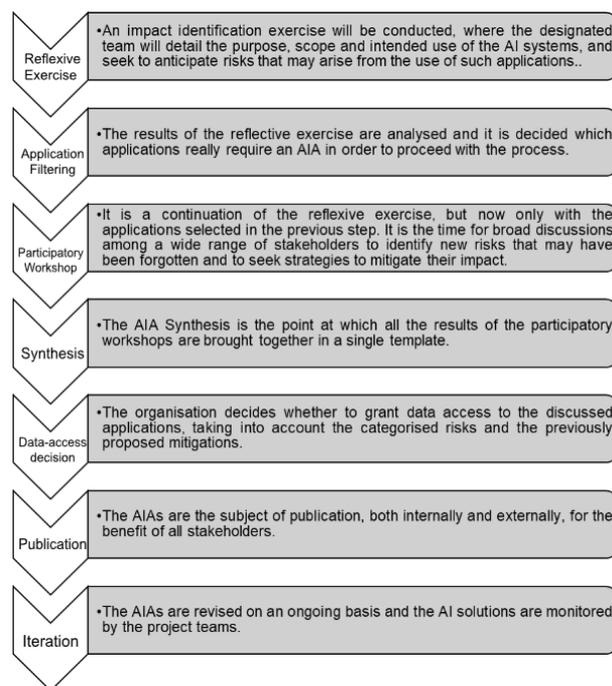
406 ADA LOVELACE INSTITUTE; AI NOW INSTITUTE; OPEN GOVERNMENT PARTNERSHIP. Algorithmic accountability for the public sector. Available at: <https://www.opengovpartnership.org/wp-content/uploads/2021/08/executive-summary-algorithmic-accountability.pdf>. Accessed on: Mar. 07, 2023.

AIAs are a way of assessing the performance and impact of AI systems before their implementation and during their operation⁴⁰⁷.

The main aim of the AIAs is to “build public trust in the use of these systems (algorithmic systems), mitigate their potential to cause harm to people and groups, and maximize their potential for benefit”⁴⁰⁸. The AIA addresses issues of ethical and social relevance (similar to the Ethical Impact Assessment), but also considers environmental and technical factors that could determine the success or failure of the solution, such as stable internet connectivity and the infrastructure within which the AI solution is located⁴⁰⁹.

Although the AIAs are not yet widely applied in practice and are only found in a few organizations, such as the Treasury Board of Canada Secretariat’s Directive on Automated Decision-Making⁴¹⁰, institutions such as the English National Health Service (NHS) already suggest a possible way of applying the AIAs in both the public and private sectors, dividing it into 7 steps:

Figure 1 - AIA process



Source: ADA LOVELACE INSTITUTE. **Algorithmic impact assessment: a case study in healthcare.**

London: Ada Lovelace Institute, 2022, p. 44.

407 ADA LOVELACE INSTITUTE. **Algorithmic impact assessment: a case study in healthcare.** London: Ada Lovelace Institute, 2022. p. 14.

408 *Ibidem.*

409 *Ibidem*, p. 15.

410 GOVERNMENT OF CANADA. **Algorithmic Impact Assessment tool.** (2019) Available at: <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>. Accessed on: Mar. 06, 2023.

The step-by-step approach outlined above is very didactic and has a wide range of applications. In summary, it starts with a kind of brainstorming activity involving all the different stakeholders of the AI solution, in order to anticipate the risks of the technology application and to define its scope and mode of operation. The first 4 steps are related to these risk and impact assessment activities, so that the organization reaches the “Synthesis” phase with a clear decision on which solutions are appropriate to be applied in practice.

The organization can subsequently decide on the amount of data to be made available for the solution to be put into practice. The AIA should then be applied and the results published both internally and externally so that the public has open access to the assessment. Finally, the organization needs to keep track of the development and deployment of the application so that it can continually revise and update the AIA.

As mentioned above, the AIA is an important mechanism to ensure the correct development and deployment of algorithmic systems and AI solutions, providing trust and security for users while promoting greater transparency and accountability for stakeholders. However, it must be clear that AIAs are not a complete solution for AI accountability, and in order to ensure the best development and implementation of AI solutions, the AIA should be complemented by audits, transparency registers and all the other measures related to Ethics by Design that have already been presented⁴¹¹.

In addition, there are two important factors that need to be taken into account when applying AI. First, there is the need to consider the context in which the AI will operate, taking into account the geopolitical and historical scenario in which the solution is embedded. Secondly, there are some errors that have already been identified in other types of impact assessment processes (such as the Data Protection Impact Assessment - DPIA) that should be addressed and corrected in order to achieve better results from the use of AIA.

In terms of the relevance of the context in which AI operates, it is crucial to reinforce Kranzberg’s first law, which states that technology is neither good nor bad; nor is it neutral:

Kranzberg’s First Law reads as follows: Technology is neither good nor bad; nor is it neutral. By that I mean that technology’s interaction with the social ecology is such that technical developments frequently have environmental, social, and human consequences that go far beyond the immediate purposes of the technical devices and practices themselves, and the same technology can have quite

411 ADA LOVELACE INSTITUTE. **Algorithmic impact assessment: a case study in healthcare**. London: Ada Lovelace Institute, 2022. p. 4.

different results when introduced into different contexts or under different circumstances⁴¹².

Such a statement is a clear affirmation of the need to consider the broad scenario related to the application of AI and, consequently, the use of AIA in each and every case. The differences in political, social and environmental aspects are key factors in the results obtained by the technologies, so this concept must be also applied to the use of accountability methods in order to obtain better and more coordinated results, which can be transformed into a clear and continuous feedback generator for the development of the application.

Most of the current computational vision is guided by hierarchies of epistemic and economic domination between the technological poles of the Global North and the Global South. In this reality, the tendency of technological solutions to be biased is even more evident, as their logic is completely distorted by the dynamics of domination that go back to colonialism:

I argue that this data-centric rationality should be understood as an expression of the coloniality of power, manifested as the violent imposition of ways of being, thinking, and feeling that leads to the expulsion of human beings from the social order, denies the existence of alternative worlds and epistemologies, and threatens life on Earth⁴¹³.

Just as AI does not work properly in scenarios and social dynamics to which it does not belong, so it does with AIA. Such an accountability mechanism cannot be fully transplanted from one reality to another, because the understanding of what can cause impacts in the application, the extension of impact effects, and the form of risk assessment vary widely from scenario to scenario.

In parallel to the precautions mentioned above, there is another issue that needs to be outlined in order to ensure better results in the application of AIA. This has to do with the errors already identified in the application of the DPIA, which should not be repeated in the AIA:

- A description of the means for the protection of data subjects' rights is missing or insufficiently elaborated.
- The description of the technical and organisational measures adopted tends to be general and, moreover, it is unclear how the

412 KRANZBERG, Melvin. Technology and History: "Kranzberg's Laws". **Technology and Culture**, v. 27, n. 3, p. 544-560. Maryland: Johns Hopkins University Press, 1986. p. 545.

413 RICAURTE, Paola. "Data Epistemologies, Coloniality of Power, and Resistance". **Television & New Media**, v. 20, n. 4, 2019.

DPIA submitter arrived at them, e.g., the OPDP's methodology is not used, and the data controller's own methodology is not clear either.

- The problematic implementation of 'balancing tests' when the necessity, suitability and proportionality of the personal data processing cannot be verified⁴¹⁴.

All the problems mentioned above have a common root: the fact that there is no standardized model for applying the DPIA, leaving it to the discretion of the data controller to decide how to conduct the balancing tests, how to describe the data processing, and which methodology to use to assess the risks and determine the results of the process.

This circumstance leads to a lack of efficiency in the DPIA application, as it does not produce credible and integrated results that show data subjects and authorities the risks and impacts that the data processing may cause. The same scenario could be replicated in the AIA application, especially if the algorithmic black box context is taken into account, which may lead companies to try to cover some details in the impact assessments.

In this context, a clear methodology and risk categorisation (specific to each country and sector) should be established for the best application of AIA in the use of new AI solutions, in order to guarantee the safety, integrity and transparency of these technologies for the general public.

After all, risk assessment is a fundamental step in any process, and even more so in the technology ecosystem, where every single decision has the potential to affect the lives of countless people, positively or negatively. Therefore, the relevance of AIAs and the benefits they bring, if applied appropriately, are unquestionable in any context involving algorithmic systems and the development and deployment of AI solutions.

3. CONCLUSION

The information society and all the innovations that surround it pose a recurring paradigm that is discussed every time a new technology is created. Such a paradigm refers to the two sides of any innovation, because while it may have good results for society, it also has consequences that can be negative for those involved.

In the present discussion, it is impossible not to recognise the high value and public good that AI and the use of algorithmic systems have created for society. Many new solutions are based on these technologies and they have the extreme power to revolutionize many aspects of society, both on an individual level and in a broader context related to States and governments.

414 KARAS, Ladislav. Beware of errors in personal data protection impact assessments. **KPMG**. Available at: <https://danovky.cz/en/news/detail/1020>. Accessed on: Jul. 05 2023.

Alongside all the positive aspects of such technologies, this article has attempted to analyze how the negative consequences of AI and algorithmic systems, such as a lack of transparency and accountability and the perpetuation of inequalities, can be addressed through a strategy based on the application of Ethics by Design in combination with an Algorithmic Impact Assessment - AIA.

The main objective of the Ethics by Design approach is to guarantee that the AI solution, from its conception to its deployment, takes into account issues related to ethics and respect for fundamental rights, so that its decision-making process is trustworthy and does not reproduce biased opinions that could undermine the positive potential of the platform in society.

In order to ensure the correct implementation of Ethics by Design, the AIA has been presented as a tool that can be used by both the public and private sectors, with the aim of categorizing the risks of the algorithmic system, analyzing them and promoting ways to mitigate their negative consequences, while at the same time presenting all the results of such an assessment to the stakeholders.

In order to satisfactorily implement a strategy that includes these aspects, it is extremely important that multiple sectors and stakeholders are involved in the process of developing, deploying and monitoring AI solutions. In such a diverse environment, the range of issues to be considered when applying the AIA, for example, would be exponentially greater, allowing for a broader identification of risks and a greater variety of mitigation strategies.

In fact, for the best application of AIA, two aspects also need to be taken into account throughout its use in order for it to produce better results: the broad geopolitical context in which AIA is to be applied, and a clear and established method for risk assessment, in order to produce more transparent and credible results. If these two aspects are truly taken into account, the use of the assessments would be broadened and their impact would contribute to the sustainable and trustworthy development of AI solutions.

Finally, the development of policies aimed at ensuring ethical digital solutions is something fundamental to ensuring a safe and fair digital ecosystem. Society is at the perfect moment to promote and implement such strategies, as effective solutions are still in their infancy, which makes it crucial that every stakeholder is involved in discussions on the ethical parameters for the development of new technologies, in order to create a better and more trustworthy future.

4. Referências

ADA LOVELACE INSTITUTE. **Algorithmic impact assessment: a case study in healthcare**. London: Ada Lovelace Institute, 2022.

ADA LOVELACE INSTITUTE; AI NOW INSTITUTE; OPEN GOVERNMENT PARTNERSHIP. **Algorithmic accountability for the public sector**. Available at: <https://www.opengovpartnership.org/wp-content/uploads/2021/08/executive-summary-algorithmic-accountability.pdf>. Accessed on: Mar. 07, 2023.

BIGONHA, Carolina. Panorama setorial da Internet: Inteligência Artificial em Perspectiva. **Inteligência Artificial e Ética**, Year X, nº 2. São Paulo: NIC.br, Oct. 2018.

BUSUIOC, Madalina. Accountable Artificial Intelligence: Holding Algorithms to Account. **Public Administration Review**, Vol. 81, Iss. 5, pp. 825-836. Hoboken: Wiley Periodicals LLC, Aug. 2020.

EUROPEAN COMMISSION. **Ethics By Design and Ethics of Use Approaches for Artificial Intelligence**. Brussels: European Commission, Nov. 2021.

EUROPEAN COMMITTEE FOR STANDARDIZATION. **Ethics assessment for research and innovation - Part 2: Ethical impact assessment framework**. Brussels: CEN-CENELEC, 2017.

GOVERNMENT OF CANADA. **Algorithmic Impact Assessment tool**. (2019) Available at: <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>. Accessed on: Mar. 06, 2023.

IMDA/PDPC. **Model Artificial Intelligence Governance Framework**. Singapore: IMDA/PDPC, 2020.

KARAS, Ladislav. Beware of errors in personal data protection impact assessments. **KPMG**. Available at: <https://danovky.cz/en/news/detail/1020>. Accessed on: Jul. 05 2023.

KRANZBERG, Melvin. Technology and History: “Kranzberg’s Laws”. **Technology and Culture**, v. 27, n. 3, p. 544-560. Maryland: Johns Hopkins University Press, 1986.

LIKENS, Scott. Eight emerging technologies and six convergence themes you need to know about. **Pwc**, Jan. 12, 2022. Available at: <https://www.pwc.com/us/en/tech-effect/emerging-tech/essential-eight-technologies.html>. Accessed on: Mar. 05, 2023.

OECD. **Recommendation of the Council on OECD Legal Instruments Artificial Intelligence**. Paris: OCDE, 2019.

RAWAN, Iyad. Society-in-the-Loop: Programming the Algorithmic Social Contract. **Ethics and Information Technology**, v. 20. Berlin: Springer, 2017.

RAWAN, Iyad. Society-in-the-Loop: Programming the Algorithmic Social Contract. **Medium**, 2017. Available at: <https://medium.com/mit-media-lab/society-in-the-loop-54ffd71cd802>. Accessed on: Mar. 04, 2023.

RICAURTE, Paola. “Data Epistemologies, Coloniality of Power, and Resistance”. **Television & New Media**, v. 20, n. 4, 2019.

ROSA, Fernanda R. Entendendo os Algoritmos: Propriedades e Dilemas. In: NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO BR. **TIC Domicílios 2017: Pesquisa Sobre o Uso das Tecnologias de Informação e Comunicação nos Domicílios Brasileiros**. São Paulo: Comitê Gestor da Internet no Brasil, 2018. p. 67-73.

YU, Eileen. **Singapore touts need for AI transparency in launch of test toolkit**. ZD NET, May 26, 2022. Available at: <https://www.zdnet.com/article/singapore-touts-need-for-ai-transparency-in-launch-of-test-toolkit/>. Accessed on: Mar. 05, 2023.

CÂMERAS DE RECONHECIMENTO FACIAL NA SEGURANÇA PÚBLICA: UMA ANÁLISE SOBRE AS NOVAS FORMAS DE CONTROLE E VIGILÂNCIA⁴¹⁵

FACE RECOGNITION CAMERAS IN PUBLIC SAFETY: AN ANALYSIS ON NEW FORMS OF CONTROL AND SURVEILLANCE

Maria Vitoria Pereira de Jesus⁴¹⁶

Resumo

Neste trabalho, propomo-nos a analisar o caso de implementação de câmeras de reconhecimento facial na cidade do Rio de Janeiro – RJ, que busca introduzir novas formas de controle e vigilância para fins de fortalecimento da segurança pública. Em nossas análises, pudemos evidenciar que a adoção dessa tecnologia por governos brasileiros parece estar na contramão das ações adotadas por governos de cidades como São Francisco e Oakland (EUA) e os principais argumentos utilizado para a sua implementação, o de “fortalecimento da segurança pública” e proporcionar “eficiência” podem ser colocados em questão pelo número de erros e falsos positivos, principalmente sobre pessoas negras, mulheres e transexuais. As prisões injustas e precipitadas pelo uso de câmeras de reconhecimento facial na segurança pública revelam não só a presença de vieses raciais e de gênero, que podem ser introduzidos no processo de aprendizagem da máquina e resultar em racismo algorítmico, como também evidenciam a existência de uma imaginação carcerária.

Palavras-chave: reconhecimento facial; segurança pública; imaginários sociotécnicos; racismo algorítmico; imaginação carcerária.

Abstract

In this work, we propose to analyze the case of implementation of facial recognition

415 Este texto trata-se de um trecho de um trabalho de monografia apresentado como requisito para conclusão de curso e obtenção do título de bacharel em Ciências Sociais.

416 Graduanda em Ciências Sociais pela Universidade Estadual de Montes Claros (Unimontes).

cameras in the city of Rio de Janeiro - RJ, which seeks to introduce new forms of control and surveillance for the purpose of strengthening public safety. In our analyses, we were able to show that the adoption of this technology by Brazilian governments seems to be against the actions taken by governments of cities such as San Francisco and Oakland (USA) and the main arguments used for its implementation, or “strengthening public safety”. and providing “efficiency” can be called into question by the number of errors and false positives, especially on black people, women and transgender people. Unjust and precipitated arrests due to the use of facial recognition cameras in public security not only show the presence of racial and gender biases, which can be introduced in the machine learning process and result in algorithmic racism, but also show the existence of a prison imagination.

Keywords: facial recognition; public security; sociotechnical imaginaries; algorithmic racism; prison imagination.

1. INTRODUÇÃO

Em transportes coletivos, escolas e universidades, no Brasil, observamos uma crescente adoção de Tecnologias de Reconhecimento Facial (TRF) em instituições e espaços públicos. Em São Paulo, por exemplo, a Universidade Estadual de Campinas (Unicamp), recentemente, informou que usará a tecnologia no procedimento de heteroidentificação do Vestibular Unicamp 2023 com o objetivo de evitar fraudes no processo de seleção por cotas raciais⁴¹⁷. No estado de São Paulo, e em diversos outros do país, tais como Bahia, Espírito Santo e Santa Catarina, percebe-se que cada vez mais gestores públicos têm sido convencidos das “vantagens” da utilização do reconhecimento facial em processos de seleção, escolas e empresas privadas. Os motivos mais destacados compreendem a necessidade de segurança, a possibilidade de se evitar fraudes e do uso de uma forma de controle mais eficiente.

Já conhecemos o uso do Reconhecimento Facial na segurança pública e no policiamento preditivo. No entanto, também sabemos do seu potencial para atuar na reprodução de preconceitos e discriminações, que resultam em prisões injustas e perseguições a determinados indivíduos e grupos. O número de prisões injustas e erros de identificação por sistemas de reconhecimento facial são consideráveis, principalmente, quando se trata de mulheres negras e transsexuais⁴¹⁸. Por isso, com base na possibilidade de

417 KER, João. Vestibular da Unicamp terá reconhecimento facial para evitar fraudes em cotas. UOL. São Paulo, 01 de agosto de 2022. Disponível em: https://educacao.uol.com.br/noticias/agencia-estado/2022/08/01/vestibular-da-unicamp-tera-reconhecimento-facial-para-evitar-fraudes-em-cotas.htm?utm_source=twitter&utm_medium=social-media&utm_content=geral&utm_campaign=noticias. Acesso em: 23 de agosto de 2022.

418 FALTAY, Paulo. Parlamentares de todas as regiões do Brasil apresentam projetos de lei pelo banimento do reconhecimento facial em espaços públicos. Lavits. Rio de Janeiro, 21 de junho de 2022. Disponível em: <https://lavits.org/parlamentares-de-todas-as-regioes-do-brasil-apresentam-projetos-de-lei-pelo-banimento-do-reconhecimento-facial-em-espacos-publicos/>. Acesso em: 26 de julho de 2022.

ocorrência dessas situações, são inúmeros os pesquisadores e ativistas que pedem o seu banimento dos espaços públicos, salientando o seu caráter falho e cheio de vieses⁴¹⁹, que poderá contribuir não apenas para o aumento da violência policial sobre a população negra, como também para a ampliação do número de pessoas presas injustamente e da vigilância em massa.

Aparentemente, os usos de Tecnologias de Reconhecimento Facial pertencem a um imaginário sociotécnico, que compreende um futuro desejável e um entendimento compartilhado sobre a vida e o ordenamento social através dos avanços científicos e tecnológicos⁴²⁰. Além disso, a criação de um imaginário também envolve o que a sociedade compreende como “bom” ou “ruim”⁴²¹ em determinado setor da vida social. Nesse sentido, se considerarmos a crescente instalação de câmeras de reconhecimento facial na segurança pública, um “bom” policiamento ou uma “boa” segurança parece ser feita com o uso de tecnologias algorítmicas e, com isso, pouco se questiona sobre os seus riscos para os indivíduos e os objetivos para os quais são utilizadas, que, em alguns casos, podem se revelar controversos.

A implementação de Tecnologias de Reconhecimento Facial (TRFs), na maioria das situações, é feita sem nenhum debate público prévio e o consentimento dos cidadãos que terão seus dados biométricos coletados, tratados e analisados. Esse é o caso, por exemplo, da cidade do Rio de Janeiro – RJ, com instalação de câmeras de reconhecimento facial em vias de acesso público, para fins de fortalecimento da segurança pública, que se constitui no principal caso analisado neste trabalho. Nos documentos, relatórios e notícias sobre o uso de câmeras de reconhecimento facial no Rio de Janeiro, evidenciamos não só a falta de debates, transparência e cuidado em assegurar direitos previstos e garantidos pela Lei Geral de Proteção de Dados Pessoais – LGPD, como também a ocorrência de inúmeras prisões injustas em virtude do erro emitido pelas máquinas. Além disso, nesse contexto, a partir da consulta ao relatório *“Um Rio de câmeras com olhos seletivos: o uso do reconhecimento facial pela polícia fluminense”* de Pablo Nunes⁴²² e às notícias os falsos positivos emitidos pelas câmeras, evidencia-se que o maior número de prisões errôneas incide sobre jovens e homens negros, tornando-se necessário que investiguemos as possíveis origens desses vieses e reflitamos sobre o seu uso para fins de segurança, uma vez que pode provocar injustiças e assegurar a manutenção de um determinado perfil de “procurados”, “suspeitos” e “encarcerados”.

419 FALTAY, Paulo. Parlamentares de todas as regiões do Brasil apresentam projetos de lei pelo banimento do reconhecimento facial em espaços públicos. **Lavits**. Rio de Janeiro, 21 de junho de 2022. Disponível em: <https://lavits.org/parlamentares-de-todas-as-regioes-do-brasil-apresentam-projetos-de-lei-pelo-banimento-do-reconhecimento-facial-em-espacos-publicos/>. Acesso em: 26 de julho de 2022.

420 JASANOFF, Sheila. Future imperfect: Science, Technology and the Imaginations of modernity. In: JASANOFF, Sheila; KIM, Sang-Hyun. **Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power**. Chicago: University of Chicago Press, 2015, p. 1-33.

421 *Ibidem*.

422 NUNES, Pablo. **Um Rio de câmeras com olhos seletivos: uso do reconhecimento facial pela polícia fluminense**. Rio de Janeiro: CESeC, 2022.

Deste modo, propomo-nos a analisar o caso de implementação de câmeras de reconhecimento facial na cidade do Rio de Janeiro – RJ, que busca introduzir novas formas de controle e vigilância para fins de fortalecimento da segurança pública. No contexto atual, são vários os chefes, governadores e gestores públicos que propõem e efetivam a sua instalação sob o argumento de promover o fortalecimento da segurança e tornar os diversos processos de análise e controle mais “objetivos” e eficientes. Essas decisões, porém, na maioria dos casos, são tomadas dando pouca atenção aos inúmeros estudos e documentos que salientam a possibilidade dos erros e vieses emitidos pelas máquinas promoverem situações de discriminação e preconceito, principalmente, quando levamos em consideração o fato de que as maiores taxas de erro incidem sobre indivíduos de pele negra. Nessas situações, compreendemos que os vieses raciais e de gênero emitidos pelos algoritmos não só desfavorecem e prejudicam grupos “minorizados”⁴²³ como também contribuem para a manutenção de antigas lógicas de segregação e preconceito com base em cor/raça.

Para tanto, realizamos leituras no âmbito dos Estudos Sociais da Ciência e Tecnologia, que nos permitiram refletir sobre os interesses políticos, sociais e econômicos presentes no uso e na implementação de tecnologias e a possibilidade de que elas incorporem diferentes níveis de poder e consciência⁴²⁴. Além disso, autores como Andrew Feenberg⁴²⁵, ao considerar os dilemas em relação ao uso e aplicação de tecnologias e utilizar a noção de troca compensatória (“*trade-offs*”), nos fizeram pensar nos impasses presentes nas decisões sobre o uso das Tecnologias de Reconhecimento Facial, que confrontam agilidade e eficiência e privacidade versus segurança. No entanto, tratando-se desta última, acreditamos que ela possui um caráter contraditório, pois entendemos que ter privacidade é termos a segurança de que não estamos sendo constantemente vigiados. Portanto, se as câmeras de vigilância pretendem promover a segurança, perguntamos: em que sentido a primeira reforça a segunda? Além disso, ao considerarmos que o uso de TRFs⁴²⁶ podem resultar em insegurança para os negros que, na maioria dos casos, são os que mais sofrem com os constrangimentos decorrentes dos erros emitidos pela máquina, perguntamos ainda: vigiar para promover a segurança de quê ou de quem? Compreendemos que esses questionamentos são de suma importância para um uso transparente, seguro e responsável das TRFs em espaços públicos e instituições privadas.

Para uma melhor compreensão do funcionamento e dos resultados fornecidos por TRF,

423 SILVA, Tarcízio. **Racismo algorítmico: Inteligência Artificial e discriminação nas redes digitais**. Sérgio Amadeu da Silveira (Org.). São Paulo: Edições SESC, 2022.

424 WINNER, Langdon. **Artefatos têm política?** Traduzido por Fernando Manso. Chicago: The University of Chicago Press, 1986.

425 FEENBERG, Andrew. Racionalização democrática, poder e tecnologia. In: NEEDER, Ricardo T. (org.). **Ciclo de conferências Andrew Feenberg**, 2010.

426 Tecnologias de Reconhecimento Facial.

acessamos estudos, como o de Cathy O’neil⁴²⁷, Fernanda Bruno⁴²⁸, Matteo Pasquinelli⁴²⁹ e Tarcízio Silva⁴³⁰, que nos fornecem um breve entendimento sobre o funcionamento de uma Inteligência Artificial, dos seus algoritmos e das etapas do processo de programação de uma máquina. Para a compreensão dos vieses raciais e de gênero emitidos pelas máquinas, apoiamo-nos nos estudos de Matteo Pasquinelli e Vladar Joler⁴³¹ e Joy Buolamwini e Timnit Gebru⁴³². Essas últimas tratam especificamente dos vieses emitidos pelos sistemas de reconhecimento facial. Por fim, cientes da relação estreita entre raça e tecnologia, consultamos os estudos de Silva⁴³³, Ruha Benjamin⁴³⁴ e Simone Browne⁴³⁵, que dedicam a pensar como antigas lógicas de preconceito, discriminação e segregação racial podem ser “atualizadas” e intensificadas através de tecnologias algorítmicas, como o reconhecimento facial.

Para a análise do caso principal deste trabalho, valemo-nos da análise documental de relatórios, como o “*Um Rio de câmeras com olhos seletivos: o uso do reconhecimento facial pela polícia fluminense*” de Pablo Nunes⁴³⁶; do documento do 19º Batalhão de Polícia Militar, que apresenta a ordem de prioridade da instalação das câmeras na cidade; do histórico de ocorrência do Batalhão Especializado em Policiamento em Estádios (BEPE), que consta o número de abordagens, inclusive as resultantes de falsos positivos, e prisões feitas a partir do uso de reconhecimento facial; e do Estudo Técnico Preliminar para a implementação de sistemas de videomonitoramento no Jacarezinho. Ademais, compreendemos que, para este trabalho, o acesso às notícias e reportagens jornalísticas

427 O’NEIL, Cathy. **Weapons of math destruction: how big data increases inequality and threatens democracy**. New York: Crown Publishers, 2016.

428 BRUNO, Fernanda. **Máquinas de ver, modos de ser: vigilância, tecnologia e subjetividade**. Porto Alegre: Sulina, 2013; BRUNO, Fernanda. Monitoramento, classificação e controle nos dispositivos de vigilância digital. Revista FAMECOS. Porto Alegre, n. 36, p. 10-16, agosto, 2008.

429 PASQUINELLI, Matteo. Machines that Morph Logic: Neural Networks and the Distorted Automation of Intelligence as Statistical Inference. **Logic Gate: the Politics of the Artifactual Mind (Online)**. 2017. Disponível em: <https://www.glass-bead.org/article/machines-that-morph-logic/?lang=enview>. Acesso em: 20 de agosto de 2022.

430 SILVA, Tarcízio. **Racismo algorítmico: Inteligência Artificial e discriminação nas redes digitais**. Sérgio Amadeu da Silveira (Org.). São Paulo: Edições SESC, 2022.

431 PASQUINELLI, Matteo; JOLER, Vladan. O manifesto Nooscópio: Inteligência Artificial como Instrumento de Extrativismo do Conhecimento. **Lavits**, Rio de Janeiro, 30 de julho de 2020. Disponível em: <https://lavits.org/o-manifesto-nooscopio-inteligencia--artificial-como-instrumento-de-extrativismo-do-conhecimento/?lang=pt>. Acesso em: 25 de maio de 2023.

432 BUOLAMWINI, Joy; GEBRU, Timnit. Gender shades: intersectional accuracy disparities in commercial gender classification. **Proceedings of Machine Learning Research**, v. 81, n. 1, p. 1-15, 2018.

433 SILVA, Tarcízio. **Racismo algorítmico: Inteligência Artificial e discriminação nas redes digitais**. Sérgio Amadeu da Silveira (Org.). São Paulo: Edições SESC, 2022.

434 BENJAMIN, Ruha. Retomando nosso fôlego: estudos de ciência, tecnologia, teoria racial crítica e imaginação carcerária. In: SILVA, Tarcízio. **Comunidades, algoritmos e ativismos digitais: olhares afrodiaspóricos**. São Paulo: Literarua, 2020, p. 13-24.

435 BROWNE, Simone. **Dark matters: on the surveillance of blackness**. Durham, NC: Duke University Press, 2015.

436 NUNES, Pablo. **Um Rio de câmeras com olhos seletivos: uso do reconhecimento facial pela polícia fluminense**. Rio de Janeiro: CESeC, 2022.

sobre os falsos positivos e o perfil da maioria dos presos pelo uso de reconhecimento facial foram de suma importância para corroborar os argumentos que salientam os riscos dessa tecnologia para a segurança e segregação de determinados subgrupos da sociedade.

2. IMAGINÁRIOS SOCIOTÉCNICOS E O USO DE INTELIGÊNCIA ARTIFICIAL PARA A SOLUÇÃO DE PROBLEMAS RELACIONADOS À AGILIDADE, EFICIÊNCIA E SEGURANÇA

Não poderíamos iniciar uma discussão sobre os usos de Tecnologias de Reconhecimento Facial (TRF) sem fornecer um breve entendimento sobre Inteligência Artificial (IA) e os objetivos dos seus usos desde a sua criação. Para Oliveira⁴³⁷, ainda não há um consenso sobre o conceito de Inteligência Artificial nas literaturas especializadas sobre o assunto. No entanto, segundo ele, em linhas gerais, podemos definir a IA como uma tentativa de reproduzir diversos aspectos da cognição humana a partir do uso de mecanismos computacionais.

O surgimento da Inteligência Artificial encontra-se diretamente relacionado ao da Cibernética, que, ao conceber o mundo como um imenso banco de dados, “considera tudo o que existe ou existiu como matéria prima a ser processada por uma tecnologia que lhe agrega valor”⁴³⁸ e compreende a semelhança das funções de controle e processamento de informações de máquinas e seres vivos e vislumbram a possibilidade de que elas sejam equivalentes e passíveis de serem reproduzidas através de leis matemáticas⁴³⁹. Tais entendimentos e possibilidade permitiram não só que o nosso corpo fosse tratado como um dado a ser processado e tratado como evidência⁴⁴⁰, como também que determinadas funções desempenhadas por humanos fossem realizadas por máquinas e decisões fossem tomadas com base em cálculos matemáticos.

437 OLIVEIRA, Samuel Rodrigues de. **SORRIA, VOCÊ ESTÁ SENDO FILMADO:** tecnologias de reconhecimento facial, privacidade e proteção de dados. Juiz de Fora, 2020. (Trabalho de Dissertação do Programa de Pós-graduação em Direito – UFJF).

438 SANTOS, Laymert G. A informação após a virada cibernética. In. SANTOS, Laymert Garcia dos; KHEL, Maria Rita; KUCINSKI, Bernardo; PINHEIRO, Walter. **Revolução tecnológica, internet e socialismo.** São Paulo, SP: Editora Fundação Perseu Abramo, 2003.

439 KIM, Joon Ho. Cibernética, ciborgues e ciberespaço: notas sobre as origens da cibernética e sua reinvenção cultural. **Horiz. antropol.**, Porto Alegre, v. 10, n. 21, p. 199-219, junho, 2004.

440 BROWNE, Simone. **Dark matters:** on the surveillance of blackness. Durham, NC: Duke University Press, 2015.

Em resumo, a ideia de uma máquina “inteligente” está diretamente relacionada à sua capacidade de reproduzir ações e comportamentos que as permitam desempenhar funções e a tomar decisões “como” seres humanos. Num contexto em que o uso de tecnologias de informação está associado às noções de modernização, neutralidade, eficiência e agilidade, a IA surge com o objetivo de promover “análises neutras” e proporcionar agilidade e eficiência na realização de processos em diferentes âmbitos da sociedade.

Sendo assim, se consideramos esses objetivos e as expectativas que, geralmente, são atribuídas ao uso dessas tecnologias, não deveríamos estranhar o fato de que cada vez mais inteligências artificiais estejam sendo implementadas em instituições públicas e setores específicos da nossa sociedade, tais como o de serviços, segurança, educação e do judiciário.

3. CÂMERAS DE RECONHECIMENTO FACIAL NA SEGURANÇA PÚBLICA: O CASO DE IMPLEMENTAÇÃO NO RIO DE JANEIRO

Poderíamos ter selecionado outros casos de implementação de câmeras de reconhecimento facial para fins de fortalecimento da segurança pública, como o da Bahia, Santa Catarina e São Paulo. No entanto, neste estudo, optamos por analisar o caso do Rio de Janeiro em função da disponibilidade de documentos, principalmente, após a publicação do estudo “*Um Rio de câmeras com olhos seletivos: o uso do reconhecimento facial pela polícia fluminense*” de Pablo Nunes. O estudo elaborado por Nunes⁴⁴¹ nos mostra não só a falta de transparência e debates públicos acerca da sua implementação, como também revela a contradição presente no fato de investimentos e dinheiro público serem feitos e utilizado para aquisição dessas tecnologias, quando poderiam, por exemplo, ser destinados para a melhoria de serviços básicos de educação e assistência social⁴⁴². Além disso, as análises feitas por Pablo Nunes⁴⁴³ parecem corroborar a existência de um imaginário sociotécnico nas iniciativas de adoção e implementação de TRFs para fins de segurança pública e evidenciar a presença de vieses raciais e de gênero, que podem promover discriminações e provocar constrangimentos.

441 BROWNE, Simone. **Dark matters: on the surveillance of blackness**. Durham, NC: Duke University Press, 2015.

442 De acordo com Pablo Nunes (2022), com o dinheiro gasto na aquisição de câmeras para a favela do Jacarezinho, 103 alunos do Ensino Médio poderiam ter sido custeados durante um ano, assim como também 102 famílias poderiam ter sido contempladas pelo Auxílio Brasil no valor de R\$ 400,00 durante o mesmo período.

443 NUNES, Pablo. **Um Rio de câmeras com olhos seletivos: uso do reconhecimento facial pela polícia fluminense**. Rio de Janeiro: CEsEC, 2022.

A implementação do reconhecimento facial na segurança pública parece permear o cenário político brasileiro desde as eleições de 2018, estando presente, por exemplo, nas promessas de campanha do ex-governador do Rio de Janeiro, Wilson Witzel⁴⁴⁴, para fortalecer a segurança e prevenir a criminalidade. No Rio de Janeiro, o projeto de implementação de câmeras de reconhecimento facial ocorreu logo no primeiro ano de mandato do governador eleito, que optou por efetuar a sua instalação no período de carnaval e em locais estratégicos, como as avenidas próximas ao bairro Copacabana, que dão acesso à praia, e à estação de metrô Siqueira Campos⁴⁴⁵. No segundo semestre do mesmo ano, o projeto foi expandido, com a instalação de câmeras no entorno do estádio Maracanã e do Aeroporto Santos Drumont⁴⁴⁶. Nesse sentido, no ano de 2019, enquanto cidades famosas, como San Francisco, Oakland e Berkeley de países como os Estados Unidos optavam pelo banimento do reconhecimento facial por policiais e agentes públicos^{447,448}, no Brasil, a cidade do Rio de Janeiro, não media esforços para acelerar a sua implementação em vias e espaços de acesso público.

Nos Estados Unidos, os motivos que levaram as cidades de San Francisco, Oakland e Berkeley a banirem o uso do reconhecimento facial⁴⁴⁹ por policiais e agentes públicos foram desde a possibilidade de aumento da injustiça racial, da vigilância massiva pelo governo, ao uso indevido das imagens coletadas e armazenadas através das câmeras de reconhecimento facial. No caso de San Francisco, evidencia-se que todos os motivos ressaltados no documento, que decide pela proibição do reconhecimento facial na cidade, vão de encontro aos salientados por inúmeros pesquisadores e ativistas brasileiros que pedem e lutam pelo banimento dessa tecnologia do país.

No Brasil, é possível que uma das preocupações maiores esteja não só na possibilidade de exacerbação da injustiça racial por uma tecnologia que insiste em sugerir a abordagem

444 HERINGER, Carolina. Uma das principais promessas de campanha de Witzel, câmeras de reconhecimento facial não funcionam mais desde o fim de 2019. **Extra**. Rio de Janeiro, 20 de julho de 2020. Disponível em: <https://extra.globo.com/casos-de-policia/uma-das-principais-promessas-de-campanha-de-witzel-cameras-de-reconhecimento-facial-nao-funcionam-mais-desde-fim-de-2019-24541055.html>. Acesso em 25 de março de 2023.

445 SECRETARIA DE ESTADO DE POLÍCIA MILITAR DO ESTADO DO RIO DE JANEIRO. Câmeras de reconhecimento facial (ordem de prioridade). Rio de Janeiro – RJ, 2019. Disponível: https://docs.google.com/document/d/1dg9OxYhP9sLzlkq_UOmsdsBc-71HCHdW/edit. Acesso em: 01 de março de 2023.

446 HERINGER, Carolina. Op. cit.; NUNES, Pablo. **Um Rio de câmeras com olhos seletivos**: uso do reconhecimento facial pela polícia fluminense. Rio de Janeiro: CESeC, 2022.

447 GOMES, Helton Simões. Por que uma das maiores cidades dos EUA banuiu o reconhecimento facial?. **UOL**. São Paulo, 16 de maio de 2019. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2019/05/16/por-que-uma-das-maiores-cidades-dos-eua-baniu-o-reconhecimento-facial.htm>. Acesso em: 01 de abril de 2023;

448 ABRAMOVAY, Ricardo. Movimento por banir uso de reconhecimento facial cresce no mundo. **Folha de São Paulo**. São Paulo, 14 de dezembro de 2019. Disponível em: <https://www1.folha.uol.com.br/ilustrissima/2019/12/movimento-por-banir-uso-de-reconhecimento-facial-cresce-no-mundo.shtml>. Acesso em: 01 de abril de 2023.

449 GOMES, Helton Simões. Por que uma das maiores cidades dos EUA banuiu o reconhecimento facial?. **UOL**. São Paulo, 16 de maio de 2019. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2019/05/16/por-que-uma-das-maiores-cidades-dos-eua-baniu-o-reconhecimento-facial.htm>. Acesso em: 01 de abril de 2023.

e a prisão de indivíduos de pele negra⁴⁵⁰, como também no risco de que mulheres negras e pessoas transexuais sofram com a restrição de acesso a direitos fundamentais como o de acesso à cidade e aos serviços e locais públicos, e de manifestação da sua personalidade. De acordo com uma pesquisa realizada pela Universidade do Colorado (EUA), que analisou a precisão da tecnologia de reconhecimento facial sobre mulheres, homens não-binários e transexuais, identificaram uma precisão de 97,6% da tecnologia em mulheres e homens cisgênero (que nasceram do sexo feminino e masculino e se identificam como homens e mulheres, respectivamente), em detrimento de 38% nos casos de homens trans⁴⁵¹. Dessa forma, tais dados corroboram, por exemplo, o caráter falho da tecnologia ao analisar rostos de pessoas transexuais.

3.1. Segurança para quê ou para quem?

No Rio de Janeiro, um relatório⁴⁵² do Batalhão Especializado em Policiamento em Estádios (BEPE) sugere que, das 11 detenções ocorridas no entorno do estádio Maracanã, apenas quatro se tratavam de pessoas que tinham, de fato, mandados de prisão em seu nome⁴⁵³. Ao que tudo indica, as outras sete pessoas só foram abordadas em virtude de falsos positivos emitidos pela máquina⁴⁵⁴. Nesse sentido, compreendemos que, além dos constrangimentos, esses indivíduos vítimas dos falsos positivos poderiam ser privados de acesso, por exemplo, no caso do Maracanã, às dependências do estádio. Sendo assim, se supomos ser esse um fato recorrente, o de falsos positivos e abordagens de policiais em virtude dos *matches*⁴⁵⁵ feitos pelas câmeras, é possível que, ao invés de promover mais segurança, o uso dessa tecnologia gere, ao contrário, insegurança aos indivíduos que temem em sofrer tais abordagens e ser impedidos de assistir ao jogo no estádio.

A instalação de câmeras de reconhecimento facial em Copacabana também nos fornece um bom exemplo para pensarmos sobre a “segurança de quê ou de quem”, principalmente se fizermos uma intersecção com a raça e a possibilidade de falsos positivos estarem relacionados à presença de vieses sociais. Em 2018, antes mesmo de uma implementação “massiva” do reconhecimento facial no Rio de Janeiro, uma

450 Segundo Nunes (2022), o primeiro levantamento feito pelo CESeC, de 2018 a 2022, constatou que 90% das pessoas presas através do uso de reconhecimento facial, no Brasil, eram negras.

451 MARSHALL, Lisa. Facial recognition software has a gender problem. **Universidade do Colorado. Colorado – EUA**, 08 de outubro de 2019. Disponível em: <https://www.colorado.edu/today/2019/10/08/facial-recognition-software-has-gender-problem>. Acesso em: 05 de maio de 2023.

452 Relatório disponibilizado na pasta de documentos utilizados para o estudo realizado por Pablo Nunes (2022). Disponível no site do “O Panóptico”: <https://opanoptico.com.br/Caso/um-rio-de-cameras-com-olhos-seletivos-uso-do-reconhecimento-facial-pela-policia-fluminense/>. Acesso em: 30 de abril de 2023.

453 NUNES, Pablo. **Um Rio de câmeras com olhos seletivos**: uso do reconhecimento facial pela polícia fluminense. Rio de Janeiro: CESeC, 2022.

454 *Ibidem*.

455 Quando a máquina identifica a correspondência do rosto filmado com alguma das imagens presentes no banco de dados.

mulher foi presa injustamente ao ser identificada com outra procurada pela polícia⁴⁵⁶. Se juntamos esse acontecimento ao fato de as mulheres negras serem mais suscetíveis aos erros provenientes das análises de reconhecimento facial⁴⁵⁷, é possível que transitar pela orla de Copacabana não seja visto ou tido como algo seguro para mulheres negras. O mesmo ocorre para o caso de homens negros, que são maioria entre os presos por reconhecimento facial⁴⁵⁸. Sendo assim, tais evidências corroboram não só a nossa compreensão de que essa tecnologia possa provocar insegurança aos indivíduos de um grupo específico, como também revela a possibilidade que ela possa colaborar para a segregação e perfilização de alguns espaços de encontro, lazer e interação na cidade.

Portanto, entendemos que, neste caso, o objetivo de fortalecer a segurança na cidade e no bairro Copacabana pode se sobrepor ao de fornecer segurança às pessoas que transitam no entorno da localidade ou pretendem acessar à praia. Além disso, compreendemos que as câmeras de reconhecimento facial, na orla de Copacabana, podem não só impor limites, mas também controlar o trânsito de pessoas negras pela cidade⁴⁵⁹.

A “segurança de quê ou de quem” também pode ser pensada em relação às pessoas transexuais e não-binárias. Em uma pesquisa realizada pelo *Coding Rights*, com o objetivo de mapear o uso das tecnologias de reconhecimento facial pelo governo federal e “avaliar suas ressonâncias para as políticas de identidade de gênero e suas múltiplas expressões”⁴⁶⁰, as pesquisadoras Mariah Rafaella Silva e Joana Varon⁴⁶¹ evidenciaram que o uso do reconhecimento facial pelo setor público já é uma preocupação para a população trans, que acredita que essa tecnologia podem deixá-los mais vulneráveis a situações de constrangimento e não oferecer tanta segurança social^{462 463}, como prometem os gestores que adotam essa tecnologia para fins de fortalecimento da segurança. As respostas das pessoas que responderam ao questionário aplicado pelas pesquisadoras para conhecer a percepção desse grupo específico sobre o uso de tecnologias de reconhecimento facial não foram sem motivo, pois uma pesquisa realizada pela Universidade do Colorado (EUA) constatou que as tecnologias de reconhecimento facial são mesmo mais suscetíveis a

456 O PANÓPTICO. 2020. Por que o uso de reconhecimento facial na segurança é controverso? Disponível em: <https://opanoptico.com.br/por-que-o-uso-de-reconhecimento-facial-na-seguranca-e-controverso/>. Acesso em: 30 de novembro de 2022.

457 GOMES, Helton Simões. Por que uma das maiores cidades dos EUA banuiu o reconhecimento facial?. **UOL**. São Paulo, 16 de maio de 2019. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2019/05/16/por-que-uma-das-maiores-cidades-dos-eua-baniu-o-reconhecimento-facial.htm>. Acesso em: 01 de abril de 2023.

458 NUNES, Pablo. **Um Rio de câmeras com olhos seletivos**: uso do reconhecimento facial pela polícia fluminense. Rio de Janeiro: CESeC, 2022.

459 *Ibidem*.

460 SILVA, Mariah Rafaella; VARON, Joana. **Reconhecimento Facial no setor público e identidades trans**: tecnopolíticas de controle e ameaça à diversidade de gênero em suas interseccionalidades de raça, classe e território. Brasil: Coding Rights, 2021.

461 *Idem*.

462 Constatação feita a partir dos gráficos elaborados pelas autoras.

463 SILVA, Mariah Rafaella; VARON, Joana. tecnopolíticas de controle e ameaça à diversidade de gênero em suas interseccionalidades de raça, classe e território. Brasil: Coding Rights, 2021.

erros quando analisam rostos de pessoas trans do que de cisgêneros⁴⁶⁴. Nesse sentido, embora a pesquisa não tenha sido feita no Brasil, nem a pesquisa de Silva e Varon⁴⁶⁵ especifique o estado de residência das pessoas que responderam ao questionário, compreendemos que os dois estudos contribuem para pensarmos de que segurança se trata se as TRFs apresentam riscos de constrangimento para as pessoas trans e não é vista por essa população como algo capaz fornecer mais segurança social.

4. RAÇA E TECNOLOGIA: POSSÍVEIS “EXPLICAÇÕES” PARA A PRESENÇA DE VIESES RACIAIS E DE GÊNERO EM CÂMERAS DE RECONHECIMENTO FACIAL

Compreendemos que, numa análise mais pormenorizada do uso de câmeras de reconhecimento facial pela polícia militar do Rio de Janeiro, a relação entre raça e tecnologia se mostra evidente, assim como a possibilidade do racismo e outras formas de discriminação serem reproduzidos e exacerbados por essas e através dessas tecnologias. Diante disso, entendemos que uma compreensão de como os preconceitos de raça e gênero podem ser entrelaçados ou estar imbricados nessas tecnologias é de fundamental importância, para que possamos compreender de que modo as câmeras de reconhecimento facial podem contribuir para formas de exclusão, controle, discriminação e segregação racial.

Tarcízio Silva⁴⁶⁶, ao analisar a relação do racismo com as tecnologias de informação e comunicação a partir de estudos e notícias sobre discriminação algorítmica, propõe para essas análises o conceito de “*racismo algorítmico*”, que, segundo ele, trata-se do “modo pelo qual a disposição de tecnologias e imaginários sociotécnicos em um mundo moldado pela supremacia branca realiza a ordenação algorítmica racializada de classificação social, recursos e violência em detrimento de grupos minorizados”⁴⁶⁷.

464 ALVES, Sarah. Além do racismo, reconhecimento facial erra mais em pessoas trans. **UOL**. São Paulo, 14 de fevereiro de 2021. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2021/02/14/nao-e-so-racismo-reconhecimento-facial-tambem-erra-mais-em-pessoas-trans.htm>. Acesso em: 12 de abril de 2023.

465 SILVA, Mariah Rafaella; VARON, Joana. **Reconhecimento Facial no setor público e identidades trans**: tecnopolíticas de controle e ameaça à diversidade de gênero em suas interseccionalidades de raça, classe e território. Brasil: Coding Rights, 2021.

466 SILVA, Tarcízio. **Racismo algorítmico**: Inteligência Artificial e discriminação nas redes digitais. Sérgio Amadeu da Silveira (Org.). São Paulo: Edições SESC, 2022.

467 *Ibidem*.

Isto é, compreendemos o racismo algorítmico como a disposição de tecnologias e imaginários sociotécnicos em fornecer um ordenamento algorítmico de classificações, recursos e violência com base em raça, que desfavorecem e prejudicam, principalmente, os grupos tidos como minorias. Silva⁴⁶⁸ também o compreende essa como sendo uma forma de continuação do racismo estrutural, que delinea as relações de poder de forma ainda mais opacas em nossas sociedades. Assim, os algoritmos, ao emitir resultados enviesados, podem reproduzir pontos de vista, opiniões e conjecturas racistas que, estando presentes nas estruturas mais antigas de nossas sociedades, também podem se refletir no banco de dados.

Silva⁴⁶⁹ salienta que, tratando-se do aprendizado de máquina, o conhecimento do banco de dados a ser processado pelos algoritmos não só é de fundamental importância, como também permite compreendermos um pouco do funcionamento do sistema e seus possíveis resultados. No caso do reconhecimento facial na segurança pública, se o banco de imagens conta com a maioria de rostos de pessoas negras, é provável que as câmeras emitam sinais de alerta ao registrarem rostos de pessoas com a mesma cor e características fenotípicas que os negros.

No Brasil, as polícias de estados que já utilizam essa tecnologia, salientam que fazem uso do banco de imagens da polícia civil⁴⁷⁰, que contém os registros gerais de milhões de cidadãos do estado. No entanto, considerando a falta de transparência das secretarias estaduais em fornecer dados sobre como essa tecnologia funciona^{471 472}, não sabemos se são apenas as imagens dos registros gerais que são processadas ou se também são utilizadas outras, de outros bancos de dados, como o do poder judiciário, que abriga as imagens de presos e condenados. No país, cerca de 64% das pessoas encarceradas são negras⁴⁷³. Portanto, se o reconhecimento facial usado nas câmeras de vigilância para fins de segurança pública for treinado também com esses dados, o número de pessoas negras presas injustamente pelo uso dessa tecnologia poderia ser parcialmente explicado.

4.1. Os vieses raciais e de gênero e a “imaginação carcerária”

Os vieses raciais e de gênero emitidos pelos algoritmos podem estar contidos no banco

468 *Ibidem.*

469 *Ibidem.*

470 O PANÓPTICO. 2020. Por que o uso de reconhecimento facial na segurança é controverso? Disponível em: <https://opanoptico.com.br/por-que-o-uso-de-reconhecimento-facial-na-seguranca-e-controverso/>. Acesso em: 30 de novembro de 2022.

471 NUNES, Pablo. **Um Rio de câmeras com olhos seletivos**: uso do reconhecimento facial pela polícia fluminense. Rio de Janeiro: CESeC, 2022.

472 SILVA, Tarcízio. **Racismo algorítmico**: Inteligência Artificial e discriminação nas redes digitais. Sérgio Amadeu da Silveira (Org.). São Paulo: Edições SESC, 2022.

473 *Ibidem.*

de dados com o qual eles são treinados ou serem incorporados no processo de rotulagem dos dados. Pasquinelli e Joler⁴⁷⁴ consideram que o viés do conjunto de dados é introduzido logo na preparação dos dados a serem utilizados para o treinamento. Para eles, nesse processo, a rotulagem dos dados é a parte mais delicada, pois é feita por operadores humanos que, ao usar referências, termos ou classificações antigas e binárias, podem contribuir para uma visão enviesada do mundo, que pode restringir a diversidade social, aprofundar relações de poder e assegurar posições sociais. Além disso, assim como Silva⁴⁷⁵, os autores também compreendem que os dados de treinamento são fundamentais e determinantes para o funcionamento e para a compreensão do que se entende como sendo a “inteligência” do sistema.

Um conjunto de dados pode espelhar inúmeras das relações, formas de organização e valores de uma sociedade. Portanto, em alguns casos, é possível que eles reflitam a desigualdade de gênero, a discriminação racial e as desigualdades interseccionais de gênero, raça e classe. Segundo Cathy O’neil⁴⁷⁶, os dados armazenados em *Big Data* nada mais são do que as codificações de dados sobre o nosso passado. Dessa forma, eles podem refletir situações, contextos e condições passadas de um indivíduo, que podem não ser as mesmas do momento em que a análise foi feita por algum sistema algorítmico, que realiza a correlação dos dados.

Nesse caso, o sujeito pode ser julgado ou ter suas condutas constantemente vigiadas e controladas com base em ações passadas, que podem não refletir as que possui no momento. Essa já é uma situação evidente, por exemplo, no setor judiciário de cidades como Filadélfia, na Pensilvânia (EUA), onde jovens infratores ou pessoas cumprem liberdade condicional tem suas ações e condutas acompanhadas através de sugestões algorítmicas fornecidas aos oficiais⁴⁷⁷, com base em dados e classificações que, ao identificar os sujeitos como sendo de “alto risco” ou “baixo risco” a partir de referências e imaginações morais providas por seres humanos⁴⁷⁸, definem a regularidade do acompanhamento.

Vivemos numa sociedade cuja fé na tecnologia para a solução de problemas e em sua neutralidade ainda é fortemente sustentada. Para essas crenças, diversos autores

474 PASQUINELLI, Matteo; JOLER, Vladan. O manifesto Nooscópio: Inteligência Artificial como Instrumento de Extrativismo do Conhecimento. **Lavits**, Rio de Janeiro, 30 de julho de 2020. Disponível em: <https://lavits.org/o-manifesto-nooscopio-inteligencia--artificial-como-instrumento-de-extrativismo-do-conhecimento/?lang=pt>. Acesso em: 25 de maio de 2023.

475 SILVA, Tarcízio. **Racismo algorítmico**: Inteligência Artificial e discriminação nas redes digitais. Sérgio Amadeu da Silveira (Org.). São Paulo: Edições SESC, 2022.

476 O’NEIL, Cathy. **Weapons of math destruction**: how big data increases inequality and threatens democracy. New York: Crown Publishers, 2016.

477 Esse exemplo pode ser visualizado e ouvido no documentário “*Coded Bias*” de Shalini Kantayya (2020), disponível na Netflix: <https://www.netflix.com/watch/81328723?source=35>. Acesso em: 02 de junho de 2023.

478 O’NEIL, Cathy. **Weapons of math destruction**: how big data increases inequality and threatens democracy. New York: Crown Publishers, 2016.

já utilizam termos, como “*solucionismo tecnológico*”⁴⁷⁹, “*tecnosolucionismo*”⁴⁸⁰ e “*tecnochauvinismo*”⁴⁸¹, que compreende não apenas a ideia de que a tecnologia é a solução, mas também trata, especificamente, da noção que se tem sobre os computadores serem mais “objetivos” ou “não possuírem vieses”, uma vez que suas respostas são fornecidas com bases em cálculos matemáticos⁴⁸².

A noção destacada no conceito “*tecnochauvinismo*” sobre a “objetividade” e “ausência de vieses” dos computadores poderia explicar, em parte, a atitude de alguns policiais, que decidem pela abordagem para verificação de pessoa apenas com o sinal de alerta ou enquadramento feito pelas câmeras de reconhecimento facial⁴⁸³. No entanto, entendemos que essa ação não se limita a crença no caráter “objetivo” e “neutro” da tecnologia. Longe disso, ela também pode revelar os valores, pontos de vista e imaginários em que, possivelmente, se baseiam as presunções de “suspeito”, “não suspeito” e “procurado” dos policiais que monitoram as câmeras.

Num país onde mais da metade das pessoas encarceradas são negras e a mídia televisiva veicula imagens de suspeitos e procurados por crimes, principalmente em regiões pobres e periféricas das grandes cidades, é provável que as decisões tomadas por gestores e policiais sobre a implementação e com base nos resultados das câmeras de reconhecimento facial constituam e se baseiem em uma imaginação carcerária, que fixa sempre os mesmos sujeitos e os mesmos lugares⁴⁸⁴ como “suspeitos” ou onde os crimes seriam mais “praticáveis”. Essa imaginação também poderia nos sugerir não só possíveis explicações para a instalação de um sistema de videomonitoramento com reconhecimento facial no Jacarezinho, como também para a regularidade de abordagens e prisões injustas de jovens, que, em sua maioria, são homens negros⁴⁸⁵.

479 MOROZOV, Evgeny. Solucionismo, nova aposta das elites globais. **Outras Palavras**, São Paulo, 23 de abril de 2020. Disponível em: <https://outraspalavras.net/tecnologiaemdisputa/solucionismo-nova-aposta-das-elites-globais/>. Acesso em: 19 de maio de 2023.

480 NOBREGA, Camila; VARON, Joana. Um olhar feminista para desmascarar tecnosolucionismos ambientais. **Le monde diplomatique Brasil**. São Paulo, 23 de maio de 2022. Disponível em: <https://diplomatique.org.br/um-olhar-feminista-para-desmascarar-tecnosolucionismos-ambientais>. Acesso em: 19 de maio de 2023.

481 BROUSSARD, Meredith. **Artificial Unintelligence: How Computers Misunderstand the World**. Massachusetts: The MIT Press, 2018.

482 Ibidem.

483 Peter Fussey e Daragh Murray (2020), em seu relatório “*Impact: Report on the police use of facial recognition technology identifies significant concerns*” sobre o uso do Reconhecimento Facial no Reino Unido, destaca o fato de policiais não esperarem pelo processo de decisão da sala de controle para a realização ou não da abordagem.

484 BENJAMIN, Ruha. Retomando nosso fôlego: estudos de ciência, tecnologia, teoria racial crítica e imaginação carcerária. In: SILVA, Tarcízio. **Comunidades, algoritmos e ativismos digitais: olhares afrodiaspóricos**. São Paulo: Literarua, 2020, p. 13-24.

485 GUIMARÃES, Hellen. Nos erros de reconhecimento facial, um “caso isolado” atrás do outro. **UOL PiauÍ**, 24 de setembro de 2021. Disponível em: <https://piaui.folha.uol.com.br/nos-erros-de-reconhecimento-facial-um-caso-isolado-atras-do-outro/>. Acesso em: 01 de junho de 2023.

Simone Browne⁴⁸⁶, ao discutir sobre essas novas formas de vigilância, destaca que a vigilância não é algo novo para a população negra, uma vez que alude às relações coloniais em que os escravos negros eram constantemente vigiados pelas patrulhas policiais. Portanto, o projeto de instalação das câmeras de vigilância no Jacarezinho, por exemplo, pode ser visto como uma atualização de antigas práticas de vigilância sobre a população negra. Já no caso de recorrência de abordagens e mandados de prisões injustas, em virtude do reconhecimento facial, acreditamos que a existência de uma seletividade penal no judiciário brasileiro⁴⁸⁷, os dados sobre a cor/raça dos encarcerados⁴⁸⁸ e uma visão racista sobre criminosos na sociedade contribua para que um grupo específico, o de jovens negros, compreenda a maioria dos detidos e encarcerados.

4.2. O problema da acurácia: o caso de Joy Buolamwini e os softwares de reconhecimento facial da Microsoft, IBM e Face ++

O caso de Joy Buolamwini pode ser visto como um dos mais emblemáticos para a compreensão dos problemas de acurácia das tecnologias de reconhecimento facial. A pesquisadora, cientista de dados e ativista, que se dedica à luta contra os vieses algorítmicos emitidos por sistemas de Inteligência Artificial, utiliza o seu exemplo pessoal para demonstrar que o reconhecimento facial pode não ser tão preciso sobre rostos de pessoas negras. Segundo Buolamwini⁴⁸⁹, ainda na época da faculdade, um robô social utilizado para interação e atividade com os estudantes através da aproximação do rosto, conseguia interagir com todos os seus colegas, menos com ela. Dessa forma, estando evidente essa distinção de rostos para a interação, a estudante presume que seja por causa da cor da sua pele e das suas características fenotípicas, por isso, decide colocar uma máscara branca, para tentar novamente, e, para a sua surpresa, o sistema a reconhece como um rosto humano e inicia a interação.

Essa é a experiência que levará a pesquisadora a investigar o caráter falho dos sistemas de reconhecimento facial, que parecem não “reconhecer” o rosto de indivíduos de pele negra e com determinadas características fenotípicas como sendo o rosto de uma pessoa. No caso relatado brevemente acima, Buolamwini⁴⁹⁰, ao propor uma auditoria do sistema, descobre que a tendência da máquina em identificar exclusivamente o rosto de

486 BROWNE, Simone. **Dark matters**: on the surveillance of blackness. Durham, NC: Duke University Press, 2015.

487 SILVA, Tarcízio. **Racismo algorítmico**: Inteligência Artificial e discriminação nas redes digitais. Sérgio Amadeu da Silveira (Org.). São Paulo: Edições SESC, 2022.

488 *Ibidem*.

489 BUOLAMWINI, Joy. **How I’m fighting bias in algorithms**. New York: TEDxBeaconStreet, 2016. Disponível em: https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms. Acesso em: 02 de abril de 2023.

490 *Ibidem*.

peças brancas se devia ao fato das imagens contidas no banco de dados não possuírem tanta diversidade e se tratarem, em sua maioria, de rostos de pessoas brancas

No artigo “*Gender shades: intersectional accuracy disparities in commercial gender classification*”, Joy Buolamwini e Timnit Gebru⁴⁹¹ propõem a auditoria de sistemas de reconhecimento facial vendidos por empresas como Microsoft, IBM e Face ++ com o objetivo de conhecer os seus níveis de acurácia sobre rostos de pessoas brancas e negras. O resultado do trabalho das autoras evidenciou que todos os sistemas performaram melhor em rostos de pessoas claras/brancas do que em rostos de pessoas escuras/negras⁴⁹². Além disso, elas também identificaram que todos os sistemas tiveram uma performance ruim no “reconhecimento” de pessoas do sexo feminino com rostos escuros/negros⁴⁹³. Por fim, todos os sistemas performaram melhor em rostos de indivíduos do sexo masculino, (independentemente de cor/raça), do que em indivíduos do sexo feminino⁴⁹⁴. Dessa forma, diante desses resultados, as pesquisadoras evidenciaram que as performances e taxas de erro dos sistemas de reconhecimento facial analisados poderiam diferir de forma considerável a depender de cada subgrupo sobre o qual era aplicado.

Os resultados encontrados por Buolamwini e Gebru⁴⁹⁵, ao confirmarem o caráter falho do reconhecimento facial sobre diferentes subgrupos e, principalmente sobre indivíduos de pele escura, sugere não só o quanto a adoção dessas tecnologias pode ser arriscada, como também pode funcionar como forma de exclusão e segregação social. No caso das câmeras de reconhecimento facial na segurança pública e a recorrência de prisões errôneas sobre jovens de pele negra, pensamos que as taxas de erro e acerto também podem estar relacionadas às imagens de “suspeitos” e “procurados” contidas na base de dados para treinamento do sistema, que podem ser, em sua maioria, de sujeitos de pele negra. Por outro lado, em virtude da falta de transparência dos programas utilizados, é possível que um software genérico seja utilizado pelas polícias estaduais brasileiras e os testes para a aceitação das taxas de erro não sejam feitos com o mesmo cuidado das avaliações de acurácia⁴⁹⁶. Buolamwini e Gebru⁴⁹⁷ salientam que até mesmo a posição, iluminação e expressão podem comprometer o nível de acerto e erro das análises de biometria facial automatizadas. A iluminação, principalmente, segundo elas, é de suma importância para a identificação da cor de pele.

491 BUOLAMWINI, Joy; GEBRU, Timnit. Gender shades: intersectional accuracy disparities in commercial gender classification. **Proceedings of Machine Learning Research**, v. 81, n. 1, p. 1-15, 2018.

492 *Ibidem*.

493 *Ibidem*.

494 *Ibidem*.

495 *Ibidem*.

496 SILVA, Tarcízio. **Racismo algorítmico: Inteligência Artificial e discriminação nas redes digitais**. Sérgio Amadeu da Silveira (Org.). São Paulo: Edições SESC, 2022.

497 BUOLAMWINI, Joy; GEBRU, Timnit. Gender shades: intersectional accuracy disparities in commercial gender classification. **Proceedings of Machine Learning Research**, v. 81, n. 1, p. 1-15, 2018.

5. CONSIDERAÇÕES FINAIS

No Brasil, o uso de câmeras de reconhecimento facial é feito sem muita preocupação com direitos, como o direito à privacidade, proteção de dados pessoais, bem como com a sua segurança nos espaços públicos, uma vez que vários são os “falsos positivos” e números de pessoas presas injustamente. Dessa forma, apresentamos que a existência de um imaginário sociotécnico⁴⁹⁸ sobre o ordenamento social com o avanço da ciência e tecnologia pode menosprezar direitos fundamentais e colocar em questão a noção de segurança, principalmente em relação a alguns grupos da sociedade.

O racismo e a discriminação emitida por tecnologias de reconhecimento facial podem ser resultantes do *Machine Learning* (“aprendizagem de máquina”) e dos dados utilizados para o seu treinamento. Evidenciamos que, no processo de aprendizagem da máquina, os vieses raciais e de gênero podem estar presentes no banco de dados a ser processado pelos seus algoritmos ou serem impostos no momento de rotulagem dos dados. Essa etapa é feita por operadores humanos, que podem usar termos antigos ou classificações binárias que contribuem para uma restrição da diversidade social⁴⁹⁹, principalmente em termos de cor/raça, e para a manutenção de ordens e posições sociais.

Para a disposição das tecnologias e imaginários sociotécnicos de fornecer resultados enviesados, que desfavorecem e prejudicam sujeitos e “minorias”, como os negros, Tarcízio Silva⁵⁰⁰ dá o nome de *racismo algorítmico*, que contribui para que antigas lógicas de segregação, desigualdades e preconceito sejam asseguradas. No caso do reconhecimento facial, compreendemos que o racismo algorítmico pode colaborar para a manutenção de práticas e conjecturas racistas, que favorecem a continuação do racismo estrutural. Além disso, pensamos que, em muitos casos, essas ações e desconfianças, principalmente em relação a indivíduos de pele negra se devem, em parte, aos dados dos encarcerados, a veiculação de imagens de suspeitos e procurados feitos pela mídia televisiva e a pontos de vista preconceituosos, que constituem e se pautam em uma “imaginação carcerária”⁵⁰¹, que se reflete no número de indivíduos negros presos injustamente pelo uso de câmeras de reconhecimento facial e justifica a atitude dos policiais de efetuarem a abordagem com o menor sinal de alerta e não esperarem pela

498 JASANOFF, Sheila. Future imperfect: Science, Technology and the Imaginations of modernity. In: JASANOFF, Sheila; KIM, Sang-Hyun. **Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power**. Chicago: University of Chicago Press, 2015, p. 1-33.

499 PASQUINELLI, Matteo; JOLER, Vladan. O manifesto Nooscópio: Inteligência Artificial como Instrumento de Extrativismo do Conhecimento. **Lavits**, Rio de Janeiro, 30 de julho de 2020. Disponível em: <https://lavits.org/o-manifesto-nooscopio-inteligencia--artificial-como-instrumento-de-extrativismo-do-conhecimento/?lang=pt>. Acesso em: 25 de maio de 2023.

500 SILVA, Tarcízio. **Racismo algorítmico: Inteligência Artificial e discriminação nas redes digitais**. Sérgio Amadeu da Silveira (Org.). São Paulo: Edições SESC, 2022.

501 BENJAMIN, Ruha. Retomando nosso fôlego: estudos de ciência, tecnologia, teoria racial crítica e imaginação carcerária. In: SILVA, Tarcízio. **Comunidades, algoritmos e ativismos digitais: olhares afrodiaspóricos**. São Paulo: Literarua, 2020, p. 13-24.

verificação das salas de controle⁵⁰², que confirmam o *match* do rosto do indivíduo filmado, com o do sujeito “suspeito” ou “procurado”.

Por fim, compreendemos que as taxas de erro e prisões injustas pelo uso de câmeras de reconhecimento facial na segurança pública revelam não apenas a presença de vieses raciais, que podem ser introduzidos no processo de aprendizagem da máquina e resultar em racismo algorítmico, contribuindo para assegurar antigas lógicas de preconceito e discriminação e garantindo a continuação do racismo estrutural, como também evidenciam a existência de uma *imaginação carcerária*⁵⁰³, constituída por dados sobre cor/raça dos encarcerados, pela veiculação de imagens de “procurados” e “suspeitos” pela mídia televisiva e pelas visões racistas e preconceituosas sobre quem são esses sujeitos em nossa sociedade. Em conclusão, o relato experiência de Joy Buolamwini⁵⁰⁴ e o seu estudo com Timnit Gebru⁵⁰⁵ apresentados aqui brevemente, não só corrobora a existência de vieses raciais e de gênero nas análises feitas por reconhecimento facial, como também nos alerta sobre os riscos de que as taxas de erro contribuam para a segregação racial e desfavoreçam e prejudiquem sujeitos negros⁵⁰⁶, tidos como “minorias”.

6. Referências

ABRAMOVAY, Ricardo. Movimento por banir uso de reconhecimento facial cresce no mundo. **Folha de São Paulo**. São Paulo, 14 de dezembro de 2019. Disponível em: <https://www1.folha.uol.com.br/ilustrissima/2019/12/movimento-por-banir-uso-de-reconhecimento-facial-cresce-no-mundo.shtml>. Acesso em: 01 de abril de 2023.

ALBUQUERQUE, Ana Luiza. Em fase de testes, reconhecimento facial no Rio falha no 2º dia. **Folha de São Paulo**. São Paulo, 17 de Julho de 2019. Disponível em: <https://www1.folha.uol.com.br/cotidiano/2019/07/em-fase-de-testes-reconhecimento-facial-no-rio-falha-no-2o-dia.shtml>. Acesso em: 01 de abril de 2023.

ALVES, Sarah. Além do racismo, reconhecimento facial erra mais em pessoas trans. **UOL**. São Paulo, 14 de fevereiro de 2021. Disponível em: <https://www.uol.com.br/tilt/noticias/>

502 FUSSEY, Peter; MURRAY, Daragh. **Impact: Report on the police use of facial recognition technology identifies significant concerns**. Colchester: University of Essex, 2020. Disponível em: <https://www.essex.ac.uk/research/showcase/report-on-the-police-use-of-facial-recognition-technology-identifies-significant-concerns>. Acesso em: 01 de junho de 2023.

503 BENJAMIN, Ruha. Retomando nosso fôlego: estudos de ciência, tecnologia, teoria racial crítica e imaginação carcerária. In: SILVA, Tarcízio. **Comunidades, algoritmos e ativismos digitais: olhares afrodiaspóricos**. São Paulo: Literarua, 2020, p. 13-24.

504 BUOLAMWINI, Joy. **How I’m fighting bias in algorithms**. New York: TEDxBeaconStreet, 2016. Disponível em: https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms. Acesso em: 02 de abril de 2023.

505 BUOLAMWINI, Joy; GEBRU, Timnit. Gender shades: intersectional accuracy disparities in commercial gender classification. **Proceedings of Machine Learning Research**, v. 81, n. 1, p. 1-15, 2018.

506 SILVA, Tarcízio. **Racismo algorítmico: Inteligência Artificial e discriminação nas redes digitais**. Sérgio Amadeu da Silveira (Org.). São Paulo: Edições SESC, 2022.

redacao/2021/02/14/nao-e-so-racismo-reconhecimento-facial-tambem-erra-mais-em-pessoas-trans.htm. Acesso em: 12 de abril de 2023.

ARBEX, Pedro; SAMOR, Geraldo. Aconteceu no Leblon... Conheça a Gabriel, a startup anjo da guarda. **Brazil Journal**. Rio de Janeiro, 01 de novembro de 2020. Disponível em: <https://braziljournal.com/aconteceu-no-leblon-conheca-a-gabriel-a-startup-anjo-da-guarda/>. Acesso em: 22 de agosto de 2022.

BARBOSA, Arthur Almeida Meneses; OLIVEIRA, Celso Eduardo Lins de; SANTARÉM, Paulo Rená da Silva; SANTOS, Natane; SILVA, Fernanda Santos Rodrigues; SILVA, Valdinei Freire da. **Documento Preto I**. Brasil: Aqualtune Lab, 2022.

BARDIN, Laurence. **Análise de Conteúdo**. Tradução de Luís Antero Reto e Augusto Pinheiro São Paulo: Edições 70, 2011.

BENJAMIN, Ruha. Retomando nosso fôlego: estudos de ciência, tecnologia, teoria racial crítica e imaginação carcerária. In: SILVA, Tarcízio. **Comunidades, algoritmos e ativismos digitais: olhares afrodiáspóricos**. São Paulo: Literarua, 2020, p. 13-24.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018, Brasília – DF, 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709compilado.htm. Acesso em 13 de dezembro de 2022.

BRASIL. Lei nº 8.666, de 21 de junho de 1993, Brasília – DF, 1993. Disponível em: <https://www.gov.br/saude/pt-br/aceso-a-informacao/banco-de-precos/legislacao/lei-no-8-666-de-21-de-junho-de-1993.pdf/view>. Acesso em: 25 de maio de 2023.

BROUSSARD, Meredith. Artificial **Unintelligence**: How Computers Misunderstand the World. Massachusetts: The MIT Press, 2018.

BROWNE, Simone. **Dark matters: on the surveillance of blackness**. Durham, NC: Duke University Press, 2015.

BRUNO, Fernanda. **Máquinas de ver, modos de ser: vigilância, tecnologia e subjetividade**. Porto Alegre: Sulina, 2013.

BRUNO, Fernanda. Monitoramento, classificação e controle nos dispositivos de vigilância digital. **Revista FAMECOS**. Porto Alegre, n. 36, p. 10-16, agosto, 2008.

BUOLAMWINI, Joy. **How I'm fighting bias in algorithms**. New York: TEDxBeaconStreet, 2016. Disponível em: https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms. Acesso em: 02 de abril de 2023.

BUOLAMWINI, Joy; GEBRU, Timnit. Gender shades: intersectional accuracy disparities

in commercial gender classification. **Proceedings of Machine Learning Research**, v. 81, n. 1, p. 1-15, 2018.

CRUZ, Luã. Where is Facial Recognition Technology heading in Brazil? In: BRANDUSESCU, Ana; REIA, Jess (Eds). **Artificial Intelligence in the city: building civic engagement and public trust**. Centre for Interdisciplinary Research on Montreal, McGill University, 2022, p. 62-64.

FALTAY, Paulo. Parlamentares de todas as regiões do Brasil apresentam projetos de lei pelo banimento do reconhecimento facial em espaços públicos. **Lavits**. Rio de Janeiro, 21 de junho de 2022. Disponível em: <https://lavits.org/parlamentares-de-todas-as-regioes-do-brasil-apresentam-projetos-de-lei-pelo-banimento-do-reconhecimento-facial-em-espacos-publicos/>. Acesso em: 26 de julho de 2022.

FEENBERG, Andrew. Racionalização democrática, poder e tecnologia. In: NEEDER, Ricardo T. (org.). **Ciclo de conferências Andrew Feenberg**, 2010.

FUSSEY, Peter; MURRAY, Daragh. **Impact: Report on the police use of facial recognition technology identifies significant concerns**. Colchester: University of Essex, 2020. Disponível em: <https://www.essex.ac.uk/research/showcase/report-on-the-police-use-of-facial-recognition-technology-identifies-significant-concerns>. Acesso em: 01 de junho de 2023.

GOMES, Helton Simões. Por que uma das maiores cidades dos EUA baniu o reconhecimento facial?. **UOL**. São Paulo, 16 de maio de 2019. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2019/05/16/por-que-uma-das-maiores-cidades-dos-eua-baniu-o-reconhecimento-facial.htm>. Acesso em: 01 de abril de 2023.

GUIMARÃES, Hellen. Nos erros de reconhecimento facial, um “caso isolado” atrás do outro. **UOL PiauÍ**, 24 de setembro de 2021. Disponível em: <https://piaui.folha.uol.com.br/nos-erros-de-reconhecimento-facial-um-caso-isolado-atras-do-outro/>. Acesso em: 01 de junho de 2023.

HARDING, Sandra. Gênero, democracia e filosofia da ciência. **RECIIS – Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**. Rio de Janeiro, v. 1, n. 1, p. 163-168, janeiro-junho, 2007.

HERINGER, Carolina. Uma das principais promessas de campanha de Witzel, câmeras de reconhecimento facial não funcionam mais desde o fim de 2019. **Extra**. Rio de Janeiro, 20 de julho de 2020. Disponível em: <https://extra.globo.com/casos-de-policia/uma-das-principais-promessas-de-campanha-de-witzel-cameras-de-reconhecimento-facial-nao-funcionam-mais-desde-fim-de-2019-24541055.html>. Acesso em 25 de março de 2023.

JASANOFF, Sheila. Future imperfect: Science, Technology and the Imaginations of modernity. In: JASANOFF, Sheila; KIM, Sang-Hyun. **Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power**. Chicago: University of Chicago Press, 2015, p. 1-33.

KER, João. Vestibular da Unicamp terá reconhecimento facial para evitar fraudes em cotas. **UOL**. São Paulo, 01 de agosto de 2022. Disponível em: https://educacao.uol.com.br/noticias/agencia-estado/2022/08/01/vestibular-da-unicamp-tera-reconhecimento-facial-para-evitar-fraudes-em-cotas.htm?utm_source=twitter&utm_medium=social-media&utm_content=geral&utm_campaign=noticias. Acesso em: 23 de agosto de 2022.

KIM, Joon Ho. Cibernética, ciborgues e ciberespaço: notas sobre as origens da cibernética e sua reinvenção cultural. **Horiz. antropol.**, Porto Alegre, v. 10, n. 21, p. 199-219, Junho, 2004. Disponível em: <https://www.scielo.br/j/ha/a/R8fbcHwxmPrw3C3XmYKbg3c/?lang=pt>. Acesso em 25 de março de 2021.

MARSHALL, Lisa. Facial recognition software has a gender problem. **Universidade do Colorado**. Colorado – EUA, 08 de outubro de 2019. Disponível em: <https://www.colorado.edu/today/2019/10/08/facial-recognition-software-has-gender-problem>. Acesso em: 05 de maio de 2023.

MOROZOV, Evgeny. Solucionismo, nova aposta das elites globais. **Outras Palavras**, São Paulo, 23 de abril de 2020. Disponível em: <https://outraspalavras.net/tecnologiaemdisputa/solucionismo-nova-aposta-das-elites-globais/>. Acesso em: 19 de maio de 2023.

NOBREGA, Camila; VARON, Joana. Um olhar feminista para desmascarar tecnosolucionismos ambientais. **Le monde diplomatique Brasil**. São Paulo, 23 de maio de 2022. Disponível em: <https://diplomatique.org.br/um-olhar-feminista-para-desmascarar-tecnosolucionismos-ambientais/>. Acesso em: 19 de maio de 2023.

NUNES, Pablo. **Um Rio de câmeras com olhos seletivos**: uso do reconhecimento facial pela polícia fluminense. Rio de Janeiro: CESeC, 2022.

OLIVEIRA, Samuel Rodrigues de. **SORRIA, VOCÊ ESTÁ SENDO FILMADO**: tecnologias de reconhecimento facial, privacidade e proteção de dados. Juiz de Fora, 2020. (Trabalho de Dissertação do Programa de Pós-graduação em Direito – UFJF).

O'NEIL, Cathy. **Weapons of math destruction**: how big data increases inequality and threatens democracy. New York: Crown Publishers, 2016.

O PANÓPTICO. 2020. Por que o uso de reconhecimento facial na segurança é controverso? Disponível em: <https://opanoptico.com.br/por-que-o-uso-de-reconhecimento-facial-na-seguranca-e-controverso/>. Acesso em: 30 de novembro de 2022.

PASQUINELLI, Matteo. Machines that Morph Logic: Neural Networks and the Distorted Automation of Intelligence as Statistical Inference. **Logic Gate: the Politics of the Artifactual Mind (Online)**. 2017. Disponível em: <https://www.glass-bead.org/article/machines-that-morph-logic/?lang=enview>. Acesso em: 20 de agosto de 2022.

PASQUINELLI, Matteo; JOLER, Vladan. O manifesto Nooscópio: Inteligência Artificial como Instrumento de Extrativismo do Conhecimento. **Lavits**, Rio de Janeiro, 30 de julho de 2020. Disponível em: <https://lavits.org/o-manifesto-nooscopio-inteligencia--artificial-como-instrumento-de-extrativismo-do-conhecimento/?lang=pt>. Acesso em: 25 de maio de 2023.

SANTOS, Laymert G. A informação após a virada cibernética. In. SANTOS, Laymert Garcia dos; KHEL, Maria Rita; KUCINSKI, Bernardo; PINHEIRO, Walter. **Revolução tecnológica, internet e socialismo**. São Paulo, SP: Editora Fundação Perseu Abramo, p. 9-33, 2003.

SECRETARIA DE ESTADO DE POLÍCIA MILITAR DO RIO DE JANEIRO. Estudo técnico preliminar, sistema de videomonitoramento (Jacarezinho). Rio de Janeiro – RJ, 2020. Disponível em: https://drive.google.com/file/d/190WFEcBcZdm5ORiGjNk_a6v0IZnJcGva/view. Acesso em: 01 de março de 2023.

SECRETARIA DE ESTADO DE POLÍCIA MILITAR DO ESTADO DO RIO DE JANEIRO. Câmeras de reconhecimento facial (ordem de prioridade). Rio de Janeiro – RJ, 2019. Disponível: https://docs.google.com/document/d/1dg9OxYhP9sLzlkq_UOmsdsBc-71HCHdW/edit. Acesso em: 01 de março de 2023.

SECRETARIA DE ESTADO DE POLÍCIA MILITAR DO ESTADO DO RIO DE JANEIRO. Termo de Cooperação Técnica. Rio de Janeiro – RJ, 2019. Disponível em: <https://docs.google.com/document/d/1s1zv5PjnAJFx7DPDFCrqnC7AnAqcFUEB/edit?pli=1>. Acesso em: 01 de março de 2023.

SENA, Anelice; BORGES, Juliana. Escolas de Nova Venécia usam reconhecimento facial para controlar frequência e desperdício de merenda. **G1**. Espírito Santo, 03 de abril de 2018. Disponível em: <https://g1.globo.com/es/espírito-santo/noticia/escolas-de-nova-venecia-usam-reconhecimento-facial-para-controlar-frequencia-e-desperdicio-de-merenda.ghtml>. Acesso em: 05 de dezembro de 2022.

SILVA, Mariah Rafaella; VARON, Joana. **Reconhecimento Facial no setor público e identidades trans**: tecnopolíticas de controle e ameaça à diversidade de gênero em suas interseccionalidades de raça, classe e território. Brasil: Coding Rights, 2021.

SILVA, Tarcízio. **Racismo algorítmico**: Inteligência Artificial e discriminação nas redes digitais. Sérgio Amadeu da Silveira (Org.). São Paulo: Edições SESC, 2022.

STENTZLER, Isadora. Empresa propõe sistema chinês para a proteção de estudantes da

região de Campinas. **Correio Popular**. Campinas, 13 de julho de 2022. Disponível em: <https://correio.rac.com.br/campinasermc/empresa-prop-e-sistema-chines-para-a-protec-o-de-estudantes-da-regi-o-de-campinas-1.1263738>. Acesso em: 20 de agosto de 2022.

WINNER, Langdon. **Artefatos têm política?** Traduzido por Fernando Manso. Chicago: The University of Chicago Press, 1986, p. 19-39.



INSTITUTO
DE REFERÊNCIA
EM INTERNET
E SOCIEDADE