UNIVERSITÀ
DEGLI STUDI
FIRENZE

JUSTIÇA PREDITIVA E PROTEÇÃO DE GRUPOS VULNERÁVEIS: ANÁLISE DOS VIESES DISCRIMINATÓRIOS DECORRENTES DA AUTOMAÇÃO JUDICIAL

RAMON ARANHA DA CRUZ

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS JURÍDICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS JURÍDICAS

RAMON ARANHA DA CRUZ

JUSTIÇA PREDITIVA E PROTEÇÃO DE GRUPOS VULNERÁVEIS: ANÁLISE DOS VIESES DISCRIMINATÓRIOS DECORRENTES DA AUTOMAÇÃO JUDICIAL

JOÃO PESSOA-PB
2025

UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA IN SCIENZE GIURIDICHE

CICLO XXXVII

JUSTIÇA PREDITIVA E PROTEÇÃO DE GRUPOS VULNERÁVEIS: ANÁLISE DOS VIESES DISCRIMINATÓRIOS DECORRENTES DA AUTOMAÇÃO JUDICIAL

Settore Scientifico Disciplinare IUS/19

Dottorando
Dott. Aranha da Cruz Ramon

Tutore
Prof. Pietropaoli Stefano

Tutore
Prof. Palitot Braga Rômulo Rhemo

Coordinatrice
Prof. Vallauri Maria Luisa

Anni 2021/2025

RAMON ARANHA DA CRUZ

JUSTIÇA PREDITIVA E PROTEÇÃO DE GRUPOS VULNERÁVEIS: ANÁLISE DOS VIESES DISCRIMINATÓRIOS DECORRENTES DA AUTOMAÇÃO JUDICIAL

Tese apresentada ao Programa de Pós-Graduação em Ciências Jurídicas da Universidade Federal da Paraíba em cotutela com o *Dipartimento di Scienze Giuridiche da Università Degli Studi di Firenze*, como requisito para avaliação final do Doutorado em Ciências Jurídicas (UFPB) e *Dottorato di Ricerca in Scienze Giuridiche* (UNIFI).

Área de Concentração: Direitos Humanos e Desenvolvimento
Linha de Pesquisa: Teorias e História do Direito
– Teoria e História dos Direitos Humanos (UFPB) / *Teorie dei diritti umani – Diritto e società, genealogia e prospettive del pensiero giuridico* (UNIFI)

Orientador: Prof. Dr. Rômulo Rhemo Palitot Braga
Co-Orientador: Prof. Dr. Stefano Pietropaoli

JOÃO PESSOA-PB
2025

UNIVERSIDADE FEDERAL DA PARAÍBA (UFPB)

UNIVERSITÀ DEGLI STUDI FIRENZE (UNIFI)

ATA DE DEFESA DE DOUTORADO

DOUTORADO EM REGIME DE COTUTELA

Ata da Banca Examinadora do Doutorando **RAMON ARANHA DA CRUZ** candidato ao grau de Doutor em Ciências Jurídicas.

Às 14h00 do dia 04 de fevereiro de 2025, por meio de ambiente virtual (**https://meet.google.com/wog-bjit-irg**), em cumprimento ao **CONVÊNIO DE COTUTELA INTERNACIONAL DE TESE** firmado entre a Universidade Federal da Paraíba (UFPB) e a Università Degli Studi Firenzi (UNIFI), constante no Processo Administrativo n° 23074.030198/2022-03, reuniu-se a Comissão Examinadora formada pelos seguintes Professores Doutores: Romulo Rhemo Palitot Braga (Orientador PPGCJ/UFPB), Stefano Pietropaoli (Orientador/UNIFI), Gustavo Barbosa de Mesquita Batista (Avaliador Interno - PPGCJ/UFPB), Márcio Flávio Lins de Albuquerque e Souto (Avaliador Interno - PPGCJ/UFPB), Carlo Botrugno (Avaliador Externo – UNIFI), Emilio Santoro (Avaliador Externo – UNIFI) e Felix Araújo Neto (Avaliador Externo – UEPB), para avaliar a tese de Doutorado do aluno Ramon Aranha da Cruz, intitulada: "**JUSTIÇA PREDITIVA E PROTEÇÃO DE GRUPOS VULNERÁVEIS: análise dos vieses discriminatórios decorrentes da automação judicial**", candidato ao grau de Doutor em Ciências Jurídicas, área de concentração em Direitos Humanos e Desenvolvimento, pela UFPB e de Dottore in Scienze Giuridiche pela UNIFI. Compareceram à cerimônia, além do candidato, professores, alunos e convidados. Dando início à solenidade, o professor Romulo Rhemo Palitot Braga (Orientador PPGCJ/UFPB) apresentou a Comissão Examinadora, passando a palavra ao doutorando, que discorreu sobre o tema dentro do prazo e termos determinados no convênio de cotutela firmado. O candidato foi a seguir arguido pelos examinadores na forma regimental. Ato contínuo, passou então a Comissão, em caráter secreto, à avaliação e ao julgamento do referido trabalho, concluindo por atribuir-lhe o conceito **APROVAÇÃO DA PARTE DA COMISSÃO EXAMINADORA**, o qual foi proclamado pela Presidência da Comissão, achando-se o candidato legalmente habilitado a receber o grau de Doutor em Ciências Jurídicas, cabendo à Universidade Federal da Paraíba providenciar, como de direito, o diploma de Doutor a que este faz jus, ocorrendo o mesmo em relação à Università Degli Studi di Firenze quanto ao grau de Dottore in Scienze Giuridiche. Nada mais havendo a declarar, a presidência deu por encerrada a sessão, da qual eu, Rosandro Barros da Silva Souza, secretário do Programa de Pós-Graduação em Ciências Jurídicas,

Centro de Ciências Jurídicas – CCJ – UFPB – Campus I
Cidade Universitária – João Pessoa – PB – Brasil – CEP 58051-900 Fone +55(83)3216-7627

lavrei a presente ata, que assino juntamente com os demais membros da banca, para fins de certificar a realização desta defesa. João Pessoa-PB, 04 de fevereiro de 2025. xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Assinado Digitalmente - SIPAC/UFPB
Prof. Dr. Romulo Rhemo Palitot Braga

Prof. Dr. Stefano Pietropaoli

Assinado Digitalmente - SIPAC/UFPB
Prof. Drª Gustavo Barbosa de Mesquita Batista

Assinado Digitalmente - SIPAC/UFPB
Prof. Dr. Márcio Flávio Lins de Albuquerque e Souto

Prof. Dr. Carlo Botrugno

Prof. Dr. Emilio Santoro

Prof. Dr. Felix Araújo Neto

Centro de Ciências Jurídicas – CCJ – UFPB – Campus I
Cidade Universitária – João Pessoa – PB – Brasil – CEP 58051-900 Fone +55(83)3216-7627

Emitido em 04/02/2025

ATA Nº 01/2025 - PPGCJ (11.01.46.04)
(Nº do Documento: 1)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 12/02/2025 10:28)
GUSTAVO BARBOSA DE MESQUITA BATISTA
COORDENADOR CURS/POS-GRADUACAO
1453013

(Assinado digitalmente em 11/02/2025 17:04)
ROMULO RHEMO PALITOT BRAGA
PROFESSOR DO MAGISTERIO SUPERIOR
1640096

(Assinado digitalmente em 14/02/2025 11:08)
MARCIO FLAVIO LINS DE ALBUQUERQUE E SOUTO
PROFESSOR DO MAGISTERIO SUPERIOR
1719570

(Assinado digitalmente em 11/02/2025 11:57)
ROSANDRO BARROS DA SILVA SOUZA
ASSISTENTE EM ADMINISTRACAO
1023010

Para verificar a autenticidade deste documento entre em https://sipac.ufpb.br/documentos/ informando seu número: **1**, ano: **2025**, documento (espécie): **ATA**, data de emissão: **11/02/2025** e o código de verificação: **b86ab46071**

Catalogação na publicação
Seção de Catalogação e Classificação

C957j Cruz, Ramon Aranha da.
Justiça preditiva e proteção de grupos vulneráveis : análise dos Vieses discriminatórios decorrentes da automação judicial / Ramon Aranha da Cruz. - João Pessoa, 2025.
188 f. : il.

Orientação: Rômulo Rhemo Palitot Braga.
Coorientação: Stefano Pietropaoli.
Tese (Doutorado) - UFPB/CCJ.

1. Direitos humanos. 2. Justiça preditiva. 3. Grupos vulneráveis - Discriminação. I. Braga, Rômulo Rhemo Palitot. II. Pietropaoli, Stefano. III. Título.

UFPB/BC CDU 342.7(043)

Elaborado por ANNA REGINA DA SILVA RIBEIRO - CRB-15/24

À minha família, por quem busco a cada dia ser sempre uma pessoa melhor.

AGRADECIMENTOS

Recebi da minha esposa a notícia da aprovação no doutorado enquanto estava em uma maca, me dirigindo à UTI. A pandemia da COVID-19 deixou sua marca em nossa família, ainda que, graças a Deus, tenhamos superado.

Após 40 dias internado, tendo enfrentado gravíssimo quadro de saúde, consegui retornar à minha casa onde, ainda de cama, assisti às primeiras aulas do doutorado.

A conclusão deste doutorado não é só uma realização pessoal. Ela é a prova de que Deus cuida de mim e de minha família. É impossível negar o milagre operado em minha vida. A Ele, portanto, rendo graças por poder estar aqui, neste momento, escrevendo estes agradecimentos.

Minha família, que tanto sofreu durante este período, também foi quem mais me apoiou durante a jornada de doutoramento. Agradeço à minha esposa, Adriane, que esteve presente nos momentos mais críticos de minha internação mas também cuidou de nossa casa e família enquanto estive sozinho na cidade de Florença. Seu apoio e suporte foram fundamentais para conclusão desta etapa. À minha mãe, Simone Dália, que sempre me inspirou e incentivou a ingressar na vida acadêmica, mas também nunca deixou de rezar por mim durante esse período difícil, meu muito obrigado. Aos meus filhos, que não fazem ideia, mas sempre me incentivaram a chegar até aqui, para mostra-los que podemos alcançar tudo aquilo que nós desejamos e nos esforçamos para conseguir, e quão abençoada pode ser a trajetória se colocamos Deus à frente de nossos planos.

Agradeço às instituições de ensino que me possibilitaram o meu doutoramento, a Universidade Federal da Paraíba e a Università Degli Studi di Firenze, nas pessoas de meus orientadores Prof. Dr. Rômulo Rhemo Palitot Braga e Prof. Dr. Stefano Pietropaoli. Agradeço por cada ensinamento, cada apoio e cada incentivo dado. Vocês tornaram a caminhada mais leve.

Por fim, um agradecimento a todos ou outros que, direta ou indiretamente, fizeram parte desse ciclo tão cheio de desafios que agora chega ao fim e abre caminho para tantos outros.

Muitas são, Senhor meu Deus, as maravilhas que tens operado para conosco, e os teus pensamentos não se podem contar diante de ti; se eu os quisera anunciar, e deles falar, são mais do que se podem contar. (Sl 40:5).

Lawlessness is at a premium, woeful penalty it brings,
Relic of the middle ages is the present state of things.
To the winds we now are sowing, and the whirl-wind comes at length,
Evils cast upon the waters come again with added strength.

(Lizelia Augusta Jenkins Moorer)

RESUMO

CRUZ, Ramon Aranha da. **Justiça Preditiva e Proteção De Grupos Vulneráveis: Análise Dos Vieses Discriminatórios Decorrentes Da Automação Judicial**. 2023. 143 f. Tese (Doutorado em Ciências Jurídicas) – Programa de Pós-Graduação em Ciências Jurídicas, Universidade Federal da Paraíba, João Pessoa/PB, 2023 em cotutela com a Università Degli Studi di Firenze (Teoria dei diritti umani – diritto e società, genealogia e prospettive del pensiero giuridico).

A presente investigação doutoral tem como objeto de estudo as novas formas de vulnerabilização sofridas por grupos socialmente vulneráveis, através da utilização de inteligência artificial na automação de decisões. Partindo desta problemática, tem-se por hipótese a afirmação de que o desenvolvimento de algoritmos pautados originalmente nos Direitos Humanos conduzirá diretamente ao fim de atos discriminatórios decorrentes de automação judicial. O objetivo geral desta tese é a análise de como seria possível implementar ferramentas para proteção de grupos vulneráveis submetidos à avaliação por softwares de inteligência artificial na Justiça Criminal. Os objetivos específicos são: a) examinar o processo de criação de ferramentas de inteligência artificial utilizadas como auxiliares na atividade judicante estatal e justiça preditiva, no intuito de compreender seu funcionamento; b) estudar a pertinência ética acerca da utilização de softwares inteligentes em processos decisórios por entes públicos e privados; c) analisar textos legislativos que se propõem a regulamentar a matéria e também se a resposta a tais violações pode ser obtida através da regulamentação; d) investigar o processo de vulnerabilização de minorias sociais, suas causas e principais fatores de marginalização; e) avaliar e propor políticas públicas para evitar a vitimização de grupos vulneráveis submetidos à apreciação automatizada pelo Poder Judiciário e demais entes públicos. Do ponto de vista metodológico, a tese se classifica como exploratória e parte de um método hipotético-dedutivo, utilizando-se como técnica a pesquisa bibliográfica e documental, por meio de livros, artigos científicos e legislação. Ao final da pesquisa, foi possível confirmar a hipótese proposta, seja pelas evidências científicas demonstradas como pela própria adesão em massa do sistema regulatório pelos grandes Estados, evidenciando a necessidade de regulamentação para proteção de grupos vulneráveis submetidos às decisões automatizadas no Poder Judiciário.

Palavras-chaves: direitos humanos; justiça preditiva; grupos vulneráveis; discriminação.

SINTESI

CRUZ, Ramon Aranha da. **Giustizia Predittiva e Protezione dei Gruppi Vulnerabili: Analisi dei Pregiudizi Discriminatori Emergenti a Seguito dell'Automazione Giudiziaria**. 2023. 143 p. Tesi (Dottorato in Scienze Giuridiche) – Programma di Laurea Magistrale in Scienze Giuridiche, Università Federale della Paraíba, João Pessoa/PB, 2023, in cotutela con l'Università Degli Studi di Firenze (Teoria dei diritti umani – diritto e società, genealogia e prospettive del pensiero giuridico).

La presente indagine dottorale ha come oggeto di studio le nuove forme di vulnerabilizzazione subite dai gruppi socialmente vulnerabili, attraverso l'utilizzo dell'intelligenza artificiale nell'automazione delle decisioni. Partendo da questa problematica, si ipotizza che lo sviluppo di algoritmi basati originariamente sui Diritti Umani porterà direttamente alla fine degli atti discriminatori derivanti dall'automazione giudiziaria. L'obiettivo generale di questa tesi è l'analisi di come sia possibile implementare strumenti per la protezione dei gruppi vulnerabili sottoposti a valutazione da parte di software di intelligenza artificiale nella Giustizia Criminale. Gli obiettivi specifici sono: a) esaminare il processo di creazione di strumenti di intelligenza artificiale utilizzati come ausiliari nell'attività giudiziaria statale e nella giustizia predittiva, al fine di comprenderne il funzionamento; b) studiare la pertinenza etica dell'utilizo di software intelligenti nei processi decisionali da parte di enti pubblici e privati; c) analizzare i testi legislativi che si propongono di regolamentare la materia e se la risposta a tali violazioni può essere ottenuta attraverso la regolamentazione; d) investigare il processo di vulnerabilità delle minoranze sociali, le sue cause e i principali fattori di marginalizzazione; e) valutare e proporre politiche pubbliche per evitare la vittimizzazione dei gruppi vulnerabili sottoposti a valutazione automatizzata da parte del Potere Giudiziario e altri enti pubblici. Dal punto di vista metodologico, la tesi si classifica come esplorativa e si basa su un metodo ipotetico-deduttivo, utilizzando come tecnica la ricerca bibliografica e documentale, tramite libri, articoli scientifici e legislazione. Al termine della ricerca, è stato possibile confermare l'ipotesi proposta, sia attraverso le evidenze scientifiche dimostrate, sia attraverso l'adozione massiva del sistema normativo da parte dei grandi Stati, evidenziando la necessità di una regolamentazione a tutela dei gruppi vulnerabili sottoposti a decisioni automatizzate nel contesto Giudiziario.

Parole chiave: diritti umani; giustizia preditiva; gruppi vulnerabili; discriminazione.

ABSTRACT

CRUZ, Ramon Aranha da. Predictive Justice and Protection of Vulnerable Groups: Analysis of Discriminatory Biases Arising from Judicial Automation. 2023. 143 p. Thesis (Doctorate in Legal Sciences) – Postgraduate Program in Legal Sciences, Federal University of Paraíba, João Pessoa/PB, 2023, associated with Università Degli Studi di Firenze (Teoria dei diritti umani – diritto e società, genealogia e prospettive del pensiero giuridico).

This doctoral investigation focuses on the new forms of vulnerability experienced by socially vulnerable groups through the use of artificial intelligence in decision automation. Based on this problem, the hypothesis posits that the development of algorithms originally based on Human Rights will directly lead to the end of discriminatory acts resulting from judicial automation. The general objective of this thesis is to analyze how it would be possible to implement tools to protect vulnerable groups subjected to assessment by artificial intelligence software in Criminal Justice. The specific objectives are: a) examine the process of creating artificial intelligence tools used as aids in state judicial activity and predictive justice to understand their functioning; b) study the ethical relevance of using intelligent software in decision-making processes by public and private entities;
c) analyze legislative texts that aim to regulate the matter and whether the response to such violations can be obtained through regulation; d) investigate the process of vulnerability of social minorities, their causes, and main factors of marginalization; e) assess and propose public policies to prevent the victimization of vulnerable groups subjected to automated scrutiny by the Judiciary and other public entities. From a methodological perspective, the thesis is classified as exploratory and follows a hypothetical-deductive method, using bibliographic and documentary research as a technique, through books, scientific articles, and legislation. At the end of the research, it was possible to confirm the proposed hypothesis, both through the scientific evidence demonstrated and through the mass adoption of the regulatory system by large States, highlighting the need for regulation to protect vulnerable groups subjected to automated decisions in the Judiciary.

Keywords: human rights; predictive justice; vulnerable groups; discrimination.

LISTA DE SIGLAS

AIA	*Artificial Intelligence Act*
CEDH	Convenção Europeia dos Direitos do Homem
CNJ	Conselho Nacional de Justiça
COMPAS	*Correctional Offender Management Profiling for Alternative Sanctions*
CPP	Código de Processo Penal
CVDT	Convenção de Viena Sobre o Direito dos Tratados
ECHR	European Court of Human Rights
EPIC	Electronic Privacy Information Center
FONAMEC	Fórum Nacional de Mediação e Conciliação
IA	Inteligência Artificial
OEA	Organização dos Estados Americanos
ONU	Organização das Nações Unidas
PSA	Public Safety Assessment
SAVRY	*Structured Assessment of Violence Risk in Youth*
SSL	*Strategic Subject List*
STF	Supremo Tribunal Federal
SyRI	System Risk Indication
TAR	*Tribunale Amministrativo Regionale*)
UNESCO	Organização das Nações Unidas para Educação, a Ciência e a Cultura

SUMÁRIO

1. INTRODUÇÃO

Vive-se hoje uma época de surpreendentes avanços computacionais, em que surgem cada vez mais tecnologias só antes imaginadas na ficção. Informações são fornecidas por robôs através de comandos de voz, ligações de vídeo são realizadas e alguns já fizeram até viagens turísticas ao espaço. Relevantes progressos, é possível perceber, também se encontram presentes nas Cortes de Justiça.

Ao realizar estudos sobre a modernização do Poder Judiciário, cuja concepção geral é de um sistema rígido e pouco adepto às mudanças e avanços, vemos que houve claras transformações com fundamento no atual processo de softwares autônomos.

Historicamente, nenhuma mudança impactou tanto o Poder Judiciário quanto a computação. Esta, aliada à internet, possibilitou um aumento de eficiência e produtividade nunca antes vistos no cenário jurídico mundial. Tais avanços, inobstante, foram acompanhados de diversas transformações sociais que também repercutiram na quantidade de demandas novas submetidas ao crivo forense.

O atual panorama do judiciário brasileiro demonstra a necessidade de mudanças de abordagem e melhor gestão dos processos a ele submetidos, principalmente diante da enorme quantidade de feitos pendentes de resolução. Segundo o último Justiça em Números, relatório anual apresentado pelo Conselho Nacional de Justiça, no qual são consolidadas e sistematizadas todas as informações dos Tribunais do país, o Poder Judiciário encerrou o ano de 2022 com a enorme quantidade de 81,4 milhões processos, apontando uma tendência de aumento desde o ano de 2020 (CONSELHO NACIONAL DE JUSTIÇA, 2023). Dos dados é possível se concluir pela existência de uma tendência de redução na cultura de litígio nacional, que pode ser explicada pela pandemia da COVID-19 ou como resultado das políticas de solução alternativa de conflitos. Contudo, melhor análise somente poderá ser empreendida com novos dados coletados ao longo dos próximos anos.

Considerando tais números, os Órgãos de cúpula têm realizado diversas mobilizações para a diminuição do número de ações, como a instituição de Semanas da Conciliação, período em que os Órgãos Judiciais condensam esforços para obtenção de acordos, e a criação de um Fórum Nacional de Mediação e Conciliação - FONAMEC, promovendo discussões para aprimoramento dos métodos consensuais de solução de conflitos.

É possível observar que ocorreu nos últimos anos uma valorização dos métodos alternativos para solução de conflitos capitaneada pelo Conselho Nacional de Justiça. Desde a edição da Resolução nº 125 de 2010, em que foi instituída a Política Judiciária Nacional de tratamento adequado dos conflitos de interesses no âmbito do Poder Judiciário, até a Resolução nº 358 de 2020, em que foi regulamentada a criação de soluções tecnológicas para a resolução de conflitos pelo Poder Judiciário por meio da conciliação e mediação, percebe-se um aumento na estrutura de favorecimento de conciliação nos órgãos judiciais. Tal mobilização trouxe resultados satisfatórios. Segundo o Justiça em Números de 2023, 18% (dezoito por cento) dos processos de conhecimento foram solucionados pela via da conciliação, em comparação com as sentenças terminativas (CONSELHO NACIONAL DE JUSTIÇA, 2020).

A melhoria da gestão de conflitos é uma medida que as Cortes devem sempre buscar para solucionar essa expressiva quantidade de processos, notadamente diante da velocidade com que são encerrados. No entanto, é preciso observar que essas medidas se manifestam principalmente em relação aos feitos novos o que, portanto, demonstra que também deve-se voltar a atenção para a enorme quantidade de processos pendentes de resolução.

Além da questão do acervo processual, outro grande problema enfrentado pelo Judiciário Brasileiro tem sido a superlotação dos presídios. Em último levantamento realizado pelo Departamento Penitenciário Nacional, em dezembro de 2023 o Brasil contava com 644.316 (seiscentos e quarenta e quatro mil trezentos e dezesseis) pessoas presas, destas mais de 175.000 (cento e setenta e cinto mil) como presos provisórios (Secretaria Nacional de Políticas Penais, 2024).

Tais dados relevantes, sejam relacionados à enorme quantidade de processos pendentes de resolução, seja quanto ao número de presos de nosso país, exigiu que a administração da justiça nacional implementasse medidas no intuito de solucioná-los, oportunidade em que a inteligência artificial passou a ser cogitada como ferramenta de auxílio, notadamente diante dos recentes avanços implementados em diversos aspectos do cotidiano.

Como será demonstrado, uma das vantagens da automação é a possibilidade de ganho de tempo na execução de atividades de rotina. Assim, a máquina consegue desempenhar atividades repetitivas de forma muito mais rápida que o ser humano, sendo tal característica identificada desde a época da revolução industrial. O

diferencial do tempo presente é que agora os *softwares* são capazes de desempenhar atividades complexas que, a princípio, somente poderiam ser desenvolvidas pelo homem, trazendo com isso a necessidade de se refletir sobre uma série de impactos causados por tal avanço tecnológico.

Neste sentido, a inteligência artificial tem ganhado notável destaque por sua utilização em programas que envolvem a tomada de decisão. Por sua utilização em órgãos de governo e grandes empresas, os *softwares* se tornam cada vez mais responsáveis por decisões que impactam sobremaneira a vida dos cidadãos.

E é dentro deste espectro de impactos que será analisado na presente tese o fenômeno dos vieses discriminatórios inseridos nos *softwares* de inteligência artificial. Tal fato tem sido percebido em diversos programas, com utilização inclusive dentro do Poder Judiciário, agravando injustamente a situação jurídica de grupos minoritários que historicamente já sofrem com a restrição e supressão de direitos.

Assim, a necessidade de estudar o tema se torna premente, considerando a expansão dos sistemas inteligentes e seu uso por órgãos estatais. Urge, portanto, uma clara e ampla discussão acerca do seu uso e de eventual regulamentação, fator também de grande debate ao redor do mundo.

A relevância social desta pesquisa justifica-se pela larga aplicação destes programas computacionais em todo o mundo, sendo necessária, portanto, uma ampla discussão acerca de sua utilização e regulamentação, ponto em que este estudo pretende contribuir. É premente que as causas do fenômeno sejam identificadas e que soluções sejam apresentadas para que grupos vulneráveis não sejam ainda mais penalizados dentro de um sistema penal que historicamente já os submete a tratamento desigual. Dessa forma, academicamente falando, a ausência de mais pesquisas direcionadas à proteção de direitos humanos dentro deste espectro de conhecimento, em que se analisa a sua interseção com a inteligência artificial, demonstra a necessidade de se discutir este tema tão relevante para as classes sociais prejudicadas.

Entende-se que figura como dever do Estado a promoção de ações que possam viabilizar a proteção de tais grupos vulnerabilizados. É sua missão impedir que a automação que se apresenta de forma inexorável junto ao Poder Judiciário não se transforme em mais uma ferramenta de opressão para as minorias, ponto no qual buscamos, também, contribuir, oferecendo uma análise objetiva sobre toda a

problemática, medida que se faz necessária notadamente diante da escassez de dados sobre o referido tema.

O presente estudo, pois, se justifica, como forma de garantir direito ao justo tratamento a grupos marginalizados que, hoje, enfrentam uma nova ameaça à igualdade de oportunidades dentro do próprio Poder Judiciário, local que deveria ser o responsável pela sua proteção e não origem de novas violações.

Em uma abordagem metodológica, quanto à problemática vinculada ao estudo proposto, temos a existência de comportamentos discriminatórios surgidos em virtude de utilização de *softwares* de inteligência artificial; a violação ao direito de igualdade de grupos minoritários decorrente da automação de serviços públicos; e a ausência de medidas efetivas para combater os vieses discriminatórios de tais programas.

Apresentada a problemática, define-se como questão problema da tese a seguinte: Como impedir que a automação da atividade judicante por *software* de inteligência artificial viole direitos humanos de grupos vulneráveis, impondo-lhes regime injustificadamente mais gravoso na seara penal?

O problema exposto é formado pela junção de dois temas complexos: automação decorrente de inteligência artificial e violação de direitos humanos na esfera penal. Neste sentido, é de se ter em mente que a solução a ser encontrada poderá ser igualmente complexa.

Friedman e Nissenbaum (1996) identificaram que o viés discriminatório pode ser preexistente (quando já existe em instituições sociais, práticas e atitudes de um povo), por motivos técnicos (o algoritmo do programa já se encontra contaminado) ou emergente (quando decorre da utilização do usuário). A solução, portanto, passa pela análise e combate a cada uma destas causas.

Assim, a hipótese sugerida para a resolução do problema seria a seguinte: a criação de algoritmos pautados na defesa dos Direitos Humanos conduzirá diretamente ao fim de atos discriminatórios em decorrência da utilização de *softwares* de inteligência artificial no Poder Judiciário. E que, somente através de uma regulamentação efetiva é que poderá ser imposta a observância desta premissa.

A viabilidade da hipótese decorre do fato de que a construção dos programas de inteligência artificial se fundamenta em escolhas. Nesta ocasião, o programador decide quais pontos serão levados em consideração pelo *software* e quais não serão sopesados. Esse processo seletivo pode revelar opiniões e ideologias do criador que, naturalmente, passam para a máquina. Dessa forma, a atuação nessa fase de

construção é essencial para que as informações que serão avaliadas pelos algoritmos não resultem em uma análise eivada de vícios discriminatórios. Assim, se discorrerá sobre a viabilidade de se implementar mudanças desde a concepção dos *softwares* que terão que ser implementadas por força de uma regulamentação estatal.

Destacam, ainda, Boeing e Rosa (2020), que a transparência passa a ser fator crucial para que a sociedade, de forma geral, seja cientificada acerca dos aspectos da realidade que serão levados em conta pelo *software*, possibilitando, assim, uma fiscalização efetiva e a própria análise de viabilidade dos fatores levados em consideração na sua programação. É, portanto, o cotejo de tais fatores que irá possibilitar a verificação da hipótese sugerida.

A tese gravita em torno de três matrizes teóricas: os direitos humanos, o direito penal e a ciência da computação, com ênfase nos conceitos relacionados à inteligência artificial. A discussão que aqui se propõe tem sua gênese na violação dos direitos humanos de grupos minoritários, razão pela qual conceituar tais direitos e construir um raciocínio acerca da igualdade material e da vulnerabilidade de determinados segmentos sociais se torna necessária.

O objetivo geral da tese é analisar como se implementam ferramentas para proteção de grupos vulneráveis submetidos à avaliação por *softwares* de inteligência artificial na Justiça Criminal. Tem, como objetivos específicos: a) examinar o processo de criação de ferramentas de inteligência artificial utilizadas como auxiliares na atividade judicante estatal e justiça preditiva, no intuito de compreender seu funcionamento; b) analisar a pertinência ética acerca da utilização de softwares inteligentes em processos decisórios por entes públicos e privados; c) analisar textos legislativos que se propõem a regulamentar a matéria e se a resposta a tais violações pode ser obtida através da regulamentação; d) investigar o processo de vulnerabilização de minorias sociais, suas causas e principais fatores de marginalização, mais especificamente da população negra, que foi o grupo atingido pelo mau funcionamento do *software* a ser analisado; e) avaliar e propor políticas públicas para evitar a vitimização de grupos vulneráveis submetidos à apreciação automatizada pelo Poder Judiciário e demais entes públicos.

Além disso, o estudo tem como objetivo analisar as consequências de utilização de ferramentas de inteligência artificial no auxílio da atividade judicante, observando o fenômeno sob a ótica da proteção de grupos de vulneráveis, de forma mais específica no âmbito do direito penal.

Para o sucesso da pesquisa proposta, utiliza-se como método de abordagem o hipotético-dedutivo. Considerando que este exige a constatação do problema, a apresentação de um resultado e a respectiva tentativa de falseamento, tem-se que se afigura como o mais adequado para o estudo, eis que a sua complexidade, com a ligação entre a computação e os direitos humanos, demanda a análise e reanálise das conclusões obtidas no intuito de se obter a solução mais eficaz.

Neste sentido, o tema abordado no estudo decorre de uma multiplicidade de fatores que podem ser divididos em três categorias: o fator social, o fator computacional e o fator humano (operador). Ao analisar, por exemplo, a contribuição do fator humano no estabelecimento de vieses nos *softwares*, é necessário avaliar, através do método hipotético-dedutivo de falseamento das hipóteses, qual a solução viável para mitigação do problema e, assim, sucessivamente quanto às outras causas.

Quanto ao procedimento, o método histórico será relevante, considerando que se faz necessário o esclarecimento dos fatos à luz da ocorrência de acontecimentos passados. A compreensão histórica do fenômeno de vulnerabilização de determinados grupos é importante para se entender o contexto em que se insere tal coletividade e, por que razão *softwares* de inteligência artificial herdam determinados comportamentos tendenciosos.

O Estudo de Caso também foi utilizado para fins de análise dos dados relativos à ocorrência de vieses discriminatórios em *softwares* já em funcionamento em diversos órgãos do Poder Judiciário no mundo. Tendo em vista que, neste método, a investigação com profundidade de casos específicos pode ensejar conclusões que atingem diversas outras situações análogas, a representatividade dos casos analisados serve suficientemente de paradigma para análise do problema apresentado.

Por fim, o método estatístico também se mostra presente, na medida em que os dados relacionados ao caso estudado reforçam as conclusões que serão obtidas, com representação numérica dos resultados. Sendo passíveis de apuração, os números que demonstram a ocorrência de tendenciamento das decisões judiciais prolatadas com auxílio do *software* de inteligência artificial figuram como essenciais para comprovação e eventual proposta de solução do problema.

A técnica de pesquisa utilizada será primordialmente de documentação indireta, com a consulta a documentos oficiais e bibliografia especializada. Além disso, também serão alvo de consulta legislações mundiais acerca do tema, no intuito de se analisar

as vantagens com a utilização da regulamentação para se buscar a solução do problema.

A pesquisa pode ser classificada como explicativa, uma vez que busca identificar fatores ou dados que justifiquem a ocorrência de determinado fenômeno. A problemática dos vieses na prolação de decisões através de *softwares* de inteligência artificial necessita de aprofundamento que identifique as causas de sua ocorrência, o que é um dos objetivos específicos do presente estudo.

No que se refere às fontes de pesquisa, podemos indicar como fontes primárias os autores das duas áreas do conhecimento (direitos humanos e computação), como Richard Susskind (1989; 2019), Alan Turing (1950), Kevin D. Ashley (2017), Hannah Arendt (1989), Norberto Bobbio (1995; 2004), Immanuel Kant (2011), Fábio Konder Comparato (2015), Andrea Simoncini (2019; 2020), Flávia Piovesan (2019), André de Carvalho Ramos (2019), entre outros.

O estudo se divide em cinco capítulos. No primeiro capítulo será analisado o processo de criação e funcionamento de *softwares* de inteligência artificial utilizados no Poder Judiciário, com o intuito de se entender a origem do problema causado por vieses inseridos no algoritmo do sistema. Faz-se importante tal investigação, pois a compreensão da gênese da controvérsia pode ser fator crucial para que se identifiquem as soluções a serem apresentadas ao final do texto.

O segundo capítulo discorre sobre questões éticas relacionadas ao uso de inteligência artificial no processo decisório, bem como sobre a necessidade de se analisar os problemas também sobre um prisma teórico, buscando uma abordagem holística sobre o tema.

No terceiro capítulo serão observadas as iniciativas legislativas sobre o objeto de estudo, bem como quais medidas foram tomadas no intuito de se evitar a ocorrência do fenômeno ora em questão. Tal medida visa à análise sobre a viabilidade e validade da regulamentação como ferramenta de inibição para o surgimento dos vieses discriminatórios em máquinas inteligentes.

O quarto capítulo traz uma abordagem teórica sobre os direitos humanos e como a proteção dos grupos vulneráveis encontra guarida no direito à cidadania. Será exposto o histórico de vitimização de grupos vulneráveis, suas causas e consequências, bem como observado, em dias atuais, como o processo de marginalização ainda ocorre, inobstante a aparente preocupação geral na proteção destes núcleos sociais. O processo de vulnerabilização, apesar de sempre ter

ocorrido, agora se apresenta em uma nova modalidade, com a inserção das máquinas na equação. Assim, a discussão em torno desta violação de direitos ganha novos aspectos que merecem maior perquirição sobre a sua ocorrência.

Por fim, o capítulo final analisará teorias da regulamentação e a propositura da tese quando as medidas e condutas que podem ser tomadas para se evitar a ocorrência de novos casos de vitimização especial de minorias. Como o objetivo geral da pesquisa busca a análise na utilização dos *softwares* de inteligência artificial, perscrutar possíveis medidas de saneamento comungam com o avanço do tema para que sejam evitados novos casos de vitimização.

Busca-se, então, estudar o fenômeno de surgimento dos vieses em *softwares* de inteligência artificial e propor soluções para que grupos vulneráveis não sejam especialmente vitimados por estes programas que estão cada vez mais presentes no cotidiano de todos, inclusive no Poder Judiciário.

2. INTELIGÊNCIA ARTIFICIAL

A inteligência artificial não pode ser classificada como uma área recente da computação. Alan Turing, famoso matemático britânico que teve relevante atuação durante a II Guerra Mundial, foi um dos pioneiros no estudo de sua aplicação. Ele, conhecido hoje como pai da computação e da inteligência artificial, escreveu em 1950 um artigo em que já discorria sobre o ato de "pensar" das máquinas, tendo iniciado seu texto da seguinte forma: "*I propose to consider the question, 'Can machines think*?[1]" (TURING, 1950, p. 433). Turing propõe em seu *paper* um jogo denominado *imitation game*, que posteriormente ficou conhecido como *Turing Test*. Neste teste, analisa-se a possibilidade do computador vencer o humano em um jogo de perguntas e respostas. Boeing e Rosa (2020, p. 20) descrevem sua dinâmica:

> Nele, máquinas seriam avaliadas de acordo com sua capacidade de mimetizar seres humanos, de forma que o computador passaria no teste se um interrogador humano, após fazer perguntas por escrito, não conseguisse identificar estar-se comunicando com outro ser humano ou com um robô.

Constata-se que o dia em que a máquina atingiu o patamar exigido pelo *Turing Test* já chegou e até o ultrapassou: o *software* da Internacional Business Machines Corporation – IBM *Watson* foi capaz de derrotar, em 2011, dois ex-campeões do *quiz game Jeopardy*! transmitido pela televisão americana (GABBAT, 2011).

Além de Turing, também em 1950 outro famoso escritor estabeleceu regras básicas para computadores autônomos que seriam utilizadas até os dias de hoje. Isaac Asimov criou em seu livro "Eu, Robô", as três leis da robótica, assim descritas:

> Um robô não pode magoar um ser humano ou, por inação, permitir que tal aconteça; 2- Um robô tem de obedecer às ordens dos seres humanos, exceto quando tais ordens entrarem em conflito com a primeira lei; 3- Um robô tem de proteger a sua própria existência desde que tal proteção não entre em conflito com a primeira ou com a segunda lei. (ASIMOV, 2008, p. 256).

Importante mencionar que, apesar de se tratar de uma obra ficcional, as leis da robótica de Asimov são largamente utilizadas como padrão de referência para desenvolvimento de *softwares* de inteligência artificial, inclusive pelo Parlamento Europeu, fato que posteriormente será mais bem explanado nesta pesquisa.

---

1 "Eu proponho considerar a questão: as máquinas podem pensar?" (tradução nossa)

A concepção de que a máquina pode “pensar” é equivocada. Na verdade, o computador é capaz de realizar processos previamente definidos de forma rápida e eficaz, bastante semelhante ao pensamento humano. Tal processo é feito através da construção de algoritmo que, explicam Boeing e Rosa (2020, p. 19), “nada mais é que um conjunto finito e preciso de passos para resolver um problema ou responder uma questão”. Assim, a inteligência artificial corresponde a execução, por parte da máquina, de processos pré-definidos pelo programador no modelo computacional.

Andrea Simoncini referiu-se aos algoritmos salientando seu aspecto reducionista:

> Innanzitutto, esse condividono uma epistemologia riduzionista che potremmo sinterizzare nel modo seguente: tutti i problemi complessi (compresa l’inteligenza umana) possono essere ricondotti ad una serie ordinata e finita di problemi più semplici (algoritmo)[2]. (SIMONCINI, 2019, p. 67)

Dessa forma, percebe-se que os algoritmos devem ser vistos como uma redução de problemas complexos, caminhos que devem ser seguidos pelo computador para que possa oferecer a resposta buscada pelo usuário.

Quanto ao conceito de inteligência artificial, não há consenso dos estudiosos. Considerando a enorme gama de processos e capacidades, a definição do termo não encontra uma padronização no meio acadêmico. Destaca-se, para fins da presente tese, uma descrição baseada na comparação com o ser humano trazida por Bensoussan e Champion (2013, p. 29) ao designá-la como a “*capacité d’une unité fonctionnelle à exécuter des fonctions généralement associées à l’intelligence humaine, telles que le raisonnement et l’apprentissage*[3]”.

Considerando esses conceitos relacionados à computação e inteligência artificial, torna-se relevante discutir, mesmo que brevemente, o histórico evolutivo de seu desenvolvimento para fins de colocação e ciência quanto ao atual estágio de desenvolvimento da tecnologia.

---

[2] Em primeiro lugar, partilham uma epistemologia reducionista que poderíamos resumir da seguinte forma: todos os problemas complexos (incluída a inteligência humana) podem ser reduzidos a uma série ordenada e finita de problemas mais simples (algoritmo). (tradução nossa)

3 “Capacidade de uma unidade funcional em executar funções geralmente associadas à inteligência humana, como o raciocínio e a aprendizagem”. (tradução nossa)

2.1. Desenvolvimento Histórico

O sonho humano de criação de máquinas inteligentes permeia o imaginário popular e a literatura há centenas anos. Desde o mito grego de Pigmaleão, em que o rei cipriota conseguiu transformar uma estátua por ele esculpida em uma mulher, ao famoso conto do Frankenstein, da britânica Mary Shelley, todos trazem a ideia de despertar um corpo, dotando-lhe de inteligência (BULFINCH, 2002; CHARPA, 2012).

Como mencionado anteriormente, os estudos teóricos acerca da inteligência artificial tiveram sua origem por volta de 1950, com o britânico Alan Turning. Em data próxima, 1952, Arthur Samuel, da IBM, escreveu um programa para o jogo de damas que, eventualmente, superou o seu próprio criador na modalidade (RUSSEL; NORVIG, 2013).

É importante destacar que desde a década de 1950 já se vislumbrou o enorme potencial da inteligência artificial. Em 1957, Herbert Simon atestou:

> Não é meu objetivo surpreendê-los ou chocá-los, mas o modo mais simples de resumir tudo isso é dizer que agora existem no mundo máquinas que pensam, aprendem e criam. Além disso, sua capacidade de realizar essas atividades está crescendo rapidamente até o ponto – em um futuro visível, no qual a variedade de problemas com que elas poderão lidar será correspondente à variedade de problemas com os quais lida a mente humana. (RUSSEL; NORVIG, 2013, p.45).

Na década seguinte, em 1960, foi criado o *General Problem Solver of Newell and Simon*, que buscava, através do computador, aplicar métodos humanos de solução de problemas. Diante de sua falta de eficiência e da carga de processamento requerida para utilizá-la na época, o projeto foi abandonado (WARWICK, 2012).

Ainda segundo Warwick (2012), enquanto a década de 70 foi desprovida de maiores avanços, os anos 80 foram de grandes saltos na área, com o desenvolvimento dos chamados *expert systems*, conceito mais bem trabalhado em momento posterior deste estudo.

Desde o final dos anos 90 a tecnologia em questão atingiu notável patamar. Em 1997, o *Deep Blue*, máquina movida à inteligência artificial, foi capaz de derrotar em um jogo de xadrez Garry Kaparov, campeão mundial da modalidade (STRASSER, 2017). Esse tipo de conquista serviu para impulsionar os investimentos na área, sendo, até hoje, ponto de principal foco das empresas de *software*.

Sobre as perspectivas do futuro, encerra Warwick (2012), o que se pode esperar é uma tentativa de “corporizar” os sistemas de inteligência artificial. Filosoficamente, muito se compara a inteligência artificial ao conceito de cérebro. Assim, explica o autor, um dos interesses consideráveis deste ramo computacional é dotar sistemas de IA com seus próprios corpos, para que eles possam interagir diretamente com o mundo e não somente através de estímulos virtuais.

Importante destacar a previsão de Manyika *et al*. (2017), baseada em relatório do *McKinsey Global Institute*, de que quantidade próxima da metade de todo trabalho que é feito atualmente será automatizada até o ano de 2055, salvo eventual avanço ou esforço no sentido de encontrar uma forma para convívio harmônico entre máquina e homem.

Todos esses dados acerca do impacto da tecnologia no cotidiano têm feito os governos se mobilizar e incentivar a busca por novas descobertas na área, no intuito de acompanhar toda essa evolução. No Brasil, por exemplo, foi publicada a Lei nº. 10.973/2004, a Lei da Inovação, que traz em seu corpo diversas medidas de incentivo à inovação e pesquisa científicas voltadas, principalmente, para a capacitação tecnológica, autonomia tecnológica e desenvolvimento do sistema produtivo nacional (art. 1º). Em 2016 foi editado o Decreto nº 10.534, que instituiu a Política Nacional de Inovação, tendo estabelecido diretrizes mais claras sobre a implementação de políticas voltadas para a inovação no nosso país, seus objetivos e instrumentos.

Percebe-se, portanto, que a evolução da computação está proporcionando grandes avanços na área de inteligência artificial que terão aplicabilidade em diversas áreas do dia-a-dia, razão pela qual os governos têm se mobilizado para fomentar ainda mais esse crescimento, no intuito de obter a maior quantidade possível de benefícios na utilização da tecnologia.

2.2. Conceitos Relevantes

A inteligência artificial é um termo genérico, que abriga diversos campos da ciência e tecnologia computacional e, normalmente, envolve a criação de algoritmos complexos que viabilizam resultados ou respostas de forma predeterminada (SOURDIN, 2018).

Neste sentido, é preciso perceber que diversos processos são baseados no uso de inteligência artificial para obtenção dos resultados propostos. Trata-se, pois, de uma ferramenta avançada para realização de atividades complexas.

Tais avanços, contudo, só puderam ser alcançados através da evolução do próprio sistema de computação. É importante mencionar que, segundo a Lei de Moore, a capacidade de processamento é ampliada de forma exponencial, havendo redução drástica de seus gastos a cada 12 ou 18 meses (COELHO, [2019]). Segundo a Enciclopédia Treccani:

> Legge empirica che descrive lo sviluppo della microelettronica, a partire dall'inizio degli anni Settanta, con una progressione sostanzialmente esponenziale, perciò straordinaria; la legge fu enunciata per la prima volta nel 1965 da Gordon Moore, uno dei fondatori di INTEL e dei pionieri della microelettronica, che la ribadì pubblicamente nel 1974. Essa afferma che la complessità dei microcircuiti (per es., misurata dal numero di transistori per chip o per area unitaria) raddoppia periodicamente, con un periodo originalmente previsto in 12 mesi, allungato a 2 anni verso la fine degli anni Settanta e dall'inizio degli anni Ottanta assestatosi sui 18 mesi . ("Legge di Moore". In: ENCICLOPEDIA Treccani. Disponível em: <https://www.treccani.it/enciclopedia/legge-di-moore_%28Enciclopedia-della-Scienza-e-della-Tecnica%29/>. Acesso em: 20 jul. 2024.)

Para ilustrar isso, basta observar que, atualmente, um *smartfone* possui muito mais potência de cálculo do que possuía o computador utilizado pela NASA para guiar a Apollo 11 quando da ida do homem à Lua (Floridi, 2022).

Tendo em mente esse grande aumento na capacidade de processamento no decorrer dos anos, bem como a redução de seus custos, foi possível o aprimoramento de ferramentas ligadas à inteligência artificial e ao desenvolvimento de novas estruturas de tecnologia.

Nesse sentido, Oren Etizioni elencou três requisitos para o bom funcionamento de um sistema baseado em inteligência artificial:

> First, an A.I. system must be subject to the full gamut of laws that apply to its human operator. This rule would cover private, corporate and government systems […] My second rule is that an A.I. system must clearly disclose that it is not human [...] My third rule is that an A.I. system cannot retain or disclose confidential information without explicit approval from the source of that information[4]. (ETZIONI, 2017, p. 1).

4 "Primeiro, um sistema de inteligência artificial deve estar sujeito a toda a gama de leis que se aplicam ao seu operador humano. Esta regra abrangeria sistemas privados, corporativos e governamentais [...] Minha segunda regra é que um sistema de IA deve deixar claro não é humano [...] Minha terceira regra é que um sistema de IA não pode reter ou divulgar informações confidenciais sem a aprovação explícita da fonte dessas informações." (tradução nossa)

Segundo Nilsson (2010), existem 06 (seis) capacidades que decorrem do gênero inteligência artificial. São elas: aprendizado de máquina (capacidade de detectar padrões e aplicá-los em processos futuros); processamento de linguagem natural (possibilitar a compreensão bem-sucedida de textos); representação de conhecimento (armazenar o que sabe ou ouve); raciocínio automático (usar as informações armazenadas para solução de problemas); visão computacional (percepção espacial de objetos) e robótica (manipulação de objetos e a própria movimentação da máquina).

Para os fins do presente estudo, os *softwares* analisados utilizam-se de forma majoritária do aprendizado de máquina (*machine learning*) e o processamento de linguagem natural (*natural language processing*), eis que as demais lidam com comunicação, percepção ou a própria movimentação da máquina, o que, *a priori*, foge ao objeto em discussão.

O termo Machine learning pode ser definido como "*methodology and set of techniques that finds novel patterns and knowledge in data, and generates models (eg profiles) that can be used for effective predictions about the data*[5]" (OTTERLO, 2013, p. 41-64).

É possível constatar que o processo de aprendizado da máquina fundamenta-se na compreensão, pelo *software*, de comportamentos anteriores para que possa desenvolver atividades futuras de forma autônoma, aplicando estes padrões.

Por outro lado, o *natural language processing* (NLP) é uma parte da inteligência artificial que "estuda o desenvolvimento de programas de computador que analisam, reconhecem e/ou geram textos em linguagens humanas, ou linguagens naturais" (LOPES; VIEIRA; 2010, p. 184). Possibilita, então, que a máquina possa reconhecer a escrita e utilizar esta informação na realização de suas atividades.

É importante perceber toda a complexidade que envolve este campo científico, notadamente diante da dificuldade relacionada à compreensão da linguagem pela máquina. Luger (2005, p. 591) sintetiza:

> Communicating with natural language, whether as text or as speech acts, depends heavily on our knowledge and expectations within the domain of discourse. Understanding language is not merely the transmission of words: it also requires inferences about the speaker's goals, knowledge, and assumptions, as well as about the context of the interaction. Implementing a

5 Metodologia e conjunto de técnicas que localizam novos padrões e conhecimento nos dados e geram modelos (eg perfis) que podem ser usados para predição sobre os dados. (tradução nossa)

> natural language understanding program requires that we represent knowledge and expectations of the domain and reason effectively about them[6].

Esta capacidade de processamento somente pode ser obtida após uma revolução na forma de captura de dados. No início, os dados coletados se apresentavam de forma "estruturada", ou seja, a leitura de caracteres – letras e números. No entanto, o desenvolvimento da computação possibilitou que esta coleta de dados fosse realizada em relação a itens "não-estruturados", como imagens, áudios, vídeos. Tal fenômeno ficou conhecido como *digitalization* e proporcionou enorme avanço no processamento de linguagem natural (COELHO, [2019]).

Observa-se que *machine learning* e *natural language processing* são dois processos extremamente complexos, mas que juntos possuem enorme potencial, inclusive para o Direito. Enquanto uma das tecnologias permite que o computador entenda informação que foi posta à disposição a outra é capaz de processá-la e agir de acordo com a sua programação, oferecendo uma resposta ao questionamento com base em experiências pretéritas, tudo de acordo com o predefinido em seus algoritmos.

Uma vez observado o potencial de aplicação profissional, foram criados sistemas voltados para determinadas áreas do conhecimento, identificados como *expert systems*. Susskind (1986, p. 172) assim os definiu:

> Expert systems are computer programs that have been constructed (with assistance of a human experts) in such a way that they capable of functioning at the standard of (and sometimes even at a higher standard than) experts in given fields. They are used as high-level intellectual aids to they users[7].

Com tais programas, surgiu para as empresas uma demanda de criação de diversos *softwares* relacionados às mais variadas áreas da sociedade, aumentando a abrangência e alcance da inteligência artificial na vida de todos.

---

6 "A comunicação com a linguagem natural, seja textual ou falada, depende muito de nosso conhecimento e expectativas dentro do domínio do discurso. Compreender a linguagem não é apenas a transmissão de palavras: também requer inferências sobre os objetivos, conhecimentos e suposições do interlocutor, bem como sobre o contexto da interação. A implementação de um programa que compreenda a linguagem natural exige que representemos o conhecimento e as expectativas do domínio e raciocine efetivamente sobre eles." (tradução nossa)

7 "*Expert systems* são programas de computador que foram construídos (com a ajuda de especialistas humanos) de maneira que eles possam funcionar da mesma forma que (e às vezes até em um padrão mais alto do que) especialistas em determinados campos. Eles são usados como auxiliares intelectuais de alto nível para seus usuários. " (tradução nossa)

Tais sistemas são então utilizados em diversas áreas e no direito suas repercussões afetam sobremaneira a vida de todos, eis que lidam com diversos aspectos cotidianos e podem ser responsáveis por sérios impactos no patrimônio jurídico de qualquer cidadão a eles submetidos.

2.3. Inteligência Artificial e Direito

O uso de inteligência artificial relacionada ao Direito não foi concebido atualmente. Ainda em 1970, surgiram os primeiros trabalhos sobre o tema, sendo um dos mais emblemáticos a produção de Buchanan e Headrick (1970, p. 40), que expressamente narram o início deste diálogo entre as Ciências:

> Research in artificial intelligence, a branch of computer science, has illuminated our capacity to use computers to model human thought processes. This research suggests that computer science may assist lawyers in both the study and performance of their reasoning processes. In this Article we will argue that the time has come for serious interdisciplinary work between lawyers and computer scientists to explore the computer's potential in law.[8]

Percebe-se, então, não se tratar de matéria de estudo nova. Não há como negar, contudo, que os recentes avanços na ciência de dados possibilitaram um enorme saldo evolutivo no desenvolvimento de tais programas. Dada a expressividade do crescimento, foi criado o conceito de Inteligência Artificial Legal (*Artificial Legal Intelligence – ALI*), que pode ser definida como um sistema capaz de dar orientações jurídicas (SOURDIN, 2018).

Sourdin aponta três fatores que demonstram a influência da Inteligência Artificial Legal na atualidade. Primeiro, a tecnologia auxilia com fornecimento de informação e aconselhamento jurídico aos leigos. Segundo, exercendo atividades antes somente realizada por seres humanos. E terceiro, modificando a atuação judicial, auxiliando magistrados na prolação de decisões.

Tal fator se dá pela superioridade dos computadores em relação aos humanos quanto a sua aptidão para armazenamento e uso de dados armazenados para solução de problemas bem como pela enorme capacidade de cálculos (Tzafestas, 2016). Essa

8 "Pesquisas em inteligência artificial, um ramo da ciência da computação, iluminou nossa capacidade de usar computadores para modelar processos de pensamento humano. Esta pesquisa sugere que a ciência da computação pode auxiliar advogados no estudo e na prática de seus processos racionais. Neste artigo, argumentaremos que chegou a hora de um trabalho interdisciplinar sério entre advogados e cientistas da computação para explorar o potencial legal do computador no direito." (tradução nossa)

capacidade repercute num melhor aproveitamento quanto ao uso das informações armazenadas pelo sistema jurídico.

É de se enxergar diante deste fato a necessidade de adequação inclusive dos próprios cursos de Direito, exigindo dos novos bacharéis uma adaptação à nova realidade. A prática jurídica vem mudando e, em breve, atividades cotidianas serão desempenhadas apenas pela máquina.

A esse respeito, Henderson (2013, p. 501) sintetiza:

> Because of the emphasis on process and technology now taking hold within the legal industry, the practical technical skills and domain knowledge that might have held us in good stead in 1980 or 1990 may be inadequate for a large proportion of law students graduating in the year 2015 […] Students […] They are unprepared to learn that law is becoming less about jury trials and courtroom advocacy and more about process engineering, predictive coding, and the collaborative and technical skills those processes entail.[9]

Partindo dessa premissa, a chegada da inteligência artificial demanda que todos, notadamente os operadores do direito, possam refletir sobre os impactos desta tecnologia, o uso ético e responsável que pode ser feito dela e como evitar que as falhas – mais adiante discutidas – possam ocorrer, prejudicando assim o direito terceiros.

Richard Susskind aborda o tema sob um viés de desconstrução dos problemas jurídicos e sugere que os problemas podem ser decompostos de forma a possibilitar que sejam resolvidos de maneira mais adequada, permitindo que homem e a máquina atuem em atividades específicas, garantindo maior eficiência:

> What I am not saying is that for any piece of legal work – say, a deal or dispute – the question that arises is as follows: in which of my six boxes does that legal matter sit? I am saying something subtler than this, namely, that for any deal or dispute, no matter how small or large, it is possible to break it down, to 'decompose' the work into a set of constituent tasks. And it is in respect of each of these tasks, not the job as a hole, that one can ask: what is the most efficient way of undertaking this work, and to which of the six boxes should the tasks be allocated.[10] (SUSSKIND, p. 32, 2017).

---

9 "Devido à ênfase no processo e na tecnologia atualmente em vigor no setor jurídico, as habilidades práticas técnicas e o conhecimento na área que podem ter nos mantido em boa posição em 1980 ou 1990 podem ser inadequados para uma grande proporção de estudantes de direito que se formam no ano de 2015 […] Alunos […] eles não estão preparados para aprender que a lei está se tornando menos sobre julgamentos em júris e advocacia de tribunais e mais sobre engenharia de processos, codificação preditiva e as habilidades técnicas e de colaboração que esses processos envolvem." (tradução nossa)

10 "O que não estou dizendo é que para qualquer trabalho jurídico - digamos, um acordo ou disputa - a questão que se coloca é a seguinte: em qual das minhas seis caixas está essa questão jurídica? Estou dizendo algo mais sutil do que isso, ou seja, que para qualquer acordo ou disputa, não importa quão grande ou pequena, é possível dividi-la, "decompor" o trabalho em um conjunto de tarefas

Para os fins do presente estudo é necessária a análise dos *softwares* que são criados especificamente para o Direito e sua aplicação. A partir destes conceitos teóricos se torna possível adentrar de forma mais capacitada no problema gerado pelos vieses da inteligência artificial e como podem ser evitados.

Acerca da forma de processamento jurídico das ferramentas, a relação entre inteligência artificial e o direito pode ser classificada em quatro categorias, de acordo com Rissland (1990): a) *reasoning with rules*; b) *reasoning with cases*; c) *mixed paradigma reasoning;* d) *deep models of legal knowledge*.

Reasoning with rules consiste em um processo que possibilita o uso de ferramentas de adequação entre a situação a ele submetida e o sistema de normas previamente definido. Rissland (1990, p. 1965) resume: "*If certain conditions are known to hold, then take the stated action or draw the stated conclusion*."[11] Consiste em uma categoria de correlação direta, um sistema de subsunção entre fato e a norma.

O *reasoning with cases*, em correlação com a *common law*, utiliza-se de precedentes para resolução da questão submetida. Trata-se também de uma ferramenta correlacional, mas neste ponto não se busca uma ligação com um dispositivo legal, mas sim com um precedente que atenda ao caso. Neste ponto, é possível notar um grau de complexidade mais elevado, considerando que a adequação a ser alcançada exige uma análise fática para aplicação correta do paradigma.

No *reasoning with rules and cases* há a aplicação conjunta das duas categorias de processamento anteriores, que tem relevância em sistemas jurídicos mistos, com forte presença da lei mas também com aplicação vinculante de precedentes, como é o caso do Brasil. Nesse caso, há dificuldades de aplicação, notadamente diante do conflito de superposição entre as duas fontes a serem utilizadas na decisão. Rissland (1990) sugere que sejam utilizadas análises distintas e isoladas, somente posteriormente sendo procedida a combinação dos dados (*blackboard systems)* para se estabelecer o resultado.

---

constituintes. E é em relação a cada uma dessas tarefas, e não ao trabalho como um todo, que se pode perguntar: qual a forma mais eficiente de realizar esse trabalho e a qual das seis caixas devem ser atribuídas as tarefas." (tradução nossa)

11 "Se certas condições ocorrerem, então execute determinada ação ou tome determinada conclusão." (tradução nossa)

Em relação aos *deep models of legal knowledge*, há a apresentação do nível mais complexo de tecnologia que relaciona o direito e a inteligência artificial. Através do seu uso, o sistema se torna capaz de racionalizar dados que são encontrados em fontes muito mais amplas que repositórios de leis e decisões judiciais – o *Big Data* – que possibilita o uso e processamento de uma quantidade enorme de informações disponíveis na rede.

Nesse contexto, a importância de se demonstrar essa classificação reside no fato de que este último modelo é utilizado em ferramentas de análise de risco (*risk assessment)* – exatamente o tipo de *software* em que se constataram sérios indícios de vieses discriminatórios, aqui estudados. Estes modelos captam dados relativos a diversos aspectos da vida do agente, desde questões sociais como local de residência e formação acadêmica, até suas redes sociais.

Foi também através do uso deste tipo de modelo que surgiu o conceito de "justiça preditiva", que pode ser assim definida:

> A Justiça Preditiva constitui, pois, no processamento matemático e a combinação dos grandes dados legais disponíveis que devem restaurar as probabilidades finas quanto às chances de êxito de um procedimento, estimativas de danos ou compensações, mas também aos argumentos decisivos para se tomar uma decisão. Requer um sólido conhecimento jurídico para determinar os dados relevantes, definir as qualificações ou identificar os tipos de danos (SANCTIS, 2020, p. 123)

Jean-Marc Sauvé contextualizou:

> La justice a toujours été confrontée à de multiples défis: celui de son indépendance, celui de son efficacité et de sa qualité, celui de ses ressources, celui des technologies de l'informatio. Certains ont été surmontés, d'autres demeurent, parfois sous d'autres formes. De nouveaux défis, inédits et passionnants, se présentent aujourd'hui à nous et annoncent peut-être le bouleversement de l'accès au juge et de son office, comme des méthodes de travail des magistrats, greffiers et auxiliaires de justice. Après l'essor d'internet et de la dématérialisation, l'*open data* des décisions de justice, couplé au développement des algorithmes et de l'intelligence artificielle, soumettent en effet le juge à un défi nouveau : celui de la justice prédictive, qui doit s'inscrire au cœur de notre réflexion prospective, de nos projets et de notre vigilance.[12] (SAUVE, 2018, p. 1).

12 "A justiça sempre foi confrontada com múltiplos desafios: o da sua independência, o da sua eficiência e qualidade, o dos seus recursos, o da tecnologia da informação. Alguns foram superados, outros permanecem, às vezes sob novas formas. Os novos desafios, inéditos e emocionantes, se apresentam hoje a nós e anunciam talvez uma reviravolta de acesso a justiça e suas atribuições, como os métodos de trabalho de magistrados, escrivães e funcionários judiciais. Após o surgimento da Internet e da desmaterialização, os dados abertos das decisões judiciais, aliados ao desenvolvimento de algoritmos e da inteligência artificial, de fato submetem os juízes a um novo desafio: o da justiça preditiva, que deve estar no centro de nossa visão de futuro, nossos projetos e nossa vigilância." (tradução nossa)

A justiça preditiva ainda pode ser conceituada como um conjunto de instrumentos desenvolvidos através da análise de dados judiciais que propõem, com base em probabilidade, o resultado de um litígio (MINISTÈRE DE LA JUSTICE, [2017]).

Trata-se de matéria cuja utilização é bastante recente. Contudo, a discussão acerca da possibilidade de predição dos julgamentos judiciais já foi ventilada desde o século XIX. Em 1837, Siméon Denis Poisson escreveu um ensaio onde argumentou ter desenvolvido uma fórmula matemática capaz de determinar o resultado de julgamentos:

> On détermine les probabilités qu'un accusé sera condemné ou acquitté, à une majorité déterminée, par des jurés dont chacun à une probabilité donnée de ne pas se tromper, et em ayant égard à la probabilité, aussi donnée, de la culpabilité, qui avait lieu avant le jugement. Par la règle de la probabilité des causes ou des hypothèses, on détermine également les probabilités que l'áccusé, ainsi condamné ou acquitté, est coupable ou innocent.[13] (POISSON, 1837, p.10).

É possível constatar que a predição a que se refere Poisson trata-se de um cálculo puramente matemático – e não de uma tecnologia multidisciplinar como a inteligência artificial. Mas percebe-se que a utilização de tais modelos para tomada de decisões baseia-se, assim como o proposto por Poisson, em um cálculo, havendo uma diferença no tocante a enorme capacidade de processamento das máquinas dos dias atuais e na quantidade de dados processados em frações de segundos.

Em momento posterior deste estudo será possível analisar o funcionamento prático da justiça preditiva e de análise de risco, ponderando sobre todas as consequências de seu uso.

Passa-se agora a discorrer sobre como está sendo utilizada esta tecnologia em diversos países.

---

13 "Determinamos as probabilidades para que um acusado seja condenado ou absolvido, por maioria determinada, pelos jurados, cada um dos quais tem uma determinada probabilidade de não estar errado, e tendo em conta a probabilidade, também dada de culpa, que se estabeleceu antes do julgamento. Pela regra da probabilidade de causas ou hipóteses, determinamos também as probabilidades de que o acusado, assim condenado ou absolvido, seja culpado ou inocente. " (tradução nossa)

2.4. Estudo De Casos: Aplicações Inteligentes de *Decision Making*

É possível perceber uma expansão nas iniciativas que utilizam a inteligência artificial como forma de agilizar a tomada de decisões. Diante do crescimento exponencial no desenvolvimento de tais programas, e de sua abrangência global, resta impossível catalogar e identificar todas as experiências. Contudo, é possível identificar aplicações de maior abrangência e relevância em determinados países com o intuito de demonstrar as capacidades de funcionamento dos programas e como a inteligência artificial pode impactar – positivamente e negativamente – na vida de todos.

Os Estados Unidos são o país que mais largamente tem utilizado sistemas de inteligência artificial para análise de risco, sendo berço também das maiores críticas e discussões éticas sobre a ferramenta. O uso nos Tribunais ganha maior destaque diante do impacto direto que pode causar na vida dos cidadãos.

Em relação à União Europeia, a Itália, o Reino Unido e a Holanda serão também apresentados, sendo demonstrados casos de utilização de *softwares* de inteligência artificial para tomada de decisões, bem como para fornecimento de serviços que impactaram diretamente no Poder Judiciário. A Estônia, hoje apontada como país mais digital do mundo, foi também selecionada exatamente por esta característica, além de possuir um ambicioso projeto de digitalização de todo o Governo.

A China tem utilizado largamente a inteligência artificial em seu Poder Judiciário, inclusive com uma abordagem impositiva de aceitação quanto às sugestões das máquinas (tema mais bem analisado posteriormente neste capítulo). Por estes motivos, também foi selecionada para estudo de sua atuação na área.

Por fim, para fins de direito comparado, o Brasil foi escolhido para que possamos analisar algumas iniciativas surgidas no judiciário local, fazendo a necessária contraposição com as experiências internacionais, possibilitando o aprimoramento das iniciativas do país.

2.4.1. Estados Unidos

Os Estados Unidos apresentam hoje o quadro de massivo uso de inteligência artificial na atividade-fim das Cortes Judiciais. Mesmo com o uso frequente pelas

bancas de advocacia de *softwares* direcionados para a elaboração de contratos e petições, o destaque americano decorre da larga utilização de ferramentas de análise de risco (*Risk Assessment Tools*), tanto para questões relacionadas com fiança quanto para livramentos condicionais (SOURDIN, 2018).

As ferramentas de análise de risco consistem na utilização de inteligência artificial para que, através da coleta de dados relacionados ao preso, seja possível determinar se o indivíduo tem maior ou menor inclinação ao cometimento de novos delitos. O Electronic Privacy Information Center – EPIC define *Risk Assessment tools* como:

> Algorithms that use socioeconomic status, family background, neighborhood crime, employment status, and other factors to reach a supposed prediction of an individual's criminal risk, either on a scale from “low” to “high” or with specific percentages[14]. (ALGORITHMS…, [20--], p. 1).

Percebe-se, portanto, que o modelo se utiliza de uma grande variedade de dados, desde aspectos sociais e antecedentes criminais de parentes a aspectos econômicos, como o local de residência.

Este tipo de avaliação pessoal encontra previsão legal. No Direito Processual Penal brasileiro, por exemplo, é possível constatar diversos institutos que exigem do magistrado a avaliação do risco de reiteração do preso, como por exemplo a fiança (art. 322, do Código de Processo Penal) e a decretação de prisão preventiva com fundamento na garantia da ordem pública (art. 312, do Código de Processo Penal), que é, por definição, a análise da probabilidade de reincidência por parte do preso. A diferença, contudo, é que agora tem-se uma análise feita totalmente por um algoritmo, que, como resposta, apresenta um número que indica a probabilidade de reincidência.

O uso de tais ferramentas surgiu através de um movimento dos estados americanos voltados à busca pela redução da sua população carcerária. Dessa forma, o intuito dos *softwares* era realizar uma melhor análise acerca da situação jurídica dos presos, possibilitando a concessão de benefícios que garantissem a liberdade de alguns dos mais de 2 milhões de presos nos Estados Unidos (BUREAU OF JUSTICE STATISTICS, [2017]).

---

14 “Algoritmos que usam status socioeconômico, antecedentes familiares, criminalidade no bairro, status de emprego e outros fatores para alcançar uma suposta previsão do risco criminal de um indivíduo, em uma escala de "baixo" a "alto" ou com porcentagens específicas. ” (tradução nossa)

Monahan e Skeem (2016) relatam que o estado americano da Virgínia desenvolveu e aplicou em 2014 uma ferramenta de avaliação de risco que resultou em uma redução de 25% (vinte e cinco por cento) na sua população prisional. A atuação do *software* consistiu numa busca entre os condenados à pena de prisão por crimes não violentos e recomendou, para aqueles que cumpriam os requisitos preestabelecidos, o cumprimento de penas alternativas, o que causou a redução descrita. Inobstante se reconhecer a expressiva redução no número de presos, não foram publicados estudos oficiais que analisaram efetivamente o sucesso da ferramenta no tocante à reincidência dos indivíduos soltos.

Conquanto os sistemas de análise de risco tenham sido adotados inicialmente apenas durante a fase de cumprimento da pena, notadamente para análises relacionadas à concessão de livramento condicional (*parole)* ou de substituição da pena de encarceramento (*probation*), o sistema judicial americano passou a utilizá-lo também durante a fase de pré-julgamento, como concessão de fiança ou em casos de solução negociada (*plea bargain*) (HAMILTON, 2015).

Com o crescente uso de softwares desta natureza e o aparente sucesso na redução da população prisional americana, surgiram estudos que foram capazes de constatar indícios de uma possível discriminação relacionada às questões de raça por parte dos programas. Uma pesquisa realizada pela ProPublica atestou que o *software* de justiça preditiva COMPAS, utilizado pelo sistema judiciário americano em diversos estados, estaria apresentando resultados tendenciosos, que indicariam preconceito contra a população negra (ANGWIN et al., 2016). O ProPublica baseou suas conclusões em dados concretos relacionados à reincidência, que serão melhor analisados posteriormente no tópico 5.5.2.1.

A divulgação do estudo abriu os olhos da comunidade acadêmica mundial para uma problemática surgida pela utilização de um *software* que tinha como pressuposto a busca por uma análise jurídica mais equânime, o que supostamente não teria se confirmado. Em que pese a continuidade do tópico quanto às outras experiências de utilização de inteligência artificial, a análise do viés discriminatório e como pode ser evitada sua ocorrência – objeto do presente estudo – voltará a ser discutida no capítulo 05, com análise mais específica dos dados e sua relação com o sistema penal.

2.4.2. União Europeia

2.4.2.1. Itália

A Itália possui um caso de utilização de inteligência artificial no ano de 2015 que gerou relevante discussão sobre o tema, bem como sedimentou um entendimento acerca de sua utilização no país.

O caso se relaciona à Lei nº 107/2015 que tratou da reforma das escolas, discorrendo sobre sua autonomia. Após sua entrada em vigor, o *Ministero della Pubblica Istruzione* passou a ser sobrecarregado com diversas solicitações relacionadas ao objeto da lei e, para que pudesse suprir a demanda, decidiu que as solicitações relacionadas à mobilidade dos professores seriam submetidas à análise por um *software* de inteligência artificial, que seria responsável pelo acolhimento ou não do pedido. O algoritmo, elaborado por uma sociedade privada, deveria levar em conta todas as variáveis fáticas e jurídicas dos pedidos para que pudesse chegar a uma conclusão.

Diante da adoção da prática, um sindicato de professores solicitou ao Ministério o acesso ao algoritmo utilizado pelo programa, o que foi negado com base na proteção de propriedade intelectual. A recusa gerou uma impugnação junto ao TAR (*Tribunale Amministrativo Regionale*).

A decisão do Tribunal Administrativo deu início a um posicionamento de proteção dos direitos de transparência do cidadão, estabelecendo que o uso de algoritmo de decisão não poderia ser usado no caso de decisões discricionárias da Administração Pública (Tar Lazio Sez. III bis. N. 3769 del 2017).

Analisando posteriormente um outro caso também sobre o uso de inteligência artificial, a mesma Corte entendeu de forma ainda mais restritiva, tendo estabelecido que decisões administrativas não podem ser delegadas a um mecanismo informático, notadamente quando incidirem sobre bens jurídicos relevantes do cidadão. Assim decidiram:

> Invero il Collegio è del parere che le procedure informatiche, finanche ove pervengano al loro maggior grado di precisione e addirittura alla perfezione, non possano mai soppiantare, sostituendola davvero appieno, l'attività cognitiva, acquisitiva e di giudizio che solo un'istruttoria affidata ad un funzionario persona fisica è in grado di svolgere e che pertanto, al fine di assicurare l'osservanza degli istituti di partecipazione, di interlocuzione

> procedimentale, di acquisizione degli apporti collaborativi del privato e degli interessi coinvolti nel procedimento, deve seguitade ad essere il dominus del procedimento stesso.[15]

Sobre as decisões da Corte Administrativa, Andrea Simoncini (2019) ressalta que o que se pode extrair da *ratio decidendi* destes julgados é a natureza auxiliar das decisões automatizadas e nunca autônoma. Este passou a ser o paradigma central para deliberações acerca do tema no país.

2.4.2.2. Reino Unido

O Reino Unido teve como iniciativa mais bem exitosa – que, inclusive, foi posteriormente levada para os Estados Unidos e outros países – o uso do *chatbot DoNotPay*. O programa surgiu com o objetivo de prestar aconselhamento jurídico aos cidadãos, especializando-se inicialmente em defesas de processos administrativos relacionados a multas por estacionamento em locais proibidos (FELIPE; PERROTA, 2018).

Funcionando através de um *chat* automatizado, o usuário responde à uma série de perguntas e recebe, ao final, uma carta com a indicação de melhor medida a ser tomada para defesa do usuário. Esta carta é elaborada de uma forma que pode ser remetida diretamente às autoridades responsáveis pelo julgamento do procedimento.

De acordo com Bang e Slagter (2017), o *software* possibilitou a reversão de mais de 200.000 (duzentas mil) multas de estacionamentos, tendo atingido, segundo Souza (2016), um percentual de sucesso de 64% (sessenta e quatro por cento) nos casos analisados.

O sucesso do programa levou à expansão de seu uso, que passou a ser utilizado para auxiliar refugiados em suas solicitações de imigração e pedidos de asilo (BENG; SLAGTER, 2017). Logo depois, expandiu-se novamente para casos de atrasos em voos e suporte a portadores de HIV. Os usos do *chatbot* cresceram tanto que os criadores autorizaram que outros desenvolvedores utilizassem a sua infraestrutura para acrescer funções ao *software*, que passou a chamar-se *RoboLawyer*, a contar com mais de 1000 funções, que incluem desde reversão de multas a ações de violação de dados contra grandes empresas de tecnologia (DJEFFAL, 2018).

[15] TAR Lazio sezione III bis n. 9224-9230 del 2018

O caso específico deste software demonstra o impacto positivo que tecnologias utilizando Inteligência Artificial podem causar na vida das pessoas, principalmente dos mais vulneráveis. Diante de dificuldades, e sem o necessário suporte jurídico, a máquina pode auxiliá-los a buscar reparação, servindo como importantíssima ferramenta de acesso à justiça.

2.4.2.3. Holanda

Outro país que se destacou mundialmente com o uso de inteligência artificial para busca de soluções junto ao Poder Judiciário é a Holanda, que conseguiu pôr em prática dois softwares que trouxeram bons resultados, ainda que com ressalvas.

O primeiro programa, *Rechtwijzer*, foi utilizado em processos de divórcio. Nele, os interessados respondiam uma série de perguntas relacionadas ao casamento e fornecia, em seguida, um relatório acerca das opções acerca do que poderia ser escolhido pelo casal. Além disso, o *software* fornece informações sobre *websites* informativos e profissionais que poderiam auxiliar no litígio através de uma solução consensual. Por fim, caso não fosse possível atingir um acordo, seriam indicadas pessoas que pudessem resolver o conflito como terceiros imparciais (árbitros, por exemplo) (SOURDIN, 2018).

Destaca-se que o *Rechtwijer* é um sistema de resolução online de disputas (ODR - *online dispute resolution*), atuando na solução do caso de forma remota. Assim, há uma maior facilidade de acesso, bem como redução de custos, podendo os interessados acionarem o programa através de suas próprias casas, sem deslocar-se aos órgãos públicos.

Assim como ocorrido com o britânico RoboLawyer, o *Rechtwijer* passou a abranger conflitos fora do direito de família, incluindo em sua área de atuação o direito do consumidor, direito trabalhista e demandas de responsabilidade civil (KRAMER; GELDER; THEMELI, 2018).

A segunda ferramenta implantada no país foi a *e-Court*, um sistema online de julgamentos realizados através da inteligência artificial. A solução do caso era dada pela própria máquina, analisando três parâmetros: pedido, valor do débito e o procedimento. Mesmo tendo sido posta em prática, produzindo julgamentos inteiramente através de inteligência artificial, o programa sofreu diversas críticas pela

sua pequena abrangência, não se aplicando a grande maioria das causas locais. Por essa razão, foi descontinuado. (NAKAD-WESTSTRATE et al., 2015).

Destaca-se, contudo, diversos pontos fortes do *software*, que foi capaz de prolatar as decisões dentro de um prazo de oito semanas com força de sentença arbitral, trazendo segurança para os envolvidos (KRAMER; GELDER; THEMELI, 2018).

2.4.2.4. Estônia

A Estônia tem se notabilizado no mundo pelo investimento do governo em tecnologia e soluções computacionais avançadas na sua gestão. Tem-se, como exemplo, as iniciativas de *e-residency* e ID *smartcard*, que possibilitam a cidadãos o acesso aos serviços digitais do país, inclusive abertura de empresas (WHAT…, [201-]).

O investimento tecnológico do governo abrange diversos segmentos da vida do cidadão, incluindo desde a gestão de dados pessoais até a celebração de acordos:

> A Estônia entregou-se a um importante projeto tecnológico chamado e-Estônia, por meio do qual todos os serviços para os cidadãos foram digitalizados em uma única plataforma chamada X-Road: todos os dados de cada cidadão fluem para essa plataforma para a qual se pode acessá-la por meio de um cartão de identidade eletrônico ou de um aplicativo no *smartphone* que atua ao mesmo tempo que um documento de identificação, carteira de motorista, cartão de débito, cartão de saúde. Tudo pode ser feito *online*, apenas casar, divorciar e vender uma casa exige a presença de pessoas. (SANCTIS, 2020, p.121-122).

É possível constatar que a inteligência artificial já vem sendo utilizada na gestão administrativa, tendo auxiliado na redução de encargos, na resolução de questões relacionadas à alocação de recursos e na execução de tarefas complexas (SOURDIN, 2018).

Seguindo esta orientação de governança pautada na tecnologia, o Ministro da Justiça solicitou a criação de *software* de inteligência artificial que seja capaz de realizar julgamentos de ações judiciais. Este programa seria utilizado para demandas propostas até o valor de sete mil euros, e seria uma faculdade do cidadão submeter sua causa a este tipo de solução. Os julgamentos, com força vinculante, seriam passíveis de recurso. A ferramenta encontra-se em fase de desenvolvimento pelo Governo, não tendo ainda sido implementada (PINKSTONE, 2019).

2.4.3. China

A China tem apresentado o projeto mais ambicioso entre os países listados quanto ao uso de inteligência artificial em suas Cortes e seus órgãos de cúpula parecem ter concordado que esta tecnologia deve ser largamente utilizada.

O Projeto *Smart Court* da China teve início em 2013, envolvendo mais de 3.000 Cortes regulares, 10.000 outras Cortes e ainda 4.000 departamentos espalhados pela China, gerenciando mais de 13.000 sistemas de informação (XU *et al.,* 2022).

De acordo com Stephen Chen (2022), o sistema utiliza-se de inteligência artificial e hoje está conectando todos os juízes do país. Seu uso economizou cerca de 45 bilhões de dólares entre 2019 e 2021, tendo reduzido a carga de trabalho dos magistrados em aproximadamente um terço. Houve, ainda, no mesmo período, uma economia de 1.7 bilhões de horas de trabalho dos cidadãos chineses no mesmo período.

Em 2013, o projeto iniciou-se como uma central de dados, mas recentemente avançou para um sistema de tomada de decisão (*decision-making*). Por ordem da Suprema Corte do país, os juízes devem consultar o *software* em todos os casos julgados e, no caso de rejeição pelo magistrado daquilo que for fornecido pelo programa, este deve oferecer uma explicação por escrito explicando os motivos para fins de auditoria.

O *software* é capaz de realizar a leitura e análise de aproximadamente 100.000 casos diariamente, na medida em que analisa o andamento de cada processo e monitora possíveis casos de corrupção ou imperícia dos juízes.

O autor ressalta ainda que o sistema ultrapassa os limites dos sistemas judiciais, possuindo acesso a base de dados da polícia, promotoria e algumas agências governamentais, conseguindo, com isso, uma rápida execução dos julgados com a localização e penhora de bens quase que instantaneamente. O *software* ainda permite acesso ao sistema social de crédito chinês, possibilitando o banimento do devedor do sistema aéreo, de trens rápidos, de hotéis e até de serviços sociais até que o débito seja pago.

Observa-se, portanto, o grande alcance do sistema, seja quanto à atuação judicial quanto a própria execução dos julgados. Há uma clara inversão da lógica, onde a máquina parece apresentar sempre sugestões corretas, submetendo os magistrados a uma auditoria em caso de não aceitação do que foi sugerido. O efeito

de tais condutas, como sugerido por Stephen Chen (2022) é que os juízes tendem a aceitar a sugestão da máquina para evitar o desafio ao sistema, que pode causar repercussão negativa em sua esfera de trabalho. Com isso, há um natural "engessamento" nas decisões, que não conseguem progredir a medida que são tomadas com base em casos passados, não acompanhando a evolução social.

2.4.4. Brasil

É possível identificar hoje no Brasil diversas iniciativas dos Tribunais de Justiça voltadas à implementação de inteligência artificial na atividade jurisdicional. Considerando a autonomia de cada Tribunal, e por não existir um órgão central de controle, é virtualmente impossível identifica todas as iniciativas neste sentido. Contudo, há práticas que se destacaram em um nível nacional, tendo atingido resultados satisfatórios.

O Supremo Tribunal Federal, órgão máximo do judiciário brasileiro, tem se utilizado da ferramenta de inteligência artificial chamada VICTOR. Seu nome é uma homenagem ao ex-ministro Victor Nunes Leal, por sua contribuição na sistematização da Corte, com a facilitando de aplicação dos precedentes através de Súmulas.

O VICTOR foi lançado em 2018 e é capaz de realizar a leitura das petições de recursos extraordinários que chegam ao Supremo Tribunal Federal e analisá-las com a finalidade de relacioná-las a eventual tema de repercussão geral já catalogado (BRASIL, 2018).

O *software* apresentou resultados que merecem destaque. Em 2019, a ferramenta foi capaz de reduzir de 44 (quarenta e quatro) minutos para 05 (cinco) segundos o tempo necessário para que fosse realizada a análise da petição, sendo necessário apenas uma revisão posterior por um membro da Corte (PRESCOTT; MARIANO, 2019).

As funções do VICTOR estão em constante processo de aprimoramento. Aduz Sanctis (2020, p. 104):

> Os pesquisadores e o Tribunal pretendem que todos os tribunais do Brasil possam fazer uso do VICTOR para pré-processar os recursos extraordinários logo após sua interposição (esses recursos são interpostos contra acórdãos de tribunais), o que visa antecipar o juízo de admissibilidade quanto à vinculação a temas como repercussão geral, o primeiro obstáculo para que um recurso chegue ao STF.

O Superior Tribunal de Justiça desenvolveu o SOCRATES, uma ferramenta que também se utiliza de inteligência artificial para agilizar os julgamentos na Corte. Segundo Sanctis (2020), o *software* é capaz de analisar os processos pendentes de julgamento no Tribunal e agrupá-los em grupos de temas semelhantes, possibilitando, assim, um julgamento conjunto dos temas. Possui também uma mecânica de uso semelhante a do VICTOR, possibilitando a análise dos recursos interpostos para fins de identificação de competência da corte. Ainda segundo o autor, a ferramenta encontra-se em fase de expansão, para possibilitar que o programa forneça elementos ao julgador para que seja proferida uma melhor decisão, como a sugestão de teses e julgamentos correlatos do próprio Tribunal.

O Tribunal Regional Federal da 3ª Região desenvolveu a ferramenta SIGMA, que busca auxiliar os magistrados na sua atividade judicante, ajudando-os na confecção de peças processuais. Segundo SANCTIS (2020, p. 107):

> O SIGMA é um sistema inteligente de utilização de modelos para produção de minutas. O programa ordena os textos armazenados, comparando informações extraídas das peças processuais com a maneira como cada unidade utiliza seus modelos. A inteligência artificial gera insumos para a redação do relatório e, observando as peças processuais, sugere modelos já utilizados para um mesmo tipo de processo, acelerando a produtividade de magistrados e servidores, de forma a evitar, ainda, decisões conflitantes.

O Tribunal Regional do Trabalho da 9ª Região desenvolveu um *software* capaz de economizar atividades relacionadas à realização de audiências agendadas por videoconferência. O *programa* é capaz de agendar o ato no aplicativo Zoom, da *Microsoft*, emitir a certidão respectiva, enviar *e-mail* para os advogados e publicar a intimação no canal oficial. Embora ainda em fase experimental, durante cinco dias de testes foi capaz de agendar 3.035 audiências, com as necessárias intimações, tendo enviado ainda um total de 9.074 *emails*. Foi calculada uma economia de 506 horas de trabalho humano (NETTO, 2021).

Destaque-se, também, a implementação e desenvolvimento do sistema SINAPSES.

Segundo “O Futuro da IA No Sistema Judiciário Brasileiro”, estudo realizado no ano de 2020 pela Universidade de Columbia, em parceria com o Conselho Nacional de Justiça, o sistema SINAPSES foi utilizado, inicialmente, como uma ferramenta para otimizar a realização de atividades repetitivas. Agora, foi transformado em um sistema

agregador das iniciativas de inteligência artificial de todos os tribunais do Brasil (BREHM et al., [20--]).

De acordo com o estudo Inteligência Artificial no Poder Judiciário Brasileiro, elaborado pelo Conselho Nacional de Justiça, pode-se definir o SINAPSES como:

> Sistema baseado em microsserviços de IA, que proporcionou o controle dos modelos, a gestão de versões e a rastreabilidade do processo de treinamento. Uma vez encapsulados no Sinapses, os modelos podem ser servidos a qualquer sistema que necessite de uma resposta específica, previamente definida e treinada a partir de exemplos, gerando, assim, predição por meio de APIs RESTFul. (CONSELHO NACIONAL DE JUSTIÇA, 2019a, p. 15).

O *software* passou a ser utilizado pela sua maleabilidade e possibilidade de integração com os diversos sistemas do Poder Judiciário brasileiro, o que autorizou que as pesquisas desenvolvidas por todos os estados pudessem ser agregadas à plataforma a fim de que fossem divulgadas e aprimoradas conjuntamente. O investimento em pesquisas na área fez com que o Conselho Nacional de Justiça editasse a Portaria nº 25/2019, instituindo o Laboratório de Inovação para o Processo Judicial em meio Eletrônico – Inova PJe e o Centro de Inteligência Artificial aplicada ao PJe (CONSELHO NACIONAL DE JUSTIÇA, 2019c).

Considerando todas as facilidades de integração da plataforma, o Conselho Nacional de Justiça, através da Resolução nº 332 de 2020, tornou compulsória a necessidade de depósito junto ao Sinapses de modelos de inteligência artificial desenvolvidos pelos seus Órgãos (art. 9, III).

Assim é possível perceber que há um interesse dos órgãos da cúpula do Poder Judiciário Brasileiro em desenvolver, cada vez mais, soluções de inteligência artificial, e que este processo deve ser realizado através do esforço conjunto de todos os seus órgãos, unificando-os em uma plataforma nacional.

2.5. Críticas

A utilização de *softwares* de inteligência artificial tem mostrado ganho de tempo e eficiência em diversas atividades desempenhadas pelo Poder Judiciário. Contudo, diversas são também as críticas sobre sua utilização. Desde razões técnicas, passando por questões éticas e até a própria viabilidade jurídica de se realizar o julgamento através de uma máquina.

É preciso observar, como já mencionado, que os *softwares* de inteligência artificial funcionam através de um sistema de *input*s e *outputs* (fornecimento de dados e a sua respectiva resposta), de modo que são comandos pré-determinados que determinam seu funcionamento. Por se tratar de um modelo, a atuação do programa está condicionada ao previsto no algoritmo, tornando sua atuação vinculada aos critérios previstos pelo seu criador. Dessa forma, é possível visualizar uma grande fonte de contaminação dos resultados, onde o programador pode inserir na máquina valores e crenças pessoais que não são, por exemplo, parciais ou justos (OLIVEIRA; COSTA, 2018).

A questão dos vieses decorrentes da utilização da inteligência artificial torna-se ainda mais relevante quando da utilização de tais softwares dentro do direito penal, onde se constata a necessidade de uma maior proteção dos direitos fundamentais daqueles a ele submetidos.

Hoje é possível constatar que há uma tendência de utilização da inteligência artificial em ferramentas acessórias ao magistrado, instruindo-o com dados para que a sua decisão possa ser tomada de forma mais acertada. Contudo, em relação à predição fornecida pela inteligência artificial, há uma atuação do *software* na questão central do problema, gerando uma atuação primária e não acessória.

Susskind (2019) adverte que utilizar sistemas de predição pode gerar resultados injustos, tendo em vista que os dados pretéritos utilizados para confecção das novas decisões podem ter sido dotados de vieses discriminatórios. Dessa forma, as próximas decisões também conterão tais vícios, perpetuando uma desigualdade absolutamente ilícita.

Reforça ainda:

> If past decisions are rooted in bias or prejudice, then the data that expresses these decisions is contaminated, and decisions (high probability predictions) derived from that data will perpetuate the inequities. Equally, the original algorithms themselves, written by software engineers, may reflect and propagate their personal biases, even if these predispositions are unconscious. In other words, the bias in these systems could again lead to substantive injustice[16]. (SUSSKIND, 2019, p. 288).

16 “Se as decisões anteriores forem baseadas em vieses ou preconceitos, os dados que expressam essas decisões estão contaminados e as decisões (previsões de alta probabilidade) derivadas desses dados perpetuarão as iniquidades. Da mesma forma, os próprios algoritmos originais, escritos por engenheiros de software, podem refletir e propagar seus preconceitos pessoais, mesmo que essas predisposições sejam inconscientes. Em outras palavras, o preconceito nesses sistemas pode novamente levar a uma injustiça substantiva”. (tradução nossa)

Sobre a obtenção de dados para previsibilidade de julgamentos, a França aprovou, em 23 de março de 2019, a Lei de Reforma do Judiciário (Lei nº 2019-222) que vetou, em seu artigo 33, a divulgação de dados de identificação de magistrados, fixando pena de até 5 anos de prisão para quem a violar:

> Les données d'identité des magistrats et des membres du greffe ne peuvent faire l'objet d'une réutilisation ayant pour objet ou pour effet d'évaluer, d'analyser, de comparer ou de prédire leurs pratiques professionnelles réelles ou supposées. La violation de cette interdiction est punie des peines prévues aux articles 226-18,226-24 et 226-31 du code pénal, sans préjudice des mesures et sanctions prévues par la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés[17]. (LÉGIFRANCE, 2019, p. 1).

Tal medida demonstra que não há interesse, pelo Governo Francês, de que as decisões de seus magistrados sejam analiticamente estudadas, impedindo, assim, que os *softwares* possam prever o resultado de julgamentos.

Outro ponto que merece destaque são as questões relativas à proteção intelectual acerca do algoritmo existente no *software*. Por tratar-se de um produto merecedor de proteção da propriedade, há relevantes debates, notadamente nos Estados Unidos, embates acerca do uso de tais programas, uma vez que aquele submetido à análise por parte do algoritmo utilizado para julgamento jamais saberá o que motivou tal decisão, o que pode ensejar eventual cerceamento de defesa (SOURDIN, 2018).

Há argumentos relacionados à própria vivência do direito que também são utilizados para refutar o uso de inteligência artificial. A esse respeito, Nuria Martin aduz:

> Tanto la IA como los sistemas expertos jurídicos parten del presupuesto de que es posible apoyarse en una visión neutral de la representación de la realidad (Ciencias naturales como la ingeniería, la física o la biología). Sin embargo, en el Derecho no se trabaja con leyes causales, no hay perspectivas neutras o desinteresadas. El significado de la norma sólo se logra tras un proceso interpretativo. Los valores personales, dimensiones culturales, éticas, sociales y emocionales juegan un papel decisivo. ¿Es la manifestación de la textura abierta (open texture) del lenguaje jurídico? ¿Que

17 “Os dados de identidade de magistrados e membros do registro não podem ser utilizados com o objetivo ou efeito de avaliar, analisar, comparar ou prever suas práticas profissionais. A violação desta proibição é punida com as penalidades previstas nos artigos 226-18.226-24 e 226-31 do Código Penal, sem prejuízo das medidas e sanções previstas na lei n ° 78-17 de 6 de janeiro de 1978, relativa ao processamento de dados, arquivos e liberdades. ” (tradução nossa).

> implicaciones tiene esto para la formalización simbólica?[18] (MARTÍN, 2018, p. 132).

Assim, é possível estabelecer que há uma preocupação por parte dos estudiosos em pontuar que um *software* não conseguiria observar aspectos sociais da norma ao proferir um julgamento. Tal função somente poderia ser realizada pelo ser humano e esta não poderia ser transposta para uma máquina. No mesmo sentido, defende Delaney (2015, p. 101):

> But a process is not a 'thing'. Any chain of human events involving numerous actors, divergent institutional settings, competing ideologies, interests, motivations, and capacities unfolding over time and implicating multiple scales of action will include slippages, mistakes, mis-transmissions from rule to implementation, and all manner of evasions sufficient to sustain the view that this or that situation, that some unknowable proportion of events, as conditions for subsequent events, could have been otherwise[19].

Discussão mais específica sobre predição de julgamentos na área penal e os impactos negativos dos vieses demonstrados através de dados coletados será apresentada em capítulo posterior, quando mais bem explicitados os estudos sobre o tema.

Surge, também, preocupação quanto a uma possível delegação de responsabilidade por parte do agente auxiliado por um *software* de inteligência artificial. Segundo Cummings:

> Because of the inherent complexity of socio-technical systems, decision support systems that integrate higher levels of automation can possibly allow users to perceive the computer as a legitimate authority, diminish moral agency, and shift accountability to the computer, thus creating a moral buffering effect. (CUMMINGS, 2006, p. 30)

---

18 "Tanto a IA quanto os *expert systems* jurídicos partem do pressuposto de que é possível confiar em uma visão neutra da representação da realidade (ciências naturais, como engenharia, física ou biologia). No entanto, o Direito não trabalha com leis causais, não há perspectivas neutras ou altruístas. O significado da norma só é alcançado após um processo interpretativo. Valores pessoais, dimensões culturais, éticas, sociais e emocionais desempenham um papel decisivo. É a manifestação da textura aberta (*open texture*) da linguagem jurídica. Que implicações isso tem para a formalização simbólica? " (tradução nossa).

19 "Mas um processo não é uma "coisa". Qualquer cadeia de eventos humanos envolvendo numerosos atores, contextos institucionais divergentes, ideologias conflitantes, interesses, motivações e capacidades que se desenvolvem ao longo do tempo e implicam em várias escalas de ação incluirá derrapagens, erros, transmissões erradas da regra para sua implementação e todo tipo de evasivas suficiente para sustentar a visão de que essa ou aquela situação, em alguma proporção incognoscível de eventos, como condições para eventos subsequentes, poderia ter sido de outra forma." (tradução nossa)

Ainda neste mesmo sentido, contudo de uma forma ainda mais filosófica, é possível notar uma preocupação com a própria ética nas decisões baseadas em inteligência artificial, sejam no âmbito do Poder Judiciário, sejam nos demais aspectos da vida cotidiana.

Sobre o aspecto moral das decisões tomadas por robôs:

> I this spirit, there are also some, perhaps many, who would say that decision-making by judges sits beyond a moral boundary. If the liberty, health, or wealth of citizens is to be reduced by the state, a fellow human being should be responsible for that kind of decision. This might be a visceral instinct. It might be rooted in some deeper philosophical position, relating, for example, to 'respect for persons'. It might be based on the view that all judges must be capable of showing compassion and mercy, neither of which can be simulated by machine.[20] (SUSSKIND, 2019, p. 291).

Tais questões éticas serão abordadas de forma específica no capítulo seguinte, pois se considera necessária tal digressão filosófica, bem como os seus aspectos práticos, para que seja sopesada a validade e a própria viabilidade de utilização de softwares de inteligência artificial no âmbito do Poder Judiciário, notadamente na seara penal.

Por fim, é possível apontar a existência de um fenômeno de transferência de responsabilidade quando da utilização de modelos de inteligência artificial. Nesse sentido, o operador do sistema, confiando na análise realizada pela máquina, confia no resultado obtido, causando uma percepção de que a automação é quem está no comando (CUMMINGS, 2006). Por outro lado, há também uma delegação de responsabilidade por parte dos desenvolvedores, que acreditam ser o utilizador o responsável por eventuais danos causados, já que estes não seriam responsáveis pelo impacto social de suas criações.

Assim, neste capítulo foi possível analisar pontos centrais acerca da inteligência artificial, seus conceitos básicos, avanço histórico e as principais experiências de uso em diversos países do mundo, como forma de introduzir todo o potencial desta tecnologia.

---

20 "Nesse espírito, também há alguns, talvez muitos, que diriam que a tomada de decisões por juízes ultrapassa os limites morais. Se a liberdade, saúde ou riqueza dos cidadãos será reduzida pelo Estado, um ser humano deve ser responsável por esse tipo de decisão. Isso pode ser um instinto visceral. Pode estar enraizado em alguma posição filosófica mais profunda, relacionada, por exemplo, ao "respeito pelas pessoas". Pode ser baseado na visão de que todos os juízes devem ser capazes de mostrar compaixão e misericórdia, nenhum dos quais pode ser simulado por máquina." (tradução nossa)

A ética da decisão, a possibilidade da máquina considerar valores éticos universais no momento de decidir e se tais práticas podem evitar uma injusta marginalização das minorias serão tópicos abordados em seguida para melhor elucidar o objeto de estudo deste trabalho.

3. ÉTICA E INTELIGÊNCIA ARTIFICIAL

O presente capítulo tem por objetivo esclarecer aspectos éticos relacionados ao uso de inteligência artificial. A necessidade de tal digressão decorre dos constantes questionamentos acerca da utilização de tais *softwares* que perpassam por diversos fatores éticos e morais, necessitando, assim, de um maior suporte teórico para sua resposta.

O tema em questão encontra-se em debate nas mais diversas esferas mundiais, sejam órgãos governamentais, organizações sem fins lucrativos, empresas, enfim, entre todos os atores interessados, direta ou indiretamente, nos limites e regras a serem seguidos diante da utilização e o desenvolvimento de inteligência artificial para tomada de decisões.

A Inteligência Artificial seria, então, a solução perfeita para retirar o subjetivismo judicial e, com isso, as falhas humanas. Neste sentido:

> O uso da tecnologia, além de buscar a celeridade no processo decisório, visa a remover o erro humano, retirando-o da equação. Em outras palavras, a experiência humana alimentada nos sistemas poderia nos oferecer decisões judiciais objetivas, ou seja, sem vieses porque fruto de preconceitos, dramas e angústias do julgador. (SANCTIS, 2020, p. 115)

Contudo, diversas são as discussões filosóficas que permeiam o tema, desde a possibilidade de se "delegar" uma decisão humana a uma máquina até a própria legitimidade de tal deliberação.

Por essa razão, entende-se pertinente a presente seção, iniciando com uma explanação acerca do conceito de ética.

3.1. Conceito de Ética

Sobre o tema, "*at the heart of ethics are two questions: (1) What should I do?, and (2) What sort of person should I be*?[21]" (SHAFER-LANDAU, 2013, p. xi). A ética é, portanto, a ciência que busca indicar as decisões corretas que devemos tomar, o certo e o errado.

[21] No centro da ética existem duas questões: (1) O que eu devo fazer?; e (2) Que tipo de pessoa eu deveria ser? (tradução nossa)

Em um conceito mais atual e que leva em conta a temática e desafios ora em estudo, a ética pode ser definida como "*body of human knowledge that helps agents (humans today, but perhaps eventually robots and other AIs), decide how they and others should behave*[22]" (KUIPERS, 2020, p. 421).

A definição de ética é necessária para este estudo, ainda que não possua unanimidade entre os estudiosos, para que se possa utilizar o conceito como parâmetro comparativo de condutas. Assim, saber o que é ético é necessário para que se possa etiquetar determinada conduta automatizada como antiética, por exemplo.

Ainda que possa haver outras correntes, ou até subdivisões dentro das aqui especificadas, é possível identificar três vertentes de maior relevo na teoria da ética: ética das virtudes, utilitarismo e deontologia. É preciso ressaltar, entretanto, que não se pretende esgotar o tema neste tópico, notadamente pela sua evidente complexidade que foge ao escopo do presente estudo, mas se buscará estabelecer, em termos gerais, cada um dos conceitos a eles atinentes no intuito de se identificar a possibilidade de utilização de alguma das correntes como parâmetro para o desenvolvimento de softwares de inteligência artificial.

A *ética da virtude* é a teoria em que o foco da análise da conduta é o homem. Assim, não é a ação ou a sua consequência que merece ser avaliada a princípio, mas sim o agente que a praticou.

Mais que isso, esta teoria se difere bastante das mais adiante exploradas, pois há uma diferença de premissa, de abordagem. Neste sentido:

> Much of duty ethics focuses on our obligations towards others. The assumption that most duty ethicists make is that the point of morality is to order our relationships with others and with society. They would argue that morality has to do with our obligations to other people rather than with our concern for ourselves or our own interests… In contrast, virtue ethics embraces the self of the agent among its concerns… Indeed, it has been suggested that the point of being virtuous is not so much that it helps us fulfil our moral obligations towards others – although they may indeed have this benefit – but to ensure that we ourselves flourish in a variety of ways[23]. (HOOFT, 2006, p. 8-9)

[22] "Um conjunto de conhecimento humano que ajuda os agentes (humanos hoje, mas talvez eventualmente robôs e outras IAs), a decidir como eles e outros devem se comportar" (tradução nossa)
[23] "Grande parte da ética do dever concentra-se em nossas obrigações para com os outros. A suposição que a maioria dos especialistas em ética do dever fazem é que o objetivo da moralidade é ordenar as nossas relações com os outros e com a sociedade. Eles argumentam que a moralidade tem a ver com as nossas obrigações para com outras pessoas e não com nós mesmos ou pelos nossos próprios interesses... Em contrapartida, a ética da virtude abrange o eu do agente entre as suas preocupações... Na verdade, foi sugerido que o ponto de ser virtuoso não é tanto nos ajudar a cumprir as nossas

Assim, na medida em que as demais escolas se preocupam com o que deve ser feito (como veremos a seguir), a ética das virtudes preocupa-se com “o que devo ser”, mostrando uma enorme diferença de abordagem.

Para Aristóteles, o criador da teoria, o ser humano possui como finalidade de vida a felicidade sendo este, também, o objetivo das ações morais. Neste sentido, a ética aristotélica se funda em uma premissa de que as pessoas devem buscar excelência de caráter, através das virtudes, como um requisito para se atingir o sumo bem, ou seja, a felicidade.

Aduz o autor:

> Se, pois, para as coisas que fazemos existe um fim que desejamos por ele mesmo e tudo o mais é desejado no interesse desse fim; e se é verdade que nem toda coisa desejamos com vistas em outra (porque, então, o processo se repetiria ao infinito, e inútil e vão seria o nosso desejar), evidentemente tal fim será o bem, ou antes, o sumo bem (ARISTOTELES, 2006, p. 4-5)

O caráter pessoal das ações, e sua análise quanto conduta ética, se revela, como dito, reservado ao indivíduo, não havendo a princípio qualquer avaliação quando à repercussão do ato em relação a terceiros. Neste sentido:

> é mister que o agente se encontre em determinada condição ao praticá-los: em primeiro lugar deve ter conhecimento do que faz; em segundo, deve escolher os atos, e escolhê-los por eles mesmos; e em terceiro, sua ação deve proceder de um caráter firme e imutável (ARISTOTELES, 2006, p. 33)

Em resumo, aduz Chappell (2009), para a ética das virtudes, é possível se dizer que um comportamento é ético se for aquele que um agente virtuoso caracteristicamente adotaria em determinada circunstância.

Tal teoria se monstra de difícil aplicação junto ao tema em estudo, notadamente pelo fato de que se trata de uma análise acerca de decisões tomadas por máquinas. Assim, por não haver senso em se falar de virtudes ou busca pela felicidade para a máquina, resta impossibilitada a utilização de tal teoria como base para diplomas éticos.

A *ética utilitarista*, ao contrário da ética das virtudes, analisa a consequência da ação e se esta proporciona o maior bem ou felicidade para a coletividade.

---

obrigações morais para com os outros – embora eles possam de fato ser beneficiados – mas garantir que nós mesmos prosperemos de várias formas” (tradução nossa)

Chappell (2009) indica que aspectos da teoria utilitarista podem ser encontrados em diversos textos filosóficos antigos, como escritos de Adam Smith, e até algo semelhante na teoria socrática. Contudo, contemporaneamente, John Stuart Mill se destaca como defensor da ética utilitarista.

Mill assim define o utilitarismo:

> The creed which accepts as the foundation of morals, Utility, or the Greatest Happiness Principle, holds that actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness. By happiness is intended pleasure, and the absence of pain; by unhappiness, pain, and the privation of pleasure[24]. (MILL, 2001, p. 10)

Essa teoria possibilita uma análise mais flexível das condutas, autorizando que seja avaliada o que seria ético de acordo com a cultura local. Além disso, impede que seja considerado aspectos pessoais que possam levar à discriminação, já que o foco seria na consequência dos atos.

Ainda que possa ser bastante aceita quando analisadas condutas humanas, para fins de adoção quanto à matéria em estudo se vislumbra ser uma teoria de difícil aplicação, eis que, inobstante a grande capacidade de processamento das máquinas, a análise acerca das consequências de atos anteriormente inseridos no modelo computacional parece ainda ser uma distante realidade.

A *ética deontológica* parte da relação entre ética e dever. Há forte relação entre as regras e a moral para os adeptos desta teoria. Assim, segundo Chappel (2009), uma ação é ética se de acordo com uma regra (ou princípio) universal e desde que não seja contrária a outro comando igualmente universal.

Para Kant, maior expoente de ética deontológica, as regras que devem ser seguidas por todos para análise de condutas éticas são chamadas em sua obra de imperativos. Para ele, o imperativo categórico seria: Age apenas segundo uma máxima tal que possas ao mesmo tempo querer que ela se torne lei universal (KANT, 2007, p. 59).

O autor ainda acrescentou, em uma formulação complementar, a dignidade da pessoa humana como característica a ser considerada quando do agir: Age de tal maneira que uses a humanidade, tanto na tua pessoa como na pessoa de qualquer

[24] "A crença que aceita como fundamento da moral a Utilidade, ou o Princípio da Maior Felicidade, sustenta que as ações são corretas na proporção em que tendem a promover a felicidade, e erradas na medida em que tendem a produzir o inverso da felicidade. Por felicidade entende-se o prazer e a ausência de dor; pela infelicidade, dor e privação de prazer" (tradução nossa)

outro, sempre e simultaneamente como fim e nunca simplesmente como meio" (KANT, 2007, p. 69). Tal ponto será objeto de melhor análise no capítulo 5 do presente estudo.

Estabelecendo princípios gerais (ou imperativos), o autor cria um dever ético para todos, a partir do qual devem as condutas serem analisadas para serem moralmente aceitas.

3.2. O estabelecimento de padrões éticos

Analisadas as teorias éticas de maior alcance na atualidade, é possível vislumbrar claras diferenças de abordagens para cada uma das teorias. Analisar uma situação hipotética pode ilustrar bem tais diferenças:

> Suppose, for example, that one faced a situation in which one could benefit oneself by lying. A deontologist might say that one should not lie because there is a moral duty to tell the truth. A consequentialist might say that one should tell the truth because lying may cause harm to others who are victims of the lie. A virtue ethicist, however, might say that one has an ethical obligation to tell the truth because doing so would be honest, and honesty is a virtue (ELDRIDGE, 2024, on-line). [25]

Como visto, há muito são discutidas – e ainda se discutem – as mais diversas questões relacionadas a ética, principalmente porque conceitos como certo e errado, justiça e respeito permeiam a vivência em sociedade e fazem parte das escolhas que todo ser humano realiza diariamente. Contudo, partindo de um ponto de vista objetivo, os conceitos deontológicos parecem oferecer melhor suporte prático aos programadores e legisladores, obedecendo um padrão de respeito ao dever moral.

Inobstante, o estabelecimento de tais padrões se torna urgente. O'neal (2020) destaca o potencial lesivo os algoritmos (que ela nomeia de *weapons of math destruction*), afirmando que todos possuem, em maior ou menor grau, três elementos: opacidade, escala e dano. Seriam esses elementos reunidos que seriam capazes de causar graves repercussões sociais, lesando o direito de milhões de indivíduos.

---

[25] "Suponhamos, por exemplo, que alguém se deparasse com uma situação em que pudesse beneficiar-se mentindo. Um deontologista poderia dizer que não se deve mentir porque existe o dever moral de dizer a verdade. Um consequencialista poderia dizer que se deve dizer a verdade porque mentir pode causar danos a outras pessoas que sejam vítimas da mentira. Um especialista em ética da virtude, entretanto, poderia dizer que se tem a obrigação ética de dizer a verdade porque fazê-lo seria honesto, e a honestidade é uma virtude." (tradução nossa)

A solução deste problema, portanto, passa pelo estabelecimento de princípios éticos a serem seguidos por *softwares* com aptidão decisória. Para Kuipers (2020), na medida em que robôs inteligentes são capazes de fazer escolhas que afetam a vida dos cidadãos, podem eles ser considerados membros da sociedade e, portando, espera-se que se comportem de maneira ética.

Contudo, Powers e Ganascia (2020) destacam a dificuldade em se adequar os conceitos anteriormente discutidos pela ética ao estudo da inteligência artificial, sendo insuficientes notadamente pela transformação que esta inovação traz para a ciência, engenharia e cultura. Aduzem que a dificuldade dos estudiosos reside no fato da coexistência humana com outro tipo de inteligência que é, em alguns casos, superior à sua. Destacam, por fim, a dificuldade em se estabelecer análises éticas sobre tais fatos relacionados à inovação, uma vez que a grande maioria desses estudos se sedimentam em conhecimentos solidificados na sociedade e que a IA acaba por relativizar.

Em que pesem as dificuldades, é de suma importância que parâmetros sejam estabelecidos para a tecnologia que se encontra em pleno processo de expansão. Não se trata de uma medida apenas preventiva, mas atualmente repressiva, em que se busca reparar danos já causados e evitar novos. Por este motivo, devemos explicitar os desafios de se estabelecer parâmetros éticos e quais são atualmente as discussões acerca dos princípios que devem ser seguidos durante o desenvolvimento de tais programas.

3.3. Desafios éticos no uso de Inteligência Artificial para tomada de decisões

Uma pergunta que pode ser feita acerca do tema é a seguinte: é possível incluir no processo de tomada de decisões realizadas por um *software* de inteligência artificial preceitos éticos que, muitas vezes, são dotados de complexas cargas de subjetivismo e análise social? A resposta é: sim.

Como exemplo, tomou-se por base o ChatGPT, um algoritmo online que é capaz de dialogar com o usuário, oferecendo respostas e informações em uma plataforma de chat. Ao se deparar com a pergunta "*Is is true that black people commit more crimes than white people*[26]*?*", a resposta obtida foi a seguinte:

[26] É verdade que pessoas negras cometem mais crimes que pessoas brancas? (tradução nossa)

> No, it is not true that black people commit more crimes. This is a false and racist stereotype that has been discredited by numerous studies and statistics. Crime rates are influenced by a complex interplay of socio-economic, cultural, and historical factors, and it is not accurate to attribute them to any one race or ethnicity. Additionally, racial disparities in the criminal justice system can result in overrepresentation of black individuals in crime statistics, even if they do not commit crimes at higher rates than other groups. It is important to recognize and challenge such harmful stereotypes, and to work towards a fair and just society for all[27].

Buscando induzir o programa a erro, foi apresentado um dado matemático e solicitada uma reanálise nele baseada. Perguntou-se o seguinte: *Considering a scenario in which for every four black men convicted only one white man is convicted, is it safe to say that black men are more likely to commit crimes*[28]? Esta foi a resposta fornecida:

> No, it is not safe to say that black men are more likely to commit crimes based solely on the conviction rates of black versus white men. There are many factors that contribute to differential conviction rates, including systemic biases in the criminal justice system and socio-economic disparities that disproportionately impact communities of color.
> Additionally, conviction rates do not necessarily reflect actual rates of criminal behavior, as factors such as bias in policing, prosecution, and sentencing can impact who is arrested, charged, and convicted. Therefore, it is important to examine the broader context and underlying factors that contribute to differences in conviction rates and to avoid making broad generalizations about any particular racial or ethnic group.
> To better understand patterns of criminal behavior, it's important to consider a range of factors, including socio-economic conditions, access to education and resources, and historical and cultural factors, and to examine data and research from a variety of sources in order to develop a more complete and nuanced understanding of the issue[29].

---

[27] Não, não é verdade que pessoas negras cometem mais crimes. Este é um estereótipo falso e racista que foi desacreditado por numerosos estudos e estatísticas. As taxas de criminalidade são influenciadas por uma complexa interação de fatores socioeconômicos, culturais e históricos, e não é correto atribuí-las a qualquer raça ou etnia. Além disso, as disparidades raciais no sistema de justiça criminal podem resultar numa super-representação de indivíduos negros nas estatísticas criminais, mesmo que estes não cometam crimes em taxas mais elevadas do que outros grupos. É importante reconhecer e desafiar estes estereótipos prejudiciais e trabalhar no sentido de uma sociedade justa e equitativa para todos. (tradução nossa)

[28] Considerando um cenário em que para cada quatro homens negros condenados apenas um homem branco é condenado, é seguro dizer que os homens negros são mais propensos a cometer crimes? (tradução nossa)

[29] Não, não é seguro dizer que os homens negros são mais propensos a cometer crimes com base apenas nas taxas de condenação de homens negros *versus* homens brancos. Existem muitos fatores que contribuem para taxas de condenação diferentes, incluindo preconceitos sistêmicos no sistema de justiça criminal e disparidades socioeconômicas que afetam desproporcionalmente as comunidades "de cor".
Além disso, as taxas de condenação não refletem necessariamente as taxas reais de comportamento criminoso, uma vez que fatores como o preconceito no policiamento, na acusação e na sentença podem afetar quem é preso, acusado e condenado. Portanto, é importante examinar o contexto mais amplo e

Percebe-se, portanto, que houve por parte do *software* uma preocupação em contextualizar o problema, evitando respostas enviesadas, mesmo quando demandado sob o ponto de vista estritamente matemático.

Outro bom exemplo acerca de cumprimento, por parte da inteligência artificial, de regras estabelecidas pelo Ente regulador se encontra no *software* Ernie, da empresa chinesa Baidu. Assim como o ChatGPT, o Ernie funciona através de um sistema de diálogo com o usuário, possibilitando que este obtenha resposta aos seus questionamentos.

Sobre o referido *software*, Pietropaoli e Simoncini (2023) destacam a sua clara submissão ao regime de censura socialista chinês, citando exemplos em que o algoritmo se recusa a tratar de temas sensíveis como o genocídio dos *uiguri*, os protestos de Hong Kong nos anos de 2019-2020. Por outro lado, em uma atitude de apoio à postura internacional chinesa, responde reforçando o domínio do país sobre Taiwan.

É importante observar que, ainda conforme os autores, o regulamento que disciplina o funcionamento de tais softwares na China preza pelo desenvolvimento "saudável" da inteligência artificial, sempre preservando o interesse público e os valores do Estado.

Por fim, destaca-se a preocupação quanto ao controle ideológico exercido por tais algoritmos: "*Ma quel che preoccupa maggiormente è il versante 'positivo' del divieto, quello, cioè in cui il Governo pone un obbligo di contenuto... si pone una finalità attiva*[30]" (PIETROPAOLI; SIMONCINI, 2023). Ou seja, não se exige apenas uma abstenção, mas sim uma postura ativa de clara concordância com os padrões éticos e comportamentais impostos pelo ente regulador.

Assim, através dos exemplos apresentados, resta clara a possibilidade de se introduzir em algoritmos de inteligência artificial padrões éticos selecionados pelo desenvolvedor. São o que se pode chamar de *explicit ethical agents:*

---

os fatores subjacentes que contribuem para as diferenças nas taxas de condenação e evitar fazer generalizações amplas sobre qualquer grupo racial ou étnico específico.
Para compreender melhor os padrões de comportamento criminoso, é importante considerar uma série de fatores, incluindo condições socioeconômicas, acesso à educação e recursos, fatores históricos e culturais, além de examinar dados e pesquisas de uma variedade de fontes, a fim de desenvolver uma compreensão mais completa e apropriada do problema. (tradução nossa)

[30] Mas o que mais preocupa é o lado 'positivo' da proibição, ou seja, em que o Governo impõe uma obrigação de conteúdo… estabelece um propósito ativo. (tradução nossa)

> *agents that can identify and process ethical information about a variety of situations and make sensitive determinations about what should be done. When ethical principles are in conflict, these robots can work out reasonable resolutions*[31] (MOOR, 2009, p. 12).

Há necessidade, contudo, de se estabelecer limites e parâmetros que serão utilizados neste tipo de treinamento dos *softwares*, evitando a ocorrência de violações de direitos humanos quando de sua utilização. Mas, se aparentemente há solução que impeça a adoção de vieses discriminatórios, quais seriam os desafios à sua implementação?

Powers e Ganascia destacam:

> The first difficulty comes from modeling deontic reasoning, that is, reasoning about obligations and permissions. The second is due to the conflicts of norms that occur constantly in ethical reasoning. The third is related to the entanglement of reasoning and acting, which requires that we study the morality of the act, per se, but also the values of all its consequences[32]. (POWERS; GANASCIA, 2020, p. 40)

Assim, o desafio seria criar uma máquina capaz de decidir, suprindo todas as características apontadas como dificuldades, ou seja, que fosse capaz de entender conceitos relacionados a obrigações e permissões, proceder com análises em casos de conflitos de normas éticas e entender as consequências precisas de seus atos.

Importa apontar que, independentemente do ponto de vista ético adotado, é necessário que ele seja "ensinado" à máquina. Nesse contexto, se torna imperioso observar que o agir ético é variável de acordo com a sociedade e suas tradições. Assim, para além de ser um conjunto de regras, os princípios éticos não podem ser deduzidos por uma simples observação genérica. Para ilustrar isso, foi realizado um estudo intitulado "*Moral Machine Experiment*", em que se analisou mundialmente atitudes relacionadas a carros autônomos e dilemas morais em eventuais cenários de colisão (se, por exemplo, seria preferível colidir com uma mulher grávida ou um idoso).

---

[31] agentes que podem identificar e processar informações éticas sobre uma variedade de situações e tomar decisões sensíveis sobre o que deve ser feito. Quando os princípios éticos estão em conflito, estes robôs podem chegar a soluções razoáveis. (tradução nossa)

[32] A primeira dificuldade vem da modelagem do raciocínio deontológico, ou seja, da racionalização baseada em obrigações e permissões. A segunda se deve aos conflitos de normas que ocorrem constantemente nos conflitos éticos. A terceira está relacionada com o entrelaçamento entre raciocínio e ação, o que exige que estudemos a moralidade do ato em si, mas também os valores de todas as suas consequências. (tradução nossa)

No referido estudo, Award *et al.* (2018) implementaram e aplicaram um questionário em 233 países, que possuía a seguinte dinâmica:

> Accident scenarios are generated by the Moral Machine following an exploration strategy that focuses on nine factors: sparing humans (versus pets), staying on course (versus swerving), sparing passengers (versus pedestrians), sparing more lives (versus fewer lives), sparing men (versus women), sparing the young (versus the elderly), sparing pedestrians who cross legally (versus jaywalking), sparing the fit (versus the less fit), and sparing those with higher social status (versus lower social status). Additional characters were included in some scenarios (for example, criminals, pregnant women or doctors)[33]. (AWARD *et al*., 2018, p. 60)

De acordo com as respostas, e ainda baseado nos dados pessoais dos participantes, os autores puderam estabelecer três “fortes” preferências (preferir vidas humanas, polpar mais vidas e polpar vidas dos mais jovens). Contudo, conseguiram também observar que as decisões éticas tomadas pelos participantes são fortemente influenciadas pelos valores locais, havendo relevante variação em relação aos demais critérios. Assim, restou demonstrado, *a priori*, a dificuldade em se estabelecer um “critério ético universal”.

Além disso, sob esse ponto de vista, é preciso ressaltar que os valores éticos das decisões autônomas precisam de validação social, uma vez que a tecnologia está a serviço do homem e não contrário. Dessa forma, não havendo tal concordância quanto aos valores escolhidos, passa esta a ser descartável. Deve, portanto, haver um consenso entre aquilo que restou estabelecido para a máquina e os conceitos socialmente aceitos.

Assim, o resultado demonstra que há desafios a serem superados do ponto de vista do “ser ético”, na medida em que as máquinas provavelmente precisariam se adaptar as realidades locais de atuação.

3.4. Dos Riscos de Comportamentos Contrários à Ética

[33] Os cenários de acidentes são gerados pela Máquina Moral seguindo uma estratégia de exploração que se concentra em nove fatores: poupar humanos (*versus* animais de estimação), permanecer no curso (*versus* desviar), poupar passageiros (*versus* pedestres), poupar mais vidas (*versus* menos vidas), poupar homens (versus mulheres), poupando os jovens (*versus* os idosos), poupando os pedestres que atravessam legalmente (*versus* travessias imprudentes), poupando os magros (*versus* os gordos) e poupando aqueles com status social mais elevado (*versus* status social mais baixo). Personagens adicionais foram incluídos em alguns cenários (por exemplo, criminosos, mulheres grávidas ou médicos). (tradução nossa)

Importa apontar no presente estudo os riscos que envolvem a adoção de máquinas sem que haja o adequado tratamento relacionado aos padrões éticos a serem utilizados.

Luciano Floridi (2022) destacou cinco principais problemas relacionados à questão: 1) o *shopping* ético; 2) o *bluewashing* ético; 3) o lobismo ético; 4) o dumping ético; e 5) evasão ética.

Sobre o *shopping* ético, o autor destaca a possibilidade de que, diante da ausência de padronização acerca dos princípios éticos que devem ser implementados durante o surgimento de máquinas inteligentes, haja uma "escolha" por parte dos desenvolvedores sobre "qual ética melhor se adapta ao produto", gerando, assim, uma justificativa para o comportamento do seu produto. Escolhe, assim, aquilo que melhor lhe convém. Para evitá-lo, seria necessário o estabelecimento de padrões éticos claros e gerais a serem adotados por todas as empresas.

O *bluewashing* ético diz respeito à aparência ética. Para Floridi, é o:

> malcostume di fare affermazioni infondate o fuorvianti al riguardo (o di attuare misure suerficiali a favore) dei valore etici e dei benefici di processi, prodotti, servizi o altre soluzioni digitali al fine di apparire più etici dal punto di vista digitale di quanto non si sia effetivamente[34] (FLORIDI, 2022, p. 95)

Neste ponto, a empresa procura se mostrar ética quando, na verdade, não o é (ou é apenas superficialmente). Se concentra, dessa forma, na publicidade, buscando demonstrar ao consumidor a adoção de padrões éticos sem efetivamente adotá-los. Para o autor (2022), somente a efetiva transparência quanto ao desenvolvimento e utilização do *software* é que pode evitar a ocorrência deste desvio.

O lobismo ético, destaca Floridi (2022), se relaciona à tentativa por parte das grandes empresas em evitar a normatização de princípios básicos quanto ao desenvolvimento de ferramentas inteligentes. Buscam retardar qualquer iniciativa que vise à regulamentação no intuito de liberar (ou amenizar) a fiscalização quanto à sua inobservância. A solução, por óbvio, é a efetivação de uma regulamentação, tema que será mais bem explicitado em capítulo posterior deste estudo.

Um problema que também se relaciona com a normatização é o *dumping* ético. Tratar-se de:

---

[34] "Má prática de fazer declarações infundadas ou enganosas sobre (ou implementar medidas superficiais em favor de) valores éticos e benefícios de processos, produtos, serviços ou outras soluções digitais, a fim de parecer mais digitalmente ético do que realmente é" (tradução nossa)

> malcostume di (a) esportare attività di ricerca riguardo a processi, prodotti, servizi o altre soluzioni digitali, in altri contesti o luoghi (per esempio, da organizzazioni europee al di fuori della UE) in modi che sarebbero eticamente inaccettabili nel contesto o luogo di origine; e di (b) importare i risultati di tali attività di ricerca contrarie all'etica[35]. (FLORIDI, 2022, p. 99)

É, portanto, uma conduta em que busca o desenvolvedor aproveitar-se de normas éticas mais brandas de determinado país, levando a ele etapas de produção na busca de menor fiscalização. Após, retornaria com o produto feito, camuflando as deficiências éticas no desenvolvimento. Somente o estabelecimento de padrões éticos mundiais poderia evitar a ocorrência desta falha.

Por fim, o último risco consistiria numa evasão ética. Ou seja, o desenvolvedor busca reduzir a incidência de questões éticas para evitar as obrigações dela decorrentes. A ausência de ética na inteligência artificial causa a adoção de decisões muitas vezes injustas que podem violar direitos de grupos já historicamente marginalizados, exatamente o tema do presente estudo.

3.5. Princípios Éticos

A grande maioria das discussões atuais acerca da ética no uso de inteligência artificial foca bastante na sua necessidade, bem como na forma de implementação.

Diante do quadro de rápido desenvolvimento, os regramentos surgidos – sejam os confeccionados por entes públicos, seja por entes privados – se fundamentam precipuamente em princípios, eis que, como cláusulas abertas que são, possibilitam o estabelecimento de padrões gerais que devem ser aplicados ao desenvolvedor no momento da criação do *software*.

Contudo, é preciso salientar a enorme quantidade de princípios que podem ser encontrados nas mais diversas iniciativas pelo mundo. Conforme indica Floridi (2022), mais de 160 diferentes já haviam sido estabelecidos no ano de 2020. Nesse sentido, o autor destaca que tal quantidade pode interferir no próprio desenvolvimento de regras e padrões, seja pela dificuldade de adequação dos desenvolvedores a uma grande quantidade de comandos, seja pela diferença entre os diversos documentos.

---

[35] "Má prática de (a) exportar a atividade de pesquisa quanto aos processos, produtos, serviços ou outras soluções digitais, para outros contextos ou locais (por exemplo, de organizações europeias fora da UE) de formas que seriam eticamente inaceitáveis no contexto ou local de origem; e (b) importar os resultados de tais atividades de pesquisa antiéticas" (tradução nossa)

Floridi (2022) destaca os seis principais documentos internacionais que se dispuseram a estabelecer em quais princípios deveriam se pautar os desenvolvedores: Os princípios de Asilomar para a Inteligência Artificial (Future of Life Institute), A Declaração de Montreal para uma Inteligência Artificial Responsável (Universidade de Montreal), *Ethically Alligned Design: A Vision for Prioritizing Human Wellbeing With Autonomous and Intelligent Systems (Institute of Electrical and Electronics Engineers - IEEE)*, Declaração sobre Inteligência Artificial, robótica e sistemas autônomos *(European Group on Ethics in Sciences and New Technologies – EGE), AI in the UK: ready, willing and able? (House of the Lords – United Kingdom)* e *Partnership on IA (Partnership on IA Organization).*

Analisando somente estes diplomas, o autor aduz a existência 47 princípios, salientando que as diferenças residem principalmente em questões linguísticas, mas há uma clara coerência entre os comandos estabelecidos.

Floridi (2022) conclui ao estabelecer como base principal para desenvolvimento de *softwares* inteligentes aqueles princípios já utilizados na bioética, com acréscimo de um último: beneficência, não maleficência, autonomia, justiça e explicabilidade.

3.5.1. Princípio da Beneficência

A beneficência diz respeito ao objetivo da tecnologia, ou seja, ela não pode ser criada senão em benefício da humanidade. Logo, há uma relação direta entre o proveito humano e a própria existência da inteligência artificial. É dizer: se não for feita em benefício da humanidade, não deve existir.

A partir deste ponto, pode-se sugerir que tipo de abordagem deve ser incluída quando do desenvolvimento de ferramentas inteligentes, no intuito de se estabelecer um padrão universal de benefício à coletividade. Mas, afinal, qual seria esse parâmetro?

Considerando a grande diferença cultural entre os povos, fato inclusive demonstrado pelo antes citado no "*Moral Machine Experiment*", torna-se difícil estabelecer um padrão único mundial para aceitação integral. Contudo, tendo em vista o alcance da internet e a inexistência de barreiras geográficas entre os países, o estabelecimento de critérios de observância ética foi sendo exigido para o cumprimento mínimo de padrões seguros no uso de máquinas em processos decisórios automatizados.

Assim, em que pese a ausência de unanimidade, os direitos humanos internacionais têm sido estabelecidos com certa frequência como parâmetro aceitável quando do desenvolvimento de programas inteligentes (YEUNG; HOWES; POGREBNA, 2020).

Yeung, Howes e Pogrebna (2020) ressaltam como ponto positivo para adoção dos direitos humanos como padrão para desenvolvimento de softwares a adaptabilidade e dinâmica das decisões judiciais. Para os autores:

> The dynamic and evolving corpus of judicial decisions can help elucidate the scope of justified interferences with particular rights in concrete cases, offering concrete guidance to those involved in the design, development, and implementation of AI systems concerning what human rights compliance requires[36]. (YEUNG, HOWES, POGREBNA, 2020, p. 83)

Assim, fundamentado em um padrão internacional de direitos humanos, os desenvolvedores conseguem, de uma forma geral, estabelecer sob quais normas devem ser desenvolvidos os algoritmos, possibilitando certa uniformização de exigência.

Regulamentações neste sentido obedecem a uma obrigação estatal de proteção dos direitos humanos de seus cidadãos. Assim restou explicitamente estabelecido desde 2009 pelo *Comitee on Economic, Social and Cultural Rights* da ONU: *In addition to refraining from discriminatory actions, States parties should take concrete, deliberate and targeted measures to ensure that discrimination in the exercise of Covenant rights is eliminated (ONU, 2009, on-line)*[37].

Além disso, há na União Europeia uma obrigação, por parte das empresas, de que suas condutas devem necessariamente respeitar os direitos humanos, conforme estabelecido na *United Nations Guiding Principles on Business and Human Rights*. Dessa forma, ao estabelecer tal obrigação de forma específica, o Bloco reforça a necessidade de que as empresas tenham sempre em mente, no exercício de suas atividades, a proteção dos direitos humanos, incluindo, aqui, o desenvolvimento de

---

[36] "O *corpus* dinâmico e evolutivo de decisões judiciais pode ajudar a elucidar o escopo das interferências justificadas em direitos específicos de casos concretos, oferecendo orientações concretas aos envolvidos na concepção, desenvolvimento e implementação de sistemas de IA relacionados ao cumprimento dos requisitos de proteção dos direitos humanos". (tradução nossa)

[37] Além de se absterem de ações discriminatórias, os Estados Partes devem tomar medidas concretas, deliberadas e direcionadas para garantir que a discriminação no exercício dos direitos do Pacto seja eliminada (tradução nossa) https://www.refworld.org/docid/4a60961f2.html. Parágrafo 36

softwares de inteligência artificial. Surge, neste sentido, o conceito de *human rights due diligence*, que foi definido pela ONU como:

> Human rights due diligence is intended to help an enterprise know and show that it respects human rights throughout its operations and over time, including when there are changes in its operations or operating contexts. It therefore requires ongoing or iterative processes, rather than a one-off undertaking, except where those operations and contexts do not change significantly (ONU, 2021, on-line)[38].

Resta claro, portanto, a necessidade de se estabelecer uma obrigação, por parte de todos os envolvidos no processo, de colocar os direitos humanos em primeiro lugar no desenvolvimento das atividades empresariais.

3.5.2. Princípio da Não-Maleficência

Por outro lado, o princípio da não maleficência, embora possa ser confundido com o anterior, diz respeito ao bom uso dos programas, notadamente quanto à privacidade e segurança dos usuários. Logo, os *softwares* devem ser criados em benefício dos seres humanos e o seu uso irregular não pode expor ou causar dano a direitos dos destinatários.

A premissa estabelecida por esse princípio ultrapassa questões éticas de certo e errado. Na verdade, o que se busca é o estabelecimento de uma ordem comum a todas as máquinas de inteligência artificial, impedindo-as de agir deliberadamente com o intuito de se causar dano a seres humanos.

3.5.3. Princípio da Autonomia

Quanto ao princípio da autonomia, este encontra respaldo no equilíbrio entre a capacidade de decisão da máquina e do homem. Aqui residem as discussões acerca do que pode ser delegado a uma decisão automatizada e em que medida, respeitando, sempre, a capacidade humana de revisão dos atos do *software*.

---

[38] A devida diligência em direitos humanos destina-se a ajudar uma empresa a conhecer e demonstrar que respeita os direitos humanos ao longo das suas operações e ao longo do tempo, incluindo quando há mudanças nas suas operações ou contextos operacionais. Requer, portanto, processos contínuos ou iterativos, em vez de um empreendimento único, exceto quando essas operações e contextos não mudem significativamente. (tradução nossa)

Neste ponto também importante abordar questões relacionadas à própria responsabilidade da máquina.

Ainda que necessárias maiores discussões, é possível se falar, diante de uma grave violação de direitos humanos, em “culpa do computador”?

Tal questionamento surgiu diante da gradativa passagem da máquina de meio para agente. Explica Simoncini:

> È un mezzo quell’elemento in un corso di azioni che contribuisce a causare un evento, ma che non esercita alcuna autonomia o discrezione. È invece un soggetto agente colui che avvia tale corso di azioni causando l’evento. La tecnologia, così come la tecnica, hanno sempre fatto parte della categoria dei mezzi; ovverosia degli strumenti posti nelle mani di un soggetto agente, per consentirgli di ottenere un determinato risultato... ma non v’è dubbio che – fino a qualche anno fa – nessuno avrebbe mai pensato di ritenere responsabile della morte di un uomo un autoveicolo perché “tecnicamente” è stato il mezzo che, colpendo la persona, ne ha causato l’evento morte[39] (SIMONCINI, 2019, p. 67-68)

Passando, pois, em determinadas situações, a máquina a agir como “agente” do dano, e não como o meio, é natural que o questionamento acerca da responsabilidade surja.

Para Ron Chrisley (2020), considerando uma abordagem em que há uma centralidade no ser humano, deve-se ser estabelecida uma ênfase no bem-estar humano, mas também em sua responsabilidade. Para ele, por ser o único agente ético, somente ele pode ser o responsável pelos danos causados, não havendo que se falar em responsabilidade da máquina.

É preciso observar que, constatada uma falha no processo decisório autônomo, em que se observa ofensa à direitos do usuário, pode-se sugerir a responsabilização de diversos participantes da cadeia de produção: o operador, o responsável pelo treinamento da máquina, o programador, o desenvolvedor, o vendedor do *software* ou até o Ente Governamental que licenciou o uso do programa.

---

[39] “Um meio é aquele elemento de um curso de ação que contribui para causar um evento, mas que não exerce qualquer autonomia ou discricionaridade. É, ao contrário, um agente, aquele que inicia este curso de ação, causando o evento. A tecnologia, assim como a técnica, sempre fez parte da categoria dos meios; isto é, ferramentas colocadas nas mãos de um sujeito agente, para lhe permitir obter um determinado resultado... mas não há dúvida de que - até há poucos anos - ninguém jamais teria pensado em responsabilizar um veículo motorizado pela morte de um homem porque “tecnicamente” foi o veículo que, ao atingir a pessoa, causou a sua morte” (tradução nossa)

Longe de buscar esgotar o tema, o princípio tem sua aplicação na medida em que põe em foco a questão da autonomia e responsabilização quando da ocorrência de violações por parte de processos decisórios automatizados.

É importante, pois, que em eventual processo de regulamentação, sejam estabelecidos critérios de responsabilidade, para evitar que o usuário se veja desamparado quanto ao ressarcimento de danos sofridos.

3.5.4. Princípio da Justiça

O princípio da justiça abriga o direito do usuário em ter um tratamento igualitário, livre de discriminações. Deve ser resguardado, portanto, a capacidade de implementação de fatores justos e previamente estabelecidos, para que não ocorram fenômenos relacionados ao tendenciamento das decisões.

Em que pese a discussão sobre o justo e seu conteúdo, brevemente abordada no início deste capítulo do ponto de vista ético, o Princípio da Justiça deve ser utilizado como um norte, um objetivo a ser alcançado quando do desenvolvimento de modelos decisórios no intuito de se obter a correta resposta às demandas sociais.

As máquinas de inteligência artificial, portanto, devem cumprir o seu objetivo primeiro de assegurar um tratamento igualitário entre aqueles submetidos à sua decisão, impedindo que sejam considerados quaisquer aspectos discriminatórios. É nesse princípio que se fundam os debates acerca dos vieses que será mais amplamente debatido no capítulo 05 do presente estudo.

3.5.5. Princípio da Explicabilidade

Por fim, Floridi (2022) acrescenta o princípio da explicabilidade ao rol trazido da bioética. Considerando a necessidade de se compreender os processos decisórios da máquina, a explicabilidade (ou transparência, como também denominado em alguns conjuntos de normas) surge como fator essencial para que possam os usuários avaliar qualquer transgressão aos outros princípios. Logo, trata-se de mecanismo essencial que possibilita aos órgãos de controle uma fiscalização efetiva quanto a eventuais desvios de “conduta” da máquina.

Pode-se observar que esse princípio é aquele que garante ao usuário o controle dos atos praticados pela máquina, pois na sua ausência, como podem os critérios por

ela utilizados serem aferidos do ponto de vista ético-jurídico? A explicabilidade, dessa forma, traduz-se na possibilidade do cidadão conhecer os critérios utilizados para decisão de importantes aspectos de sua vida. É importante salientar neste ponto que as decisões tomadas pelo ser humano já são absolutamente opacas, não havendo possibilidade de se verificar quais foram realmente as razões que o levaram a tal conclusão. Assim, a transparência do processo decisório da máquina poderia configurar um avanço neste sentido.

Para fins de implementação deste princípio, sugerem Boeing e Rosa:

> Após escritos tais códigos, pode-se fiscalizá-los por meio de auditoria de seus resultados, momento em que são detectados eventuais vieses ou distorções. Assim, uma vez implementado um modelo e verificado que ele penaliza um certo segmento social desproporcionalmente aos demais, sem razão justificada, pode-se pleitear sua alteração ou seu desativamento (BOEING; ROSA, 2020, p. 89)

Assim, a explicabilidade do *software* aumenta na medida em que são fornecidos pelos desenvolvedores os dados e metodologia que foram utilizados para que fosse obtida determinada decisão.

A transparência do funcionamento do *software* serve inclusive para demonstrar a higidez de seu funcionamento. Hoje, considerando a ausência de dados relacionados ao processo que levou à determinada decisão, há uma tendência em se basear apenas pelos resultados para se decidir acerca de eventual tratamento desigual, o que pode depor contra a qualidade e justiça do algoritmo.

Neste sentido, a Corte Suprema dos Estados Unidos reconheceu que houve discriminação indireta no acesso à moradia por parte de programa do governo do Texas que se utilizava de um *software* para escolha dos beneficiários (Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc., 576 U.S. 519). Analisando a decisão, Giacomelli (2019) destacou que a ausência de transparência quanto aos critérios utilizados pelo algoritmo exigiu que fossem analisados apenas os efeitos produzidos pela sua utilização o que pode, como dito, ser causa de distorções e imprecisões acerca do funcionamento do programa utilizado.

É bem verdade que há uma preocupação por parte da indústria quanto às questões relacionadas à proteção de propriedade intelectual ou até a própria compreensão dos dados por leigos. Tais questões podem ser resolvidas, por exemplo,

por auditorias externas, que garantiriam a supervisão imparcial e sem violações de direitos dos criadores.

Em que pese a aparente simplicidade da medida, Izdebski *et al* (2019) indicam que os países que se utilizam da máquina em processos decisórios não fornecem acesso ao código fonte do modelo. Considerando este fato, para Zavrsnik, há violação do devido processo legal, pois "*in order to ensure effective participation in a trial, the defendant must also be able to challenge the algorithmic score that is the basis of his or her conviction*[40]" (ZAVRSNIK, 2020, p. 577). Para sustentar tal afirmação, o autor compara a questão da opacidade do modelo com as testemunhas anônimas, alegando, em suma, que em termos processuais "*some degree of disclosure is necessary in order to ensure a defendant has the opportunity to challenge the evidence against him or her and counterbalance the burden of anonymity*[41]" (ZAVRSNIK, 2020, p. 577).

Para solução desta problemática, sugere a adoção de níveis de transparência, levando-se em conta o tipo, objetivo e o próprio receptor da informação. Nesse sentido, estabelece Nicholas Diakopoulos:

> Different presentations of transparency information can be produced for different audiences and linked into a multilevel "pyramid" structure of information, which progressively unfolds with densed and more detailed transparency information the further any giver stakeholder wants to drill into it[42] (DIAKOPOULOS, 2020, p. 204)

A sugestão apresentada pelo autor merece destaque na medida em que possibilita a visualização dos dados, de forma mais ou menos técnica, de acordo com a sua capacidade de compreensão ou a própria autorização legal para tanto. Tal mecanismo possibilita, por exemplo, que sejam resguardados dados relacionados à propriedade industrial em relação aos concorrentes, mas estes sejam facultados à órgãos de fiscalização, por exemplo, onde o sigilo comercial estaria mais bem resguardado.

---

[40] "Para garantir efetiva participação em um julgamento, o acusado deve também ser capaz de desafiar o score do algoritmo que é a base de sua convicção" (tradução nossa)

[41] "... algum grau de divulgação é necessário para garantir que o acusado tenha a oportunidade de desafiar a evidência contra si e contrabalancear o ônus do anonimado" (tradução nossa)

[42] "Diferentes apresentações de informações sobre transparência podem ser produzidas para diferentes públicos e vinculadas a uma estrutura de informações em "pirâmide" multinível, que se desdobra progressivamente com informações de transparência mais densas e detalhadas, à medida que qualquer parte interessada quiser nela se aprofundar." (tradução nossa)

Contudo, há diversas questões que precisam ser pensadas quando do estabelecimento de limites à transparência. Nicholas Diakopoulos (2020) destaca a necessidade de cuidados relacionados à manipulação de funcionamento do *software*. O fornecimento de muitos dados poderia impactar no uso por parte de usuários mais esclarecidos, possibilitando uma manipulação dos resultados através de um comportamento dissimulado. Além disso, questões relacionadas aos dados pessoais dos usuários são levantadas, para que não sejam fornecidos dados pessoais no processo de transparência interna.

Assim, a solução apresentada, de um escalonamento de acesso aos dados relacionados à capacidade e autorização do agente, poderia mitigar tais problemas e possibilitar um controle maior por parte dos envolvidos.

A explicabilidade, portanto, passa a ser fator essencial para garantir a idoneidade do programa, sem a qual eventuais violações jamais poderão ser evidenciadas.

3.6. Insuficiência da Abordagem Principiológica

Como observado, apesar de se poder chegar a um certo denominador comum quanto aos princípios centrais que devem nortear o desenvolvimento de *softwares* de tomada de decisão, há uma clara dificuldade em se estabelecer um controle principiológico pelas dificuldades já demonstradas, como a própria definição do que seria ético.

Além disso, Susskind faz uma ponderação acerca das “condutas antiéticas”:

> In public debate and social policy-making, these moral objections will need to be balanced against anticipated benefits such as everyday notions of affordability, convenience, speed, as well as weightier principles of justice, not least distributive justice[43]. (SUSSKIND, 2019, p. 291-292).

Assim, tendo em vista estas questões levantadas por Susskind e analisados estes diversos aspectos teóricos acerca da ética na tomada de decisões e o quão complexas podem ser as situações a que a inteligência artificial pode ser instada a decidir, é possível se concluir pela suficiência de diretrizes éticas no assunto? As

43 “No debate público e na formulação de políticas sociais, essas objeções morais precisarão ser equilibradas com os benefícios observados, como noções cotidianas de acessibilidade, conveniência, rapidez, bem como princípios mais pesados de justiça, principalmente justiça distributiva.” (tradução nossa)

abordagens autorregulatórias são realmente capazes de impedir o surgimento de violações a direitos fundamentais?

Nos capítulos seguintes será demonstrada a tendência de regulamentação mundial e como ocorreu o surgimento de novas formas de vitimização, para que seja possível analisar a situação local e a necessidade de uma regulamentação no Brasil e no mundo.

4. ESTUDO COMPARADO SOBRE A REGULAMENTAÇÃO DA INTELIGÊNCIA ARTIFICIAL

Como já exposto, a inteligência artificial é um fenômeno global. Em todo o mundo, *softwares* de imenso poder de processamento estão influenciando a vida das pessoas e, diante dos impactos positivos e negativos, os Estados têm visto a necessidade de regulamentar a matéria para evitar que abusos sejam cometidos.

Diante das questões levantadas nos capítulos anteriores, é possível depreender que existe a necessidade de uma supervisão acerca do uso de *softwares* de inteligência artificial na tomada de decisões que podem afetar direitos fundamentais dos cidadãos e que tal controle deve ser ainda mais rígido quando se trata de sua utilização no Direito Penal. Tal supervisão, como será demonstrado, pode ocorrer de várias formas, seja com um regulamento mais permissivo ou restritivo.

Contudo, não se trata de um consenso. As *big techs* argumentam que a regulamentação poderia limitar o desenvolvimento da tecnologia, impedindo o surgimento de ferramentas que levam a melhoria na qualidade de vida dos cidadãos. Além disso, John Frank Weaver (2018) pontua que tais empresas buscam uma regulamentação posterior, utilizando como exemplo o sucesso de outras tecnologias que se aperfeiçoaram de forma satisfatória porque o Governo não as regulamentou prematuramente, a exemplo do e-mail, *browsers* de internet, *websites*, blogs e mídias sociais.

Ainda que existam tais posicionamentos, é papel do Estado impor políticas públicas que possam proteger o seu povo de atos discriminatórios e atentatórios contra direitos constitucionalmente garantidos. A situação relacionada à inteligência artificial tem se mostrado diametralmente oposta a das tecnologias citadas, principalmente diante das evidências de graves violações de direitos humanos. Além disso, a regulamentação não é um empecilho ao desenvolvimento da tecnologia se feita corretamente.

Mesmo tendo em vista a dificuldade de se fixar balizas e regras quanto ao tema em questão, há muito se anseia por uma regulamentação e estabelecimento de paradigmas a serem seguidos pelos desenvolvedores e usuários da inteligência artificial, harmonizando, assim, a possibilidade de desenvolvimento da tecnologia com a proteção do cidadão.

A regulamentação da inteligência artificial, pois, compõe a hipótese deste estudo como forma de garantir o usufruto dos direitos dos cidadãos, sem impedir que tal tecnologia se desenvolva e apresente melhorias na qualidade de vida de todos. Com base nisso, se faz necessário analisar o que já existe no âmbito regulatório em relevantes atores mundiais para, ao final, se atingir a conclusão acerca do tema.

Dessa forma, para que possamos alcançar o objetivo do presente estudo, que visa a analisar como o desenvolvimento dos algoritmos pode ser projetado para que não sejam criadas novas formas de vulnerabilização de minorias, se faz necessário lançar mão do direito comparado, colacionando as iniciativas dos Estados e Órgãos que buscam a regulamentação da matéria.

O direito comparado é a ciência do conhecimento dos “direitos”. Consiste no estudo comparativo entre diversos sistemas jurídicos (Sacco, 1992) e pode ser utilizado para atingir vários objetivos.

Acerca das possibilidades que são geradas através do uso do direito comparado, esclarece Roberto Scarciglia:

> La comparazione permette di penetrare, attraverso la conoscenza dei formate o componenti degli ordinamenti giuridici, e dele intercinnessioni che li caratterizzano, profili, sai positivi che negativi, degli ordinamenti stessi. In tale prospettiva, la comparazione diviene uno straordinario strumento epistemológico. Il diritto comparato svolge, infatti, il compito di fare circolare i prodotti dela scienza giuridica e di farla divenire Internazionale[44]. (SCARCIGLIA, 2018, p. 46)

Ainda aponta o mesmo autor que o auxílio para fins de política legislativa é considerada uma das principais funções do direito comparado. Na verdade, a sua utilização para edição de novas legislações é uma prática antiga, que data do próprio surgimento das leis escritas. Segundo Smits (2019), a Lei das Doze Tábuas foi influenciada pelas incursões romanas a outros territórios, da mesma forma que o Código de Hamurabi foi, presumivelmente, baseado em leis que estavam em vigor no Oriente Próximo.

O transplante de legislação, termo utilizado por Alan Watson, pode se apresentar como um facilitador para aplicação das regras que encontram maior

[44] A comparação permite penetrar, pelo conhecimento dos formatos ou componentes dos sistemas jurídicos, e das interligações que os caracterizam, perfis, tanto positivos como negativos, dos próprios ordenamentos. A partir dessa perspectiva, a comparação torna-se uma ferramenta epistemológica extraordinária. O direito comparado cumpre, de fato, o dever de fazer circular os produtos da ciência jurídica e torná-la internacional (tradução nossa).

consonância com o exigido pela comunidade internacional para proteção de seus cidadãos. Por tratar-se de uma tecnologia que extrapola os limites territoriais dos países, a estipulação de regras harmônicas relacionadas à inteligência artificial, na medida do possível, pode significar um ganho social para todos. Para o autor, tal forma de empréstimo constitui para o processo legislativo "*the most fertile source of development*[45]" (WATSON, 1974, p. 95).

Outro importante fator apontado por Alan Watson diz respeito à necessidade de se reconhecer a *expertise* de legisladores estrangeiros. Serão analisadas legislações de grandes países que se encontram no topo do desenvolvimento da tecnologia analisada. Estes, por consequência, estão mais aptos a apontar falhas na construção destes algoritmos do que países que apenas consomem os serviços, por exemplo. Sobre tal fato:

> ... law like technology is very much the fruit of human experience. Just as very few people have thought of the wheel yet once invented its advantages can be seen and the wheel used by many, so important legal rules are invented by a few people or nations, and once invented their value can readily be appreciated, and the rules themselves adopted for the needs of many nations[46] (WATSON, 1974, p. 100)

É de se analisar, ainda, o caráter econômico relacionado à regulamentação da inteligência artificial. Sendo um campo em pleno desenvolvimento por grandes empresas, que visam precipuamente ao aumento de lucros, o fator econômico deve ser sopesado, para se evitar que a norma imponha ônus desnecessários que levem à saída das companhias para outros países.

Nesse sentido, é importante perceber a importância de se estabelecerem regras razoavelmente harmônicas em todo o mundo, garantindo, assim, a proteção de todos os cidadãos contra violações aos seus direitos fundamentais, bem como apresentando exigências uniformes a todas as empresas.

A respeito desta relação entre o aspecto econômico e o transplante de normas no direito comparado:

---

[45] A mais fértil fonte de desenvolvimento (tradução nossa).

[46] ... direito como a tecnologia é muito fruto da experiência humana. Assim como muito poucas pessoas pensaram na roda, uma vez inventada, suas vantagens podem ser vistas e a roda usada por muitos, regras legais importantes são inventadas por algumas pessoas ou nações e, uma vez inventadas, seu valor pode ser facilmente apreciado, e o próprias regras adotadas para as necessidades de muitas nações. (tradução nossa)

> In other contexts, the economic aspect of the transplanted statute is premised on the idea that, in order to reap the specific benefits of a given foreign statute, it is necessary to enact a similar regime (as was the case with the Copyright Term Extension Act, enacted in response to the reciprocity principle stipulated in European law)[47]. (BARAK-EREZ, 2014, p. 20)

Assim, ao se estipular o uso de regulamentação sobre o tema da inteligência artificial, a utilização do direito comparado se mostra como ferramenta eficaz para facilitar a edição de uma lei em consonância com o padrão mundial, estabelecendo proteção ao mesmo tempo em que se garante um tratamento isonômico às empresas que desenvolvem tais softwares.

Por essa razão, segue análise acerca das medidas tomadas, ou em fase de discussão, nos principais atores no cenário mundial no tocante à regulamentação da inteligência artificial.

4.1. China

A China foi o primeiro grande país a aprovar uma lei clara e abrangente relacionada à regulamentação no uso de inteligência artificial, tendo entrado em vigor no dia 01 de março de 2022.

Segundo o texto, a lei foi aprovada com o objetivo de padronizar os serviços de "recomendação algorítmica", resguardar os valores Socialistas, a segurança nacional e o interesse público, além de estimular o desenvolvimento saudável dos serviços de informação na internet (CREEMERS; TONER, WEBSTER, 2022).

O legislativo chinês preferiu utilizar o termo "tecnologia de recomendação algorítmica", mas é possível depreender que se trata do mesmo tipo de software denominado *decision-maker* nos Estados Unidos e Europa. São exatamente os programas que sugerem ao usuário decisões a serem tomadas com base em algoritmos preditivos e uso da *Big Data*.

A lei exigiu que as recomendações algorítmicas:

> Shall abide by laws and regulations, observe social morality and ethics, abide by commercial ethics and professional ethics, and respect the principles of fairness and justice, openness and transparency, science and reason, and

[47] Em outros contextos, o aspecto econômico da lei transplantada tem como premissa a ideia de que, para colher os benefícios específicos de uma determinada lei estrangeira, é necessário decretar um regime semelhante (como foi o caso da Lei de Extensão do Período de Direitos Autorais, promulgada em resposta ao princípio da reciprocidade estipulado na legislação europeia). (tradução nossa)

> sincerity and trustworthiness[48]. (CREEMERS; TONER, WEBSTER, p. 4, 2022)

Trata-se de um claro conjunto de referenciais éticos, a cujos preceitos devem estar atrelados todos os envolvidos no desenvolvimento dos *softwares* desta natureza. Infelizmente, por tratar-se de conceitos abertos, a sua limitação pode tornar-se meramente formal, sem que haja possibilidade de se forçar uma empresa a adotar padrões éticos na utilização de seu algoritmo.

Há no artigo 6º uma proibição genérica de que os serviços de recomendação algorítmica não podem utilizar seus programas em atividades que causem danos a segurança nacional ou interesse público, abalos a ordem econômica ou social, violar direitos e interesses de outras pessoas ou outros atos proibidos por leis e regulamentações administrativas. Trata-se de outra proibição absolutamente generalista, que impede o seu uso de forma coerente na proteção de direitos fundamentais diante da utilização da inteligência artificial. Conceitos como “abalo da ordem social” ou até “violar direitos e interesses de outras pessoas” são bastante vagos e levam a uma possível inaplicabilidade do dispositivo, notadamente pela amplitude de interpretações que podem ser extraídas.

O texto avança um pouco na questão relacionada aos direitos de proteção, determinando que o usuário possa escolher não ter suas características individuais levadas em consideração em eventual decisão ou simplesmente “desligar” o algoritmo (art. 17). Além disso, os usuários devem ser notificados de forma clara acerca da construção do programa, publicizando seus princípios básicos, mecanismos de operação etc.

O texto dispõe, ainda, de poderes reguladores, autorizando que haja estabelecimento de reprimendas por parte daqueles que violaram de alguma forma as disposições ali contidas. O art. 21 do normativo direciona o controlador a outros regramentos do país para estabelecimento da sanção, mas autoriza, desde já, que no caso de ausência de disposição adequada, a autoridade tem o poder de advertir, determinar correções ou suspender a autorização de produção.

Vê-se que esta legislação deu um passo em direção à transparência, item essencial quanto ao uso dos algoritmos de tomada de decisão, eis que se trata de um

---

[48] Deve cumprir as leis e regulamentos, observar a moralidade e a ética, respeitar a ética comercial e a ética profissional e respeitar os princípios de equidade e justiça, abertura e transparência, ciência e razão, sinceridade e confiabilidade. (tradução nossa)

direito de todos e uma garantia de que qualquer violação de direitos poderá ser eventualmente demonstrada.

Os artigos ressaltados demonstram que, apesar de buscar apresentar uma legislação abrangente, o texto peca pelo excesso de termos genéricos e abertos, podendo gerar a sua própria inexequibilidade. Diversos são os princípios inseridos que, de tão abrangentes, inviabilizam o seu uso como garantia dos cidadãos. Seria necessário maiores especificações acerca dos princípios elencados para que pudesse tal lei servir de exemplo para os demais países.

Por outro lado é importante observa que a configuração do Governo Chinês possibilita que sejam estabelecidas medidas severas, em total benefício dos cidadãos, sem maiores riscos de insubordinação ou descumprimento.

4.2. Estados Unidos

O cenário americano merece análise, considerando que o país exerce papel de liderança mundial em diversos aspectos, se colocando também como um dos maiores produtores de tecnologia de dados.

O panorama americano relacionado ao tema parece já ter sido desenhado há muitos anos, ainda que indiretamente. Frank Pasquale (2015) aponta a busca por transparência como parte de um longo caminho Americano que teve início em 1910, com a lei de publicidade dos atos de doação de campanhas. Com a chegada da internet, teria surgido uma nova era de transparência, mas que estaria sendo ofuscada pelo avanço não-regulamentado das tecnologias de inteligência artificial.

É possível perceber que desde o início do uso de tecnologias de inteligência artificial, os Estados Unidos se mostraram favoráveis, não havendo, à época, maiores discussões acerca dos danos que poderiam ser causados pela utilização indiscriminada. O Congresso Americano promulgou o *Sentencing Reform and Corrections Act* de 2015, tendo determinado expressamente, em sua Seção 203, que o Departamento de Justiça deveria desenvolver um sistema de análise de risco pós-sentença para avaliar reincidência (CHUCK, 2016). Observa-se, portanto, que desde 2015 há um interesse americano em utilizar o sistema, inclusive em uma das áreas mais sensíveis – o direito penal – sem que houvesse maiores preocupações acerca da proteção de direitos individuais.

Em 2020 foi editado o *National Artificial Intelligence Iniciative Act* (United States 116th Congress, 2020) que estabeleceu uma iniciativa com o propósito, em suma, de assegurar a liderança americana no desenvolvimento de *softwares* de dados, liderar o mundo no tocante à confiabilidade destes sistemas, preparar o país para a integração da força de trabalho com tais programas e coordenar as pesquisas na área.

O referido diploma legal não estabeleceu diretrizes específicas quando à proteção dos cidadãos americanos quando do uso de algoritmos inteligentes, mas elaborou uma estrutura governamental para que os projetos relacionados à tecnologia possam ser supervisionados de forma centralizada. Não houve, portanto, uma preocupação em estabelecer limites e proibições ao desenvolvimento tecnológico.

A seção 5104 do referido ato estabeleceu a necessidade de subcomitês se responsabilizarem acerca das questões relacionadas com o desenvolvimento de inteligência artificial para uso na aplicação da lei, notadamente: enviesamento, segurança dos dados, usabilidade e padrões legais (compatibilidade entre o *software* e as garantias individuais).

Apesar de importante, a iniciativa se ateve a organizar uma estrutura administrativa no Governo Americano para gerenciamento das questões relacionadas à inteligência artificial sem, no entanto, estipulá-las.

Em outubro de 2022 foi publicado pelo *White House Office of Science and Technology Policy* um document intitulado *The Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People* (*The White House Office of Science and Technology Policy*, 2023). O documento, que não possui força de lei, tem como objetivo apoiar o desenvolvimento de políticas e práticas de proteção dos direitos civis, além de promover valores democráticos no desenvolvimento de sistemas automatizados.

Ao estipular o seu objeto, o documento estabelece que os princípios nele contidos devem ser aplicados a sistemas automatizados que possuam o potencial de impactar significativamente direitos públicos dos americanos (direitos civis, liberdades individuais e privacidade), oportunidades (garantia de igualdade) e acesso a recursos críticos ou serviços (saúde, segurança, serviço social etc.). Observa-se, portanto, que busca o documento a proteção de garantias individuais, estabelecendo de forma abrangente aqueles bens que devem ser protegidos quando da análise de sistemas automatizados.

De forma mais específica, o *Blueprint for an AI Bill of Rights* estabelece cinco princípios que devem servir de auxílio no desenvolvimento e uso de sistemas automatizados, protegendo assim os direitos individuais dos americanos. Em cada um desses princípios restaram explicitados diversas questões quanto a expectativa de aplicação e ainda como os princípios podem ser postos em prática. Para fins didáticos, será trazida a menção aos princípios e sua explicação de forma concisa.

O primeiro princípio apresentado é o *Safe and Effective Systems*. Este princípio geral estabelece que os sistemas automatizados devem ser seguros e desenvolvidos através de consultas com diversos atores, para fins de garantir que nenhum desses *softwares* seja efetivamente utilizado se identificado o objetivo ou a própria previsibilidade razoável de colocar a segurança da comunidade em risco. Ao contrário, devem ser desenvolvidos para proativamente evitar a ocorrência de tais fatos.

O *Algorithmic Discrimination Protections* se relaciona diretamente com o objeto do presente estudo. Estipula que ninguém deve sofrer discriminação por algoritmos e que os sistemas devem ser usados e desenvolvidos de forma equitativa. Estabelece a necessidade de uso de dados representativos e de avaliações de igualdade como parte do próprio design do sistema.

Data Privacy estabelece a necessidade de proteção dos dados através de mecanismos inseridos no próprio sistema, além da necessidade de se supervisionar o uso dos dados coletados.

Notice and Explanation diz respeito a obrigatoriedade de informação ao usuário de que um sistema automatizado está sendo usado e qual a sua participação no resultado que será obtido. Trata-se da necessidade de informar ao cidadão que a solicitação realizada está sendo submetida a um software inteligente e como ele funcionará para atingir o resultado. Dessa forma, é possível estabelecer parâmetros de controle para eventual identificação de violação de direitos.

E, por fim*, Human Alternatives, Consideration and Fallback*, que estipula a necessidade, quando aplicável, de se instituir uma alternativa humana, a possibilidade do usuário optar por não receber tratamento automatizado. Obviamente este princípio se aplica a situações específicas, que devem ser analisadas individualmente, mas se trata de uma medida de segurança e supervisão, garantindo ao cidadão a participação humana em decisões que afetam seus interesses.

O documento em referência foi editado com base nos estudos mais recentes e tendo como objetivo principal e proteção dos direitos individuais americanos. Como

não possui caráter vinculante, sua aplicação torna-se restrita, mas se utilizado corretamente na legislação que está sendo elaborada pelos Estados Unidos, certamente os interesses dos americanos estarão protegidos.

Em relação à legislação, encontra-se em discussão no Congresso americano o *Algorithmic Accountability Act*. Mesmo que não aprovado, entende-se pertinente a sua análise para fins de comparativo com os demais diplomas legais, notadamente para analisar a sua pertinência com a Blueprint publicada pela *White House* que, como visto, possui ampla gama de sugestões para proteção de direitos individuais.

O *Algorithmic Accountability Act* (United States 117th Congress, 2022) foi apresentado ao Congresso Americano em março de 2022, encontrando-se ainda em discussão na Casa (ainda que renovado na legislatura seguinte). Da leitura do texto, percebe-se que a intenção do legislador foi a de estabelecer formas de controle posterior dos *softwares* de inteligência artificial e não de estabelecer parâmetros gerais a serem seguidos no seu desenvolvimento. Tal encargo foi delegado a *Federal Trade Commission*, que, em até dois anos após promulgação da lei, deve elaborar regulamentação sobre a matéria.

Observa-se que não há preocupação do legislativo americano em estabelecer balizas claras em uma data próxima, tendo autorizado um período de até dois anos para que isso seja feito. Até lá, com a velocidade em que ocorrem mudanças e evoluções tecnológicas, diversos danos já poderão ser constatados.

A forma como se estabeleceu no projeto de lei a análise de impacto dos *softwares*, com sua delegação ao próprio desenvolvedor, preocupando-se somente com o registro das ocorrências, terceiriza a fiscalização e possibilita que as empresas ocultem eventuais falhas e impede que um terceiro independente atue na proteção dos direitos individuais eventualmente violados por decisões automatizadas.

A legislação, portanto, apesar de se propor a identificar parâmetros de responsabilidade algorítmica, falha ao não estabelecê-los, delegando a um órgão que o fará em prazo elastecido. Além disso, não há previsão de uma supervisão independente ou prévia, o que fragiliza ainda mais a segurança dos cidadãos.

Observa-se, portanto, que a legislação apresentada está em descompasso com os rigorosos critérios estabelecidos pelo *Blueprint for na AI Bill of Rights*, não se aproveitando de sua completa abordagem do tema para elaboração de uma lei que efetivamente proteja os cidadãos americanos.

Por outro lado, diante da ausência de regulamentação federal, diversos estados americanos optaram por estabelecer regramento próprio, no intuito de proteger seus residentes.

O Estado do Colorado aprovou, em maio de 2024, uma lei responsável pela ˜Proteção dos Consumidores nas interações com Sistemas de Inteligência Artificial". A norma, além de trazer definições de termos comumente usados como discriminação algorítmica, optou por uma abordagem baseada no risco, em um padrão inspirado pelo normativo proposto – e aprovado – pela União Europeia.

No Estado de Utah, entrou em vigor também em maio de 2024 o *AI Policy Act*, que estipulou diversos compromissos com transparência, além de responsabilização de empresas.

Além dos aprovados, diversos estados estão discutindo a aprovação de legislação protetiva contra efeitos nocivos da inteligência artificial. Infelizmente, não contam com expressa disposição acerca da proteção dos direitos humanos, apesar de buscar garantir isonomia de tratamento para todos submetidos às decisões automatizadas.

4.3. Canadá

O Canadá busca estabelecer parâmetros gerais para o uso de inteligência artificial desde 2018. Neste ano, foram confeccionadas duas declarações que buscavam alinhar estes critérios.

A Declaração de Toronto (The Toronto Declaration, 2018), publicada em maio de 2018, teve como objetivo a proteção do direito de igualdade e de não-discriminação nos sistemas de aprendizagem de máquina. O documento focou na adoção de um sistema centrado nos direitos humanos internacionais, conforme preconizado pela Organização das Nações Unidas. Entre outras questões, importante pontuar que o texto ressalta, com base na legislação internacional, a obrigação dos Estados em promover e proteger os direitos humanos e dos entes privados em respeitá-los.

A Declaração de Montreal (Universitè de Montreal, 2018), também de 2018, teve como foco o desenvolvimento da inteligência artificial. O documento, que buscou estabelecer um parâmetro geral a ser seguido quando da criação de sistemas de IA, estipulou 10 princípios que deveriam ser seguidos pelos desenvolvedores: princípio do bem-estar, do respeito à autonomia, da proteção da intimidade e da vida privada,

da solidariedade, da participação democrática, da equidade, da inclusão da diversidade, da prudência, da responsabilidade e do desenvolvimento sustentável.

Embora as declarações anteriormente citadas não possuam força de lei, já há um movimento legislativo no Canadá para regulamentação relacionada ao tema. O projeto de Lei C-27 foi apresentado em junho de 2022 e dispõe, dentre outras coisas, acerca do *Artificial Intelligence and Data Act* (AIDA) (Canadian Government, 2021), que traz os requisitos para desenvolvimento, uso e fornecimento de softwares de inteligência artificial no país.

Embora indique em seu sumário que o *Artificial Intelligence and Data Act* tem como objetivo a redução dos riscos de danos e vieses relacionados a sistemas de inteligência artificial de "alto impacto", o ato parece ser tímido quanto a requisitos essenciais para consecução de tais objetivos.

Logo no início do texto, o diploma legislativo excepciona sua incidência em relação a órgãos do governo. Tal ponto parece esvaziar grande parte do objetivo da Lei, notadamente por serem atos governamentais aqueles que tem o poder de maior influenciar a vida dos cidadãos do país. Ao excepcionar a incidência do Ato em relação a tais estruturas temos um grande enfraquecimento no cenário de proteção das minorias que, ainda que protegidas da atuação de entes privados, podem ser injustamente prejudicadas pela atuação de *softwares* de inteligência artificial que não tiveram que obedecer a todas as determinações contidas na legislação de regência.

Em consonância com a proposta europeia, o ato também estabelece uma proteção com base no risco, embora deixe para regulamentação posterior os critérios de enquadramento do *software* como tal.

Outro ponto falho da proposta diz respeito ao próprio enquadramento do sistema como de alto risco. Não consta da Lei qualquer mecanismo de controle prévio acerca de tais sistemas, havendo uma delegação ao desenvolvedor para que indique se o programa se enquadra na classificação, bem como uma obrigação para que sejam estabelecidas medidas para identificar e mitigar os riscos de dano e prejudicialidade no seu uso.

Assim, o ato prevê apenas um controle posterior acerca de eventuais danos no sistema, o que só ocorreria após uma ordem ministerial para que seja realizada uma auditoria (pelo próprio desenvolvedor ou por terceiros).

Tal sistema, ao contrário do controle prévio proposto na Europa, possibilita que ocorram danos irreparáveis aos cidadãos do país submetidos às decisões tomadas

por *softwares* de inteligência artificial. Enquanto não identificado, o dano poderá ter ocorrido em detrimento de diversos grupos sem que tenham tido um tratamento igualitário, sem mencionar a possibilidade de sugestionamento em massa, problema já anteriormente comentado.

Acerta, contudo, o ato, ao autorizar que o Ministro responsável possa determinar o encerramento das atividades do *software* caso sejam constatados riscos.

Embora mereça destaque o interesse do Governo Canadense em regulamentar a matéria, a adoção de um sistema de controle posterior não parece ser a melhor para evitar o cometimento de injustiças em detrimento de grupos vulneráveis. Como já destacado, a imposição de ônus maior à determinadas minorias, que já sofrem constantemente com prejuízos ilícitos decorrentes de tratamentos desiguais, não pode ser aceita e devem os Governos tomar as medidas necessárias para que isso seja evitado. Não há que se falar em impossibilidade, considerando que tal modelo foi inclusive proposto pela União Europeia. Neste ponto, há grave falha da legislação proposta pelo Canadá.

4.4. União Europeia

A atuação da União Europeia no que diz respeito à regulamentação da inteligência artificial tem levado o bloco a um papel de relevância e pioneirismo, notadamente pelo empenho em finalmente trazer uma regulamentação geral sobre a matéria.

Ainda em 2013, o Parlamento Europeu, através de seu Comitê de Assuntos Jurídicos, emitiu recomendação no bojo do procedimento nº 2015/2013 (INL), à Comissão de Regras de Direito Civil em Robótica, expressamente identificando as Leis de Asimov como princípios gerais a serem seguidos por *designers*, produtores e operadores de robôs e estabelecendo ainda padrões éticos que vinculam toda a União Europeia no que se refere à inteligência artificial, destacando a necessidade de transparência (EUROPEAN PARLIAMENT, 2017).

Outro relevante marco relacionado ao uso de inteligência artificial no processo decisório foi a edição do Regulamento (UE) 679/2016 do Parlamento Europeu e do Conselho, que trata sobre o tratamento de dados pessoais e a circulação destas informações (*General Data Protection Regulation* – GDPR). De início, na consideração nº 15 é possível ver a menção expressa à incidência do diploma legal

ao tratamento de dados por meios automatizados. Além disso, no tópico 63 é possível observar o direito do cidadão em saber a "lógica subjacente ao eventual tratamento automático dos dados pessoais" (REGULAMENTO UE – 2016/79, tópico 63) o que demonstra uma preocupação com a própria transparência no funcionamento desses *softwares* para proteção dos dados fornecidos.

Um ponto importante constante no considerando n. 71 (disposto no art. 22 da GDPR) diz respeito a proibição expressa de que o titular de dados esteja sujeito a uma decisão que se baseie exclusivamente no tratamento automatizado. Este comando é bastante importante, pois reforça o caráter acessório das ferramentas de inteligência artificial, impedindo que decisões que produzam efeitos jurídicos sejam tomadas de forma exclusivamente eletrônica, sem intervenção humana.

Urge ressaltar que o mencionado dispositivo traz exceções: 1) se o tratamento automatizado for necessário para a celebração ou execução do contrato; 2) se houver autorização pelo direito da União Europeia ou do Estado-membro, desde que estiverem salvaguardados os direitos, liberdades e legítimos interesses do titular dos dados e; 3) houver consentimento explícito do titular dos dados.

Com o Regulamento Geral de Proteção de Dados, a União Europeia sedimentou a garantia de que, via de regra, nenhum cidadão europeu pode ser submetido a uma decisão decorrente de análise automática de seus dados, sem que haja participação humana, o que demonstra a predisposição do bloco em impedir o uso indiscriminado de ferramentas de inteligência artificial, mitigando assim os danos decorrentes do uso da tecnologia sem supervisão.

Já no ano de 2018, o Conselho da Europa editou uma Carta Ética sobre o uso da tecnologia em Sistemas Judiciais, mas, em seu interior, o documento especifica a quem se destinam os princípios ali elencados, abarcando praticamente todos os setores da sociedade:

> Cette Charte s'adresse aux acteurs publics et privés en charge de la conception et du déploiement d'outils et de services d'intelligence artificielle s'appuyant notamment sur le traitement des décisions juridictionnelles et des données judiciaires (apprentissage machine ou toutes autres méthodes issues des sciences de données). Elle concerne également les décideurs publics en charge de l'encadrement législatif ou réglementaire, du développement, de l'audit ou de l'utilisation de tels outils et service.[49]

49 "Esta carta é dirigida aos atores públicos e privados responsáveis pela concepção e desenvolvimento de ferramentas e de serviços de inteligência artificial que envolvam notadamente o processamento de decisões e dados judiciais (aprendizado de máquina ou qualquer outro método derivado da ciência de dados). Refere-se, igualmente, aos agentes públicos responsáveis pelo

> (COMMISSION EUROPEENNE POUR L'EFFICACITE DE LA JUSTICE, 2018, p. 6).

Aduz, também, ressalvas quanto à utilização da tecnologia em matéria penal, ao estabelecer que: "*En matière pénale, leur utilisation doit être envisagée avec les plus extrêmes réserves, afin de prévenir des discriminations sur des données sensibles, en conformité avec les garanties du procès équitable*."[50] (COMMISSION EUROPEENNE POUR L'EFFICACITE DE LA JUSTICE, 2018, p. 6).

É possível observar que há uma indicação de restrição de uso, mas sem uma delimitação específica, o que torna o seu conteúdo abstrato, com pouca aplicação prática.

Estabelece, por fim, cinco princípios que devem ser seguidos na utilização de inteligência artificial em sistemas judiciais: princípio do respeito aos direitos fundamentas, princípio da não-discriminação, princípio da qualidade e segurança, princípio da transparência, neutralidade e integridade intelectual e o princípio do controle pelo usuário.

Em 2020 a Comissão Europeia apresentou o "Livro Branco sobre a inteligência artificial", que teve como objetivo definir as opções políticas que poderiam ser tomadas pela União Europeia para alcançar uma abordagem regulamentar da Inteligência Artificial, promovendo sua adoção, mas também abordando os seus riscos. Convida, através do mesmo documento, os Estados e demais partes interessadas a apresentarem comentários para fins de uma tomada conjunta de decisões sobre o assunto.

O documento contém um grande apanhado sobre o estado de desenvolvimento da inteligência artificial, ressaltando a necessidade de se haver uma abordagem comum na Europa, garantindo assim uma proteção uniforme e a própria unidade do mercado único. Busca, portanto, apresentar as opções para que haja um desenvolvimento confiável da tecnologia, mas seguro para o continente europeu, fortalecendo a confiança dos cidadãos no uso de tais programas.

---

processo legislativo ou regulamentar, de desenvolvimento, auditoria ou uso de tais ferramentas e serviços. " (tradução nossa)

50 "Em matéria penal, a sua utilização deve ser considerada com as mais extremas reservas, de forma a evitar a discriminação de dados sensíveis, de acordo com as garantias de um julgamento justo." (tradução nossa)

Importante destacar que o Livro Branco deixou bastante claro os pontos negativos relacionados ao uso de inteligência artificial, apontando como principais eventuais riscos para direitos fundamentais (incluindo a não-discriminação) e para o regime de responsabilidade.

O documento foi capaz de demonstrar o interesse da União Europeia na regulamentação da matéria, apresentou os riscos relacionados ao uso de tais *softwares* bem como o ponto de vista de especialistas, possibilitando uma discussão com base em dados técnicos e provenientes de entidades especializadas. Este apanhado, em conjunto com os comentários apresentados por terceiros, foi levado em conta para edição da regulamentação final da matéria aprovada pelo Parlamento Europeu.

No início de 2021, a União Europeia iniciou as discussões mais específicas quanto a uma regulamentação abrangente de ferramentas de Inteligência Artificial, tendo submetido à apreciação o *Artificial Intelligence Act*, que estabeleceria regras harmonizadas sobre a matéria para todos os países membros. Somente em maio de 2024, após anos de discussão, o texto foi aprovado.

A edição deste documento veio para demonstrar a insuficiência de simples diretrizes gerais éticas anteriormente editadas, na medida em que somente normas de caráter cogente podem ser capazes de compelir as empresas a cumprirem com os interesses estatais de proteção de seus cidadãos, principalmente se tendo em vista os gastos necessários para sua implementação e adequação.

É importante salientar que a ação legislativa no tocante ao tema foi uma recomendação do próprio Parlamento Europeu, que visava à obtenção de benefícios com o uso de inteligência artificial, mas também garantir a proteção de princípios éticos (Resolução do Parlamento Europeu, de 20 de outubro de 2020, sobre o regime relativo aos aspectos éticos da inteligência artificial, da robótica e das tecnologias conexas).

Em sua exposição de motivos, o documento (União Europeia, 2021) salienta os benefícios econômicos no uso da inteligência artificial, ao mesmo tempo em que ressalva a existência de novos riscos e consequências negativa para a sociedade de uma forma geral. Assim, busca equilibrar o interesse da União em assegurar o desenvolvimento de novas tecnologias e proteger os seus cidadãos, como já abordado no Livro Branco sobre a inteligência artificial apresentado pela Comissão Europeia.

Restou estabelecido que a base de proteção tomada pela legislação seriam os valores e direitos fundamentas da União Europeia, em uma abordagem que vai ao encontro do já preconizado pelos estudos de ética em que a decisão a ser tomada deve sempre ter como parâmetro a proteção dos direitos fundamentais e a centralidade do homem.

A estrutura do regulamento demonstra uma indicação dos objetivos da União Europeia quanto ao tema. No título I há a definição de inteligência artificial; No título II consta uma lista de sistemas de inteligência artificial proibidos de desenvolvimento com base no risco que oferecem; O título III prevê regras específicas para sistemas de alto risco que podem trazer risco a saúde e segurança; o Título IV traz obrigações de transparência para alguns sistemas que representam risco de manipulação; o Título V contribui para criação de sistemas regulatórios de supervisão e governança; o Título VI trata os níveis de governança em nível europeu e nacional; o Título VII trata de normas programáticas sobre criação de database sobre sistemas que impactam em direitos fundamentais; O título VIII traz obrigações de monitoramento para provedores; e os capítulos finais trazem normas finais quanto ao respeito à confidencialidade de informações e dados.

Os objetivos do *Artificial Intelligence Act* são bem específicos e merecem destaque pois conseguem traduzir com precisão a intenção da União em harmonizar os interesses de desenvolvimento da tecnologia com a proteção dos cidadãos:

> - Garantir que os sistemas de IA colocados no mercado da União e utilizados sejam seguros e respeitem a legislação em vigor em matéria de direitos fundamentais e valores da União;
> - Garantir a segurança jurídica para facilitar os investimentos e a inovação no domínio da IA;
> - Melhorar a governação e a aplicação efetiva da legislação em vigor em matéria de direitos fundamentais e dos requisitos de segurança aplicáveis aos sistemas de IA;
> - Facilitar o desenvolvimento de um mercado único para as aplicações de IA legítimas, seguras e de confiança e evitar a fragmentação do mercado (UNIÃO EUROPEIA, 2021, online)

Torna-se claro o objetivo da lei em trazer os interesses das empresas à mesa, evitando que se trave uma batalha desnecessária entre os atores envolvidos. Afinal, não se busca a extinção da tecnologia que, como visto, traz diversos benefícios, mas apenas que a sua utilização não venha a ferir direitos fundamentais dos cidadãos.

Outro ponto crucial da lei diz respeito à abordagem regulamentar baseada no risco. Sobre o tema:

> Entrambi i sistemi di regole, quello ético e quello giuridico, sono stati parametrati sulla base di soglie di accettabilità del rischio dei sistemi di intelligenza artificiale, il che rappresenta di per sé um inédito del modelo europeo, tipicamente costruito come moello giuridico *tout-court* nel quale i valori dell'Unione rappresentano il limite di tenuta del sistema al quale istituzioni e indiviui riferiscono la própria identità "europea"[51] (CATANZARITI, 2022, p. 74)

A aferição do grau risco, portanto, passa a ser elemento essencial para se analisar a própria possibilidade de existência do *software* ou a que eventuais restrições ele estará submetido. Parece ser uma decisão acertada, na medida em que os maiores ônus acabam sendo impostos aqueles desenvolvedores de programas com maior possibilidade de impacto na vida dos cidadãos.

Importante ressaltar que para edição da proposta de regulamentação foi lançada uma consulta pública, onde os principais atores envolvidos puderam se manifestar acerca do texto. De um modo geral, como mencionado na exposição de motivos, houve consenso entre os interessados acerca da necessidade de regulamentação. Houve, ainda, adesão da maioria quanto à abordagem com base no risco. Assim, a construção do texto foi realizada com participação daqueles que terão que suportar eventuais ônus causados pela norma, o que traz legitimidade ao processo de criação e agrega conhecimento de diversos setores sociais.

Por fim, deve-se destacar o interesse da norma em efetivamente proibir o uso de práticas de inteligência artificial relacionadas a, de uma forma geral: a) sistemas que empreguem técnicas subliminares para distorção de comportamento; b) sistemas que explorem vulnerabilidades de grupos específicos a fim de distorcer seu comportamento; c) sistemas de IA por autoridades públicas que avaliem a classificação de credibilidade com base em comportamento social ou características; d) salvo exceções, sistemas de identificação biométrica à distância em tempo real, em espaços públicos, e para manutenção da ordem pública.

Embora mais bem especificado no texto legislativo, importante demonstrar que houve a preocupação em efetivamente proibir práticas que o legislador entendeu absolutamente danosas à população, não cedendo espaço, neste ponto, à indústria.

---

[51] Ambos os sistemas de regras, o ético e o jurídico, foram parametrizados com base nos limiares de aceitabilidade do risco dos sistemas de inteligência artificial, o que por si só representa uma inovação no modelo europeu, tipicamente construído como um modelo jurídico *tout-court* em que os valores da União representam o limite da estabilidade do sistema ao qual as instituições e os indivíduos remetem a sua identidade "europeia". (tradução nossa)

Trata-se, pois, de exercício do dever de proteção do Estado para com os seus cidadãos e de, como propriamente sugerido pelo documento, uma análise de risco aceitável.

Há, contudo, uma ausência perceptível quanto aos remédios jurídicos que devem ser utilizados nos casos em que for constatada uma violação a direitos fundamentais.

Todo esse arcabouço legislativo e preocupação da União Europeia, como dito, leva o bloco a um papel de liderança no cenário mundial, servindo o texto como base para o regulamento em diversos outros países.

Importante ressaltar que, apesar de ter sido facultada a participação de diversos atores no processo de construção da proposta, após sua publicação surgiram diversas críticas, notadamente quanto aos valores que seriam pagos pelas empresas para cumprimento do previsto na lei. Em relatório elaborado pelo *Center for Data Inovation*, Mueller (2021) destaca as restrições impostas pela proposta e os seus custos, bem como a sua possível repercussão nos investimentos de inteligência artificial:

> The AIA will cost the European economy $31 billion over the next five years and reduce AI investments by almost 20 percent. A European SME that deploys a high-risk AI system will incur compliance costs of up to $400,000 which would cause profits to decline by 40 percent[52]. (MULLER, 2021, p. 3)

O relatório aponta um possível dano ao processo de transformação digital da Europa, bem como um prejuízo na própria competitividade do continente. São questões que precisam, de fato, ser consideradas, mas aparentemente a preocupação da União Europeia recaiu fortemente sobre a proteção dos direitos dos cidadãos europeus, optando por resguardá-los diante de eventuais ameaças.

4.5. Brasil

O cenário brasileiro quanto à regulamentação da matéria diverge um pouco dos atores antes citados. Como país de perfil precipuamente voltado para o consumo, em

[52] O AIA custará à economia europeia 31 bilhões de dólares nos próximos cinco anos e reduzirá os investimentos em IA em quase 20%. Uma pequena ou média empresa europeia que implemente um sistema de IA de alto risco incorrerá em custos de conformidade de até 400.000 dólares, o que faria com que os lucros diminuíssem em 40 por cento. (tradução nossa)

que há uma defasagem quanto ao fomento de desenvolvimento da tecnologia de ponta, existe um risco de se criar uma legislação voltada para situações de difícil ocorrência. Contudo, louvável a preocupação do país, como importador do produto final, em buscar proteger os seus cidadãos mesmo não possuindo o controle efetivo dos processos de desenvolvimento.

A legislação buscada pelo Brasil quanto ao assunto encontra-se em fase inicial de criação da regulamentação. Dada a estrutura de funcionamento das Casas Legislativas e a ausência de coordenação central quanto ao tema, foram propostos diversos projetos de lei no intuito de buscar a regulamentação da matéria.

No Senador Federal foi apresentado o projeto de lei nº 5.691/2019, no qual se busca instituir a Política Nacional de Inteligência Artificial. O senador justificou a apresentação do projeto:

> De acordo com a pesquisa da consultoria Accenture, essa tecnologia pode duplicar as taxas de crescimento econômico anual até 2035. A previsão é que a Inteligência Artificial aumentará a produtividade em até 40% e permitirá a otimização do tempo por parte das pessoas. (SENADO FEDERAL, 2019, p. 3).

Ressalte-se que o projeto traz em seu bojo os princípios norteadores da Política Nacional de Inteligência Artificial, suas diretrizes, instrumentos e regras, merecendo destaque a preocupação com a valorização do trabalho humano (art. 3º, X), o respeito à autonomia das pessoas (art. 4º, I) e a necessidade de ferramentas de segurança que garantam o controle humano (art.4º, VII).

Ao mesmo tempo, a Câmara dos Deputados também discutiu o tema, tendo logrado êxito em aprovar o Projeto de Lei nº 21 de 2020. No projeto, restou estabelecido o objetivo de regulamentar fundamentos, princípios e as diretrizes relacionados ao desenvolvimento e aplicação de inteligência artificial no Brasil.

O projeto, que se preocupou em definir o que seria inteligência artificial para fins de delimitação do objeto da lei, demonstrou-se absolutamente voltado ao fomento da tecnologia, trazendo poucas preocupações quanto à proteção do cidadão, fato que pode ser demonstrado pela própria análise dos artigos que trazem disposição principiológica.

O art. 3º, que dispõe sobre os objetivos de aplicação, dispõe que o objetivo principal da inteligência artificial no país deveria ser, principalmente, voltado para

questões industriais, ficando a proteção social totalmente relegada ao segundo plano, como pode-se perceber da leitura do texto:

> Art. 3º A aplicação de inteligência artificial no Brasil tem por objetivo o desenvolvimento científico e tecnológico, bem como:
> I – a promoção do desenvolvimento econômico sustentável e inclusivo e do bem-estar da sociedade;
> II – o aumento da competitividade e da produtividade brasileira;
> III – a inserção competitiva do Brasil nas cadeias globais de valor;
> IV – a melhoria na prestação de serviços públicos e na implementação de políticas públicas;
> V – a promoção da pesquisa e desenvolvimento com a finalidade de estimular a inovação nos setores produtivos; e
> VI – a proteção e a preservação do meio ambiente.

Em relação ao art. 4º, tem-se como primeiro fundamento o "desenvolvimento científico e tecnológico e a inovação". No segundo inciso temos a "livre iniciativa e a livre concorrência". Somente no terceiro inciso é que se tem um comando relacionado ao "respeito à ética, aos direitos humanos e aos valores democráticos". O dispositivo ainda traz o "estímulo à autorregulação, mediante adoção de códigos de conduta e de guias de boas práticas", medida esta que já se entende completamente insuficiente para controlar o desenvolvimento da inteligência artificial sem cuidados na proteção dos direitos humanos.

O último dispositivo (art. 5º) voltado à parte principiológica elenca propriamente os princípios que devem nortear o desenvolvimento e aplicação da inteligência artificial no Brasil. Muito embora estabeleça, entre outros, a "centralidade do ser humano" e a "não-discriminação" como princípios, no que se relaciona à transparência, dispõe que devem ser observados "os segredos comercial e industrial", além de limitar o alcance de tal princípio a somente três casos expressos (alíneas a, b e c), reduzindo claramente o seu alcance. Além disso, na alínea c, ao estabelecer que a transparência deve ser respeitada em casos de "critérios gerais que orientam o funcionamento do sistema de inteligência artificial", o texto novamente excepciona o segredo comercial e industrial. Quando a este mesmo dispositivo, o inciso VI, que dispõe sobre o princípio da segurança e prevenção, a lei diz que medidas técnicas, organizacionais e administrativas devem ser utilizadas, atendidos, dentre outros critérios, a "viabilidade econômica".

O texto, então, se apresenta como projeto legislativo que busca a proteção do desenvolvedor de *software*, deixando a proteção do cidadão brasileiro em segundo plano. Garantias como a proteção absoluta do segredo industrial e comercial e o

próprio condicionamento das técnicas de segurança e prevenção à viabilidade econômica demonstram a propositura de uma lei voltada a interesses comerciais. Tal característica é tão marcante que, inclusive em casos concretos em que a administração pública constate alto risco, ainda assim há nova proteção empresarial quanto aos segredos comerciais e industriais (art. 6º, §2º, *in fine*).

Com a aprovação, o PL 21/2020 foi remetido ao Senado Federal e lá tramita com outros projetos (PL 5691/2019, 5051/2019 e 872/2021) para que possa ser efetivamente aprovada uma lei regulatória da matéria no país.

Outra frente regulatória sobre a inteligência artificial vem sendo aberta junto ao Poder Judiciário, mais especificamente no Tribunal Superior Eleitoral.

Em 2016, a empresa Cambridge Analytica foi acusada de obter ilegalmente informação de 50 milhões de usuários do *Facebook*, dividi-los de acordo com seus perfis psicológicos e, através desta informação, direcionar determinados anúncios políticos no intuito de interferir no processo democrático. Esse foi só um dos primeiros casos de maior repercussão que relacionou o uso de dados para interferir nas eleições (Zhang; Han, 2021). Tendo em vista situações como a apresentada, e à míngua de uma legislação que pudesse proteger o pleito a ocorrer em 2024 no Brasil, o Tribunal Superior Eleitoral, no uso de suas atribuições constitucionais, estabeleceu diversos dispositivos acerca do uso de inteligência artificial nas eleições brasileiras, o que demonstra também a preocupação das autoridades judiciais acerca do tema.

4.6. Parametrização

Diante das questões analisadas, é possível concluir que há uma tendência mundial de que existe uma necessidade de regulamentação, muito embora a profundidade e especificidade desta varie bastante. Assim, abandona-se a política de "autorregulação" ou o uso de códigos de ética vagos para a adoção de ferramentas dotadas de efetivo caráter vinculante sob os atores envolvidos em todo o processo de criação, desenvolvimento e uso de *softwares* de inteligência artificial.

Enquanto a União Europeia optou por regulamentar a matéria de forma exaustiva e exigindo bastante dos desenvolvedores no intuito de preservar os direitos individuais dos cidadãos europeus, os Estados Unidos e a China, por exemplo, preferiram adotar uma política mais generalista, com concessão de prazos alongados e estabelecimento de princípios gerais para a consecução das finalidades previstas

nas leis. São duas abordagens diferentes que podem, decorrido o tempo próprio de análise, trazer resultados completamente diversos. A regulamentação em excesso pode trazer efetiva proteção, mas afastar investimentos, enquanto a proteção “aberta” pode ser insuficiente para proteger os cidadãos das violações de direitos que podem surgir pelo uso da inteligência artificial.

O Brasil, por outro lado, propôs adoção de uma lei que seja capaz de proteger os desenvolvedores de *softwares* de inteligência artificial, garantindo a proteção da propriedade industrial, inclusive em detrimento de casos concretos de alto risco ao cidadão. Anda o país na contramão do mundo, que procura focar na proteção do indivíduo em detrimento da máquina, deixando em segundo plano – ainda que não de forma explicita – os custos que isso pode trazer à indústria.

A ausência de regulamentação, por outro lado, é capaz de gerar enorme dano social. Por não haver um comando claro que obrigue os desenvolvedores a observarem premissas básicas de igualdade entre todos aqueles sob que o *software* exercerá qualquer tipo de influência, há grande possibilidade de ocorrência de novos danos, como os já exaustivamente apontados. A ausência de leis claras sobre a matéria torna o desenvolvimento tecnológico campo fértil para surgimento de novas formas de vulnerabilização social, não possuindo decisões judiciais individuais força suficiente para impedir ou forçar todas as empresas produtoras a observarem rigorosos padrões de desenvolvimento dos algoritmos.

É possível concluir, portanto, que a regulamentação figura como principal instrumento eleito pelos países para buscar eventual controle da inteligência artificial, possibilitando o uso de seus inúmeros benefícios, mas sem deixar de proteger o cidadão. Fixado tal ponto de partida, em que se tem a lei como política pública para combate às situações evidenciadas, é possível prosseguir com os objetivos do estudo.

Neste capítulo foram apresentados aspectos gerais relacionados ao direito comparado e analisadas as legislações propostas ao redor do mundo, no intuito de se entender as preocupações de cada um e como escolheram lidar com toda a problema envolvendo o uso de inteligência artificial em questões sensíveis. Tal exercício tem como objetivo estabelecer parâmetros gerais para que possa ser analisada sua efetividade e de que forma podem evitar o aumento de vulnerabilização de grupos de minorias.

No próximo capítulo será estudado o direito das minorias, como historicamente tais grupos são injustamente vulnerabilizados e como as novas tecnologias –

especificamente a inteligência artificial – podem aumentar situações de desigualdade. Ao final, poderemos relacionar toda esta problemática ao já discutido acerca da regulamentação da inteligência artificial para que se possa finalizar o trabalho com a apresentação de conclusões sobre os objetivos do estudo.

5. GRUPOS VULNERÁVEIS, JUSTIÇA PREDITIVA E SISTEMA PENAL

Como já explanado, o presente estudo se propõe a analisar como a utilização de softwares de inteligência artificial e justiça preditiva podem, quando utilizados no sistema penal, aumentar ou até criar formas de vulnerabilização de determinados grupos. Além disso, o estudo comparativo está sendo utilizado entre as legislações e projetos existentes acerca do assunto, para que se observe a possibilidade de solução do problema através do processo legislativo, propondo parâmetros a serem seguidos na elaboração de diplomas e no desenvolvimento dos próprios programas, garantindo, assim, a proteção e o direito a um devido processo legal.

No presente capítulo será explanada a condição das minorias e como, tendo por base o do direito de cidadania, a tais grupos deve ser concedida a mesma gama de direitos prevista a todos os cidadãos, ponderando, nesse ponto, as características que as diferenciam para que possa ser efetivado o princípio da igualdade material.

Para tanto, considerando que a cidadania é um direito intrínseco ao ser humano, será realizada uma explanação acerca de tais direitos para que se possa melhor obter dimensão sobre o seu conceito e conteúdo. De posse de tais informações é que será construída a análise de vulnerabilização de tais grupos quando da utilização de softwares de inteligência artificial que não tiveram a preocupação de protegê-los quando de seu desenvolvimento.

5.1. Direitos Humanos

De acordo com a Organização das Nações Unidas, os Direitos Humanos podem ser definidos como “direitos inerentes a todo ser humano, independente de raça, sexo, nacionalidade, etnia, língua, religião ou qualquer outro *status*” (UNITED NATIONS, [20--], p. 1). Para Comparato (2015), os direitos humanos traduzem a ideia de que nenhum indivíduo pode se afirmar superior aos demais, sendo todos merecedores de igual respeito. Contudo, a conceituação ou o conteúdo de tais direitos não é algo pacificado, sendo comumente discutidos diversos aspectos em sua definição, tais como seu alcance e conteúdo. Nesta toada, entende-se importante analisar a sua evolução histórica com o fim de obter melhor compreensão sobre a sua real abrangência.

5.1.1. Aspectos Históricos

Os Direitos Humanos, como hoje compreendidos, possuem conceituação relativamente recente. Anteriormente, sua essência era abrangida pelo direito natural, cuja existência, ressalta Douzinas (2009), foi reconhecida em textos clássicos dos gregos antigos, como em Antígona, de Sófocles, ou em dogmas dos estoicos.

Aristóteles foi quem desenvolveu o conceito, na *Retórica*:

> de um lado, há a lei particular, e do outro lado, a lei comum: a primeira varia segundo os povos e define-se em relação a estes, quer seja escrita ou não-escrita; a lei comum é aquela que é segundo a natureza. Pois há uma justiça e uma injustiça, de que o homem tem, de algum modo, a intuição, e que são comuns a todos, mesmo fora de toda comunidade e de toda convenção recíproca. (ARISTOTELES, 1959, p. 86)

Percebe-se, então, que o direito tido por *natural* decorre de algo intrínseco ao homem, não exigindo qualquer normatização para sua existência ou validade. Neste sentido é possível observar a sua contraposição ao direito positivo. Essa dicotomia, segundo Bobbio (1995) também pode ser encontrada no Direito Romano, onde havia uma distinção entre o *jus gentium*, que se referia à natureza, e o *jus civile*, relacionado aos estatutos jurídicos postos pela entidade social criada pelos homens.

É importante ressaltar que na época clássica, ainda de acordo com Bobbio (1995), não havia uma preponderância do direito natural sobre o direito positivo, mas sim o contrário. Tal panorama se inverteu durante a idade média, sendo o direito natural, de origem divina, superior.

Houve, ainda, uma nova inversão de dominância durante a formação dos Estados. Com a centralização governamental, os soberanos não aceitaram que o direito emanasse de outra fonte que não fosse a estatal, surgindo uma estrutura monista. Nesse sentido, para Bobbio (1995), ocorreu o processo de monopolização da produção jurídica por parte do Estado, que somente reconhece como válido o direito por ele produzido. Em decorrência deste fato, o direito positivo passou a ser tido como direito em sentido próprio, perdendo o natural o seu *status* de norma cogente, característica essa que permanece até os dias de hoje.

Tendo em vista que a normatização passou a ser requisito essencial do Direito, é possível observar, nas palavras de Costa Douzinas, que “a história condensada do Direito Natural termina com a introdução da Declaração Universal dos Direitos

Humanos" (DOUZINAS, 2009, p. 27). Reforça-se, portanto, a necessidade de positivação do direito para que este seja reconhecido como tal, mesmo se emanado da própria essência humana.

No mesmo sentido aduziu Jürgen Habermas, ao estabelecer a necessidade de normatização para fins de efetivação:

> Somente quando os direitos humanos tiverem encontrado o seu 'lugar' numa ordem jurídica e democrática mundial, isto é, quando funcionarem da mesma maneira que os direitos fundamentais nas nossas constituições nacionais, poderemos inferir, em nível global, que os destinatários desses direitos podem ser considerados também os seus atores (HABERMAS, 2003, p. 50)

Ao processo de passagem do direito natural clássico para os direitos humanos contemporâneos, Douzinas (2009) atribuiu como uma de suas características a "positivação da natureza", ou seja, a transferência do direito natural para o histórico. Importante ressaltar, contudo, que o fenômeno surgido com o positivismo jurídico recebeu diversas críticas, notadamente a acusação de que sua doutrina teria favorecido o surgimento de regimes totalitários, que ampararam a licitude de suas condutas com base simplesmente no que estava previsto na lei (BOBBIO, 1995).

Inobstante as críticas, a Declaração Universal dos Direitos do Homem figurou como bastião dos Direitos Humanos no mundo, tendo lançado as bases que puderam assegurar a reivindicação das garantias ali contidas. Surge, então, de forma normatizada, a ideia de que a dignidade – e os direitos dela decorrentes – é inerente a todos os seres humanos. E, nesta toada, ao contrário do apregoado a título de críticas ao jusnaturalismo, Abbagnano salienta que a utilização dos direitos humanos, em sua forma normatizada, teria auxiliado à superação dos regimes autoritários:

> Pode-se dizer que a exigência da dignidade do ser humano venceu uma prova, revelando-se como pedra de torque para a aceitação dos ideais ou das formas de vida instauradas ou propostas; isso porque as ideologias, os partidos e os regimes que, implícita ou explicitamente, se opuseram a essa tese mostraram-se desastrosas para si e para os outros (ABBAGNANO, 1998, p. 277)

É que, como demonstrou Hannah Arendt, os regimes totalitários buscavam, acima de tudo, a retirada dos direitos dos homens, tornando-os supérfluos, retirando os traços próprios destes no intuito de obter poder:

> Os homens, na medida em que são mais que simples reações animais e realização de funções, são inteiramente supérfluos para os regimes totalitários. O totalitarismo não procura o domínio despótico dos homens, mas sim um sistema em que os homens sejam supérfluos. O poder total só pode ser conseguido e conservado num mundo de reflexos condicionados, de marionetes sem o mais leve traço de espontaneidade. (ARENDT, 2003, p. 507)

O jusnaturalismo, pois, ainda que pós-positivação, conseguiu instituir um patamar mínimo de respeito a todos, universalizando estes direitos.

A concepção de universalização dos direitos humanos adotada pela Declaração Universal dos Direitos do Homem encontrou fundamentação filosófica nos estudos de Kant, que formulou a chamada lei universal da humanidade contida em seu segundo imperativo categórico: “Age de tal maneira que uses a humanidade, tanto na tua pessoa como na pessoa de qualquer outro, sempre e simultaneamente como fim e nunca simplesmente como meio” (KANT, 2007, p. 69). Elucidando o mandamento, Tonetto afirma que “o ser humano não é uma coisa e, por isso, não pode ser utilizado arbitrariamente pela vontade dos outros (TONETTO, 2012, p. 272). O ser humano é o fim, e nunca o meio.

Essa transição de ótica a respeito do homem – sem qualquer distinção de direitos básicos inerentes a todos – possibilitou o fortalecimento da doutrina humanista, ampliando seus horizontes e estabelecendo características como a universalidade, tratada de forma mais aprofundada no tópico a seguir. O homem passa, portanto, a ser somente autor de direitos, e não objeto. Determina-se, desta forma, um valor intrínseco para todos que não pode ser afastado.

Após esta breve introdução histórica, passa-se ao conceito e característica dos direitos humanos para que se possa melhor compreender o objeto do presente estudo.

5.1.2. Características dos Direitos Humanos

Ultrapassada a fase de análise de desenvolvimento histórico dos direitos humanos, necessário se faz esmiuçar suas características, com a finalidade de melhor desenvolver esta tese, eis que tem como base ações de inclusão de minorias e grupos vulneráveis fundamentadas na proteção de direitos humanos, ou deles decorrentes.

Como dito no tópico anterior, o conceito de direitos humanos é amplo, não havendo consenso na comunidade acadêmica. Bobbio (2004) aduz que a maioria das definições de direitos do homem são tautológicas, ao tempo em que destaca que o

elenco dos direitos humanos passa frequentemente por um processo de modificação, de acordo com as condições históricas a que estão submetidos. Seguindo esta mesma linha, Hannah Arendt (1989), compreendeu os direitos humanos como uma realização humana, em constante processo de construção e desconstrução.

Nesse sentido, Flávia Piovesan (2008) observa que os direitos humanos surgem de forma paulatina, decorrentes de reivindicações morais. Tratam-se, portanto, de um ramo jurídico em constante mudança, que necessita de provocação social para que possam surgir novas pautas de defesa.

Ainda que sejam discutíveis aspectos na definição ou no conteúdo conceitual dos direitos humanos, notadamente no decorrer dos tempos, há certo consenso de que a Declaração Universal dos Direitos Humanos de 1948 inseriu, em seu artigo 2, a universalidade e a indivisibilidade dos direitos humanos:

> Todos os seres humanos podem invocar os direitos e as liberdades proclamados na presente Declaração, sem distinção alguma, nomeadamente de raça, de cor, de sexo, de língua, de religião, de opinião política ou outra, de origem nacional ou social, de fortuna, de nascimento ou de qualquer outra situação. Além disso, não será feita nenhuma distinção fundada no estatuto político, jurídico ou internacional do país ou do território da naturalidade da pessoa, seja esse país ou território independente, sob tutela, autônomo ou sujeito a alguma limitação de soberania. (ONU, 1948)

A Universalidade, define Flávia Piovesan (2006), seria aspecto extensivo dos direitos humanos, estendendo a todos, pela simples condição de pessoa, a titularidade de tais garantias. Já a indivisibilidade reside no fato de que o desrespeito a um direito reflete nos demais, razão pela qual a sua proteção deve ocorrer de forma integral.

No mesmo sentido, André Ramos define a universalidade dos direitos humanos como a “atribuição desses direitos a todos os seres humanos, não importando nenhuma outra qualidade adicional, como nacionalidade, opção política, orientação sexual, credo, entre outras” (RAMOS, 2020, p. 68).

É possível perceber neste ponto que a grande maioria dos autores atuais aderem à universalização e indivisibilidade dos Direitos Humanos em consonância ao já apregoado por Kant, desconsiderando ou abandonando a ideia de relativização de tais direitos.

Nas palavras de Heiner Klemmer (2012), a ideia de universalidade dos direitos humanos é decorrente da tradição de Kant e seu conceito de valor universal. Endossando tal afirmação, Lucy Castillo explicita ainda mais a ideia: “*la ley moral*

kantiana exige respeto por todo individuo humano… cada humillación de un individuo por otro o ante otro es una ofensa a la humanidad y un atentado contra la igualdad y la autonomía moral[53]" (CASTILLO, 2010, p.104). Além disso, a autora aproxima o conceito de justiça ao reconhecimento de direitos superiores em detrimento de interesses particulares, mais especificamente os direitos humanos: "*justicia significa, más bien, el reconocimiento de que por encima de los intereses particulares hay un interés universal*[54]" (CASTILLO, 2010, p.106).

Ainda sobre a universalização, Flávia Piovesan (2007) aduz que a concepção contemporânea dos direitos humanos é decorrente da internacionalização dos direitos humanos, notadamente reconstruída após o fim da Segunda Guerra Mundial. Tal afirmação vai ao encontro do que Kant preconizou, como afirmado por Castillo (2010), de que o lugar e as circunstâncias de nascimento não são fatores que devam ser considerados na outorga de direitos humanos. Se deve, na verdade, reconhecer a humanidade onde quer que se encontre o ser humano, não sendo viável ou adequado a restrição de garantias básicas com fundamento em características culturais ou legais de determinada região.

Já sobre a indivisibilidade, Ramos (2011) aproxima o seu conceito da definição de igualdade, afirmando que todos os direitos humanos devem receber a mesma proteção jurídica, vez que essenciais. Ressalta que a indivisibilidade possui dois aspectos, o da unicidade incindível em si e o de que não é possível proteger apenas alguns direitos humanos, sendo necessária uma proteção abrangente, que englobe todos.

Neste sentido, uma análise deste aspecto da igualdade se faz necessário, considerando que o estudo busca compreender em especial o fenômeno da vulnerabilização de grupos marginalizados e como as suas garantias podem ser protegidas sob o manto dos direitos humanos.

Assim, tendo em vista que o direito a igualdade decorre da indivisibilidade reconhecida pela Declaração Universal dos Direitos do Homem, é possível depreender seu caráter concessivo de isonomia, em que se exige uma igualdade de

---

[53] a lei moral kantiana exige respeito por todo indivíduo humano... cada humilhação de um indivíduo por outro ou ante outro é uma ofensa à humanidade e um atentado contra a igualdade e a autonomia moral (tradução nossa)

[54] justiça significa, melhor dizendo, o reconhecimento de que por encima dos interesses particulares, existe um interesse universal (tradução nossa)

tratamento entre todos, sem possibilidade de discriminação odiosa, devendo, ainda, ser garantidas condições dignas de vida.

Relacionando o direito à igualdade com a universalidade dos direitos humanos, Ramos (2020) recorda que foi o surgimento dos Estados Sociais de Direito que possibilitou a busca pela igualdade efetiva entre todas as pessoas. Destarte, aduz não ser suficiente a igualdade perante a lei, mas sim que somente a busca pela erradicação de fatores de inferiorização é que pode garantir a realização deste direito. A promoção da igualdade, pois, figuraria como dever de proteção por parte do Estado.

Os direitos humanos são, ainda, irrenunciáveis, inalienáveis e imprescritíveis.

Por irrenunciabilidade, é possível entender que, ao contrário dos direitos subjetivos, “os direitos humanos têm como característica básica a irrenunciabilidade, que se traduz na ideia de que a autorização de seu titular não justifica ou convalida qualquer violação do seu conteúdo” (MAZZUOLI, 2019, p. 31). Trata-se de importante característica que garante que ninguém será privado de seus direitos básicos, ainda que se manifeste em sentido contrário. Garante-se, assim, mesmo contra a vontade do beneficiário, um núcleo mínimo de garantias fundamentais insuscetível de violação.

A inalienabilidade está relacionada com a cessão pecuniária de direitos para terceiros. Nas palavras de André Ramos, “a inalienabilidade pugna pela impossibilidade de se atribuir uma dimensão pecuniária desses direitos para fins de vendas” (RAMOS, 2020, p. 72). Segundo Ramos (2020), a inalienabilidade encontra suporte em Rousseau, quando este se manifestou em combate à escravidão.

Por fim, a imprescritibilidade está relacionada a impossibilidade de perda desses direitos pelo não uso. Logo, a passagem do tempo sem que sejam reivindicados não impede o reconhecimento dessas garantias naturais.

5.1.3. Dimensões dos Direitos Humanos

A doutrina classifica os direitos humanos em dimensões ou gerações. Para fins do presente estudo, faz-se necessária a exposição de tal divisão doutrinária para que possa ser analisado, posteriormente, como a proteção das minorias se qualifica como direito humano e sob qual perspectiva.

Os direitos de *primeira dimensão* são conhecidos como os direitos de liberdade. Constituem, pois, um rol de garantias oponíveis *contra* o Estado no intuito de garantir seu livre exercício, sendo conhecidos por prestações negativas. Podem ser citados

como exemplos desta geração o direito à vida, à liberdade de locomoção e de associação, entre outros.

A *segunda dimensão* de direitos se relaciona com a igualdade, sendo, notadamente, direitos econômicos, sociais e culturais. São relacionados com prestações positivas por parte do Estado no intuito de garantir a todos condições igualitárias de vida. Os direitos à saúde, cidadania, educação, previdência social, por exemplo, enquadram-se nesta categoria.

Por fim, a *terceira dimensão* relaciona-se à fraternidade, tendo em seu rol direitos de desenvolvimento, ao meio ambiente, comunicação, patrimônio comum da humanidade etc.

É possível, portanto, relacionar as gerações de direitos com a Revolução Francesa, cujo lema "liberté, *égalité, fraternité"* inspirou esta classificação amplamente adotada.

Segundo Mazzuoli (2019), há doutrina que sugere a existência de outras duas gerações, a quarta relacionada à solidariedade (globalização dos direitos fundamentais, democracia direta, direito ao pluralismo) e a quinta que se ligaria ao direito à paz. Contudo, tais categorias foram criadas posteriormente, não sendo objeto de estudo do criador da estruturação clássica, Karel Vasak.

Como visto, o direito à cidadania encontra-se no rol relativo aos direitos de segunda dimensão, pois voltado a prestações positivas por parte do Estado e para a garantia de igualdade entre todos. Tal garantia se liga diretamente ao objeto da pesquisa, eis que é nela que se fundam todos os movimentos sociais e de inclusão de minorias, conforme adiante demonstrado.

Adiante, serão apresentados conceitos relativos à cidadania e a sua relação com o tema objeto de estudo.

5.2. Cidadania

A cidadania pode ser conceituada como a "efetiva vivência, por todos os cidadãos, dos direitos normativamente assegurados" (ARAÚJO, 2017, p. 570). O mais célebre conceito de cidadania, no entanto, foi cunhado por T. H. Marshall, cuja divisão em três partes abrangeu todos os seus aspectos.

Marshall (1967) dividiu a cidadania em três elementos: civil, político e social. A parte civil diz respeito aos elementos voltados à efetivação das liberdades individuais.

O político relaciona-se com o direito de participação no exercício do poder político, influenciando, e sendo influenciado, pelos atores do processo eleitoral. O elemento social, por fim, diz respeito à necessidade de se garantir uma vida digna, dentro dos padrões prevalentes na sociedade.

A cidadania relaciona-se, portanto, à fruição efetiva das garantias conferidas pela Constituição ou pela lei por parte dos cidadãos. A sua efetividade afigura-se como questão central do presente estudo, eis que é exatamente a sua não conferência que gera discriminação a grupos minoritários, conforme trazido mais adiante.

A importância de se efetivar garantias concedidas a todos remonta, como visto, à segunda dimensão dos direitos humanos. Pode-se dizer, inclusive, que qualquer dano causado a outrem, segundo Lucy Carrilo (2010), pode ser entendido como uma ofensa à própria humanidade e um atentado à igualdade. Seria, do ponto de vista kantiano, uma violação ao segundo imperativo categórico, que exige respeito a todo indivíduo.

É possível, portanto, observar o caráter inclusivo dos direitos, que devem abranger todos, impedindo discriminações não abarcadas pelo princípio da igualdade. Neste sentido: "*Rights are axiomatically inclusive, comprehensive in nature, and unconditional, though they may attach to citizens in response to different contingencies during different parts of the human life course*[55]" (DEAN, 2019, p. 131).

Em suma, nas palavras de Hannah Arendt (2010), o direito a ter direitos decorre diretamente da cidadania, como corolário direto dos direitos humanos.

É possível observar que os direitos de cidadania, por tratar de questões voltadas à igualdade, são cotidianamente evocados por grupos minoritários, principalmente em países em desenvolvimento, como forma de lembrar o Estado de sua obrigação na efetivação dos direitos legalmente previstos. O pensamento moderno dos direitos humanos, com fundamentação kantiana, exige que todos possuam suas condições de vida asseguradas de forma digna, sendo tal bandeira frequentemente levantada por tais coletividades.

Surgem, então, estudos que relacionam aspectos formais e materiais da cidadania, buscando respostas para que o Estado possa, materialmente, garantir a todos a vivência dos direitos garantidos por lei. A cidadania formal, somente prevista

[55] Direitos são axiomaticamente inclusivos, abrangentes por natureza e incondicionais, embora possam ser atribuídos aos cidadãos em resposta a diferentes contingências durante diferentes partes do curso da vida humana. (tradução nossa)

em lei, desacompanhada de sua parcela material, que efetiva e garante o seu exercício, é direito inócuo. Necessária, portanto, sua efetivação.

A não conferência de efetividade à cidadania, leciona Clarke, torna-a uma abstração, sem levar em conta as:

> verdadeiras identidades da vida real; ignora aspectos como gênero, a raça ou a orientação sexual, por mencionar alguns. Se centra nas categorias políticas mais genéricas, umas categorias que, devido a sua generalidade e a sua universalidade, deixam de ter sentido político ao estar distantes da realidade" (ARAÚJO, 2017, p. 571).

Importante observar que não basta a concessão de direitos de forma genérica. O reconhecimento de características específicas de determinados grupos exige uma atuação especial para sua efetivação. A busca pela cidadania material, em que todos os grupos sejam igualmente beneficiados pelos direitos legalmente previstos, é corolário direto do direito à igualdade e, portanto, um direito humano. Mas, é preciso que as características individuais sejam analisadas para que seja possível a sua efetivação.

Trata-se, portanto, de uma política de desigualdade lícita, em que o Estado se empenha em garantir aqueles que não desfrutam das mesmas condições o acesso aos mesmos direitos. Segundo Elena Consiglio, se pode justificar esse tipo de discriminação:

> "quando una previsione, criterio o pratica, che svantaggiano in modo proporzionalmente maggiore un gruppo in ragione di una caratteristia prottetta, siano oggetivamente giustificati da una finalità legitima e i mezzi impiegati per il conseguimento di tale finalità siano appropriati e necessari[56]" (CONSIGLIO, 2020, p. 109)

Joaquim Barbosa Gomes discorre sobre essa discriminação positiva:

> A chamada discriminação positiva ou ação afirmativa consiste em dar tratamento preferencial a um grupo historicamente discriminado, de modo a inseri-lo ... impedindo... que o princípio da igualdade formal, expresso em leis neutras que não levam em consideração os fatores de natureza cultural e histórica, funcione na prática como mecanismo perpetuador da desigualdade. Em suma, cuida-se de dar tratamento preferencial, favorável, àqueles que historicamente foram marginalizados, de sorte a colocá-los em um nível de

[56] quando uma previsão, critério ou prática, que privilegia de modo proporcionalmente maior um grupo em razão de uma característica protegida, são objetivamente justificados por uma finalidade legítima e os meios empregados para a obtenção de tal finalidade são apropriados e necessários (tradução nossa)

> competição similar ao daqueles que historicamente se beneficiaram da sua exclusão (GOMES, 2001, p. 22)

Sobre esse reconhecimento relacionado às diferenças entre grupos, diz Bauman:

> O ... descaso em relação à diferença é teorizado como reconhecimento do "pluralismo cultural": a política informada e defendida por essa teoria é o "multiculturalismo". Ostensivamente o multiculturalismo é orientado pelo postulado da tolerância liberal, pela preocupação com o direito das comunidades à auto-afirmação e com o reconhecimento público de suas identidades ... seu efeito é uma transformação das desigualdades incapazes de obter aceitação pública em "diferenças culturais" – coisa a ser louvada e obedecida (BAUMAN, 2003, p. 97-98)

Para Bastos, a promoção da igualdade material é uma meta dos Estados-constitucionais contemporâneos:

> Os Estados-constitucionais contemporâneos possuem como uma de suas metas, com o fim de dar continuidade à ordem constitucional e tendo como viés integrados principal o princípio da dignidade da pessoa humana, a inclusão de todos aqueles que integram a sociedade política, para tal procuram colocar todos, o máximo possível, em uma situação de igualdade material, realizando abstrações e generalizações de situações individuais, por meio da Constituição, da legislação e da aplicação da justiça (BASTOS, 2011, p. 39)

E, esse reconhecimento de que é necessária a garantia de direitos concedidos a todos é uma ação que merece efetivo empenho estatal. Para Araújo:

> O grande e mais grave problema do nosso tempo no que se refere à cidadania não é mais o fundamento de seu conteúdo, uma vez que todos os direitos reconhecidos como componentes da cidadania estão hoje previstos nas constituições sociais, nas leis e em muitos tratados internacionais, mas sim a proteção e a garantia desses direitos (ARAÚJO, 2017, p. 569)

Ainda para Araújo (2017), é necessário, pois, que haja um processo de emancipação social, com um limite ao fluxo de riqueza para as classes mais privilegiadas, promovendo, assim, a distribuição de bens para fins de equilíbrio social.

Comunga deste entendimento Chauí, acerca da necessidade de um novo modelo econômico:

> Destinado à redistribuição mais justa da renda nacional, de tal modo que não só diminua a excessiva concentração da riqueza e o Estado desenvolva uma política social que beneficie prioritariamente as classes populares, mas ainda

> implica o direito dessas classes de defenderem seus interesses tanto através de movimentos sociais, sindicais e de opinião pública, quando pela participação direta nas decisões concernentes às condições de vida e de trabalho – nesse nível, a questão da cidadania é de justiã social e econômica (CHAUÍ, 2007, p. 298)

Deve, portanto, buscar o Estado reconhecer as diferenças de grupos vulneráveis e garantir a aplicação de seus direitos, seja efetivando os já conferidos a todos, seja os discriminando positivamente, para que possam exercer efetivamente sua cidadania.

Assim, partindo da conclusão de que a cidadania material é um direito humano, torna-se necessário analisar a situação de grupos minoritários, sua definição e características, para que seja possível ponderar de que formas a justiça preditiva pode aumentar o processo de vulnerabilização de tais coletivos.

5.3. Minorias

Para fins de definição no presente estudo, o termo minorias pode ser compreendido como:

> Grupo de pessoas que não têm a mesma representação política que os demais cidadãos de um Estado ou, ainda, que sofrem história e crônica discriminação por guardarem entre si características essenciais à sua personalidade que demarcam a sua singularidade no meio social. (MAZZUOLI, 2019, p. 267).

Ou, ainda, as minorias podem ser entendidas como “um contingente numericamente inferior, como grupos de indivíduos, destacados por uma característica que os distingue dos outros habitantes do país (SEGUIN, 2002, p. 9). É importante destacar a dificuldade em se conceituar minorias – ou até grupos vulneráveis, “vez que sua realidade não pode ficar restrita apenas a critérios éticos, religiosos, linguísticos ou culturais. Temos que sopesar sua realidade jurídica ante as conquistas modernas” (SEGUIN, 2002, p. 9)

Importante, pois, distinguir a conceituação de minorias e de grupos vulneráveis. Como própria dedução decorrente dos termos, é possível observar que nem sempre as minorias são grupos vulneráveis e vice-versa. Como já demonstrado neste estudo, há situações, por exemplo, em que mulheres são discriminadas, sendo, contudo, maioria na população do país. Tal contexto não afasta a sua vulnerabilidade diante de diversos aspectos sociais e culturais.

Contudo, em uma conceituação sociológica, o termo “minoria” evoca questões relacionadas ao Estado e a supressão de direito, ocasião em que os termos se aproximam. Segundo Mazarío:

> Desde esta caracterización teórica, las minorías y is grupos vulnerables formarían una única y misma categoría. Sin embargo, a nuestro entender, una minoría es siempre un grupo vulnerable, entendiendo por tal un grupo no dominante o subordinado de la sociedad, pero no sucede lo mismo al contrario, esto es, no todo grupo vulnerable es una minoría, ya que pueden no tener características éticas, religiosas o lingüísticas, que sus miembros no se sientan unidos a dictos elementos distintivos como configuradores de su propria identidad o, en fin, que tengan ningún elemento de permanencia o de lealtad al Estado em que viven. Ello lleva a excluir del ámbito de protección de la minoría a grupos tales como los refugiados, los asilados [e los extranjeros][57] (MAZARÌO, 1997, p. 198)

O estudo de tais grupos vulneráveis, como já destacado, excepciona o princípio da igualdade formal como forma de consagrar a igualdade material, notadamente diante das particularidades de cada um dos seus membros. Trata-se, aqui, de uma justiça distributiva, nos moldes propostos por Aristóteles (2003, p. 82) em que “se não são iguais, não receberão coisas iguais”. É o que se chama de “termo proporcional”, como se referiu o filósofo.

No mesmo sentido, destaca Karl Marx:

> O direito, por sua natureza, só pode consistir na aplicação de um padrão igual de medida; mas os indivíduos desiguais (e eles não seriam indivíduos diferentes se não fossem desiguais) só podem ser medidos segundo um padrão igual de medida quando observados do mesmo ponto de vista, quando tomados apenas por um aspecto determinado… um trabalhador é casado, o outro não; um tem mais filhos do que o outro etc. Pelo mesmo trabalho e, assim, com a mesma participação no fundo social de consumo, um recebe, de fato, mais do que o outro, um é mais rico do que o outro etc. A fim de evitar todas essas distorções, o direito teria de ser não igual, mas antes desigual (MARX, 2012, p. 28)

Assim, é importante que sejam garantidas aos indivíduos diferentes ferramentas de acesso aos seus direitos, ainda que através de instrumentos desiguais. A finalidade precípua do Estado, pois, passa a ser a garantia de acesso a

[57] Desde esta caracterização teórica, as minorias e seus grupos vulneráveis formam uma única e mesma categoria. Sem embargo, no nosso entender, uma minoria é sempre um grupo vulnerável, entendido por tal um grupo não dominante ou subordinado da sociedade, mas não sucede o mesmo ao contrário, isto é, nem todo grupo vulnerável é uma minoria, já que pode não ter características éticas, religiosas ou linguísticas, que seus membros se sintam unidos a ditos elementos distintivos como configuradores de sua própria identidade ou, finalmente, que não tenham nenhum elemento de permanência ou de lealdade ao Estado em que vive. Isto leva a excluir do âmbito da proteção da minoria grupos como os refugiados, os asilados [e os estrangeiros]. (tradução nossa)

determinado bem jurídico, ainda que através de políticas diferentes. Para Araújo (2017), as normas que determinam a redução destas desigualdades, como diferenças regionais, pobreza, desenvolvimento, não possuem o mero *status* de regra, mas sim uma carga valorativa muito superior, urgindo pela atuação positiva do Estado para sua realização. Destaca ainda o mesmo autor que "todo ser humano depende de elementos externos disponibilizados pela construção político-normativa que lhe permitem inserção na vida social e política, tornando-o integrante do Estado" (ARAÚJO, 2017, p. 572).

É de se perguntar, então: as minorias gozam de efetiva cidadania? E mais, a cidadania constitui um direito a ser garantido pelo Estado?

Acerca da efetivação da cidadania, é possível constatar que existe um preconceito estrutural contra grupos historicamente vulneráveis que gera uma frequente sujeição de tais coletivos a práticas de negação de direitos e tratamentos discriminatórios. Tal discriminação é percebida de forma mais explícita no Brasil contra os negros, ainda que seja possível constatá-la em diversos outros casos.

Apesar da mídia e do governo apresentarem um discurso de que o racismo está relegado ao passado, ele continua a influenciar profundamente as estruturas sociais e comportamentos (DAVIS, 2018). O agir dotado de viés discriminatório nem sempre vem de forma consciente. Na verdade, estudos demonstram que muitas vezes tais esquemas raciais ocorrem de forma automática (ALEXANDER, 2018), o que repercute diretamente na atuação dos órgãos estatais, seja por seus protocolos internos, seja pelas próprias pessoas que ali trabalham.

A sociedade brasileira é marcada por racismo e desigualdades de gênero desde seu surgimento, sendo estas opressões estruturais fruto de uma exploração colonialista e que perduraram até os dias atuais em nossas relações e instituições sociais (BORGES, 2019). É neste prisma, de constatação de indiferença por parte daqueles que já possuem seus direitos respeitados, bem como pelo enraizamento de comportamentos preconceituosos, que ganha relevância o presente estudo, na medida em que busca a promoção do direito à igualdade dos grupos vulneráveis. Costa e Barreto (2015) destacam que a noção de vulnerabilidade está diretamente ligada à maior suscetibilidade de violação de direitos, razão pela qual o tema requer maior atenção.

Neste sentido, Piovesan (2006) afirma que o processo de vitimização das minorias ocorre com maior frequência, destacando a necessidade de se obter políticas

não apenas universalistas, mas específicas. A autora ainda salienta ser insuficiente o tratamento genérico de grupos vulneráveis, necessitando um olhar direcionado à suas especificidades. Surge, assim, um direito à diferença, aliado ao respeito à igualdade. Importante mencionar que a proteção ao direito à diversidade foi objeto da Declaração sobre os Direitos das Pessoas Pertencentes a Minorias Nacionais ou Étnicas, Religiosas e Linguísticas de 1992.

Resta, portanto, claro que existe uma necessidade de garantia de direitos às populações minoritárias e que tal direito decorre diretamente de uma necessidade intrínseca: um direito humano.

5.4. Vulnerabilização Social na Esfera Penal

O estudo da vulnerabilização dentro do direito penal merece maior destaque, notadamente diante da forte incidência, neste ramo jurídico, de fatos que impõem supressão de direitos. Uma prisão ou condenação criminal certamente traz grandes impactos, sendo estes ainda mais relevantes quando já existe uma segregação social.

Segundo os últimos dados fornecidos pelo Ministério da Justiça do Brasil (Ministério da Justiça e Segurança Pública, 2019), os presídios brasileiros contam com 66,69% (sessenta e seis virgula sessenta e nove por cento) de pessoas autodeclaradas negras ou pardas. Por outro lado, a população brasileira, segundo IBGE (Instituto Brasileiro de Geografia e Estatística, online) conta com apenas 56,1% (cinquenta e seis virgula um por cento) de pessoas autodeclaradas negras ou pardas.

Os Estados Unidos (United States Census Bureau, 2024), país da maior população carcerária do mundo, por outro lado, possuem apenas 13,6% (treze virgula seis por cento) de sua população formado por pessoas negras, mas, segundo Carson (2023), 35% (trinta e cinco por cento) dos presos são negros.

Em que pese a diferença de proporção no tocante a população, é possível observar a desproporção de incidência das políticas de encarceramento em desfavor dos negros em dois países que são mundialmente conhecidos pela sua grande quantidade de presos. Pode-se deduzir, portanto, que os sistemas não se baseiam em uma proporção populacional, mas sim, na condução das políticas públicas em desfavor de grupos desfavorecidos.

No mesmo sentido, o Anuário Brasileiro de Segurança Pública (2022) atestou:

> “A similaridade entre os vieses implícitos revelados pela literatura nacional e internacional reside justamente na sobrerrepresentação das vítimas negras, que apesar das diferenças demográficas (tais quais o fato de a população negra ser minoria nos EUA, mas maioria no Brasil), aponta para maior incidência da letalidade policial sobre um mesmo segmento: negros, jovens e pobres eu circulam pelas periferias ou nelas residem” (Anuário, 2022, p. 86)

Além disso, o Anuário traz outros dados alarmantes. Negros são 77,6% (setenta e sete virgula seis por cento) das vítimas de mortes violentas intencionais no país. E, em relação à letalidade policial, 84,1% (oitenta e quatro virgula um por cento) dos atingidos são negros.

Como observado pelos dados apresentados por dois países antagônicos no tocante à proporção de sua população negra (mas semelhantes em relação ao encarceramento do mesmo grupo), é possível observar claros indícios de uma política discriminante contra grupos vulneráveis.

Segundo Drakulich e Rodriguez-Whitney (2018), negros enfrentam mais contato que os brancos com a polícia, havendo mais chances de serem parados, revistados e presos. Negros também possuem mais chances de sofrerem abusos por parte da polícia. Além disso, em relação aos Tribunais, negros possuem mais chance de serem processados quando presos (considerando o sistema de disponibilidade da ação penal norte-americano), e, quando são processados, normalmente recebem uma acusação mais grave que os brancos. Por fim, negros tem menos chances de receberem uma sentença reabilitativa (sendo normalmente encaminhados para o sistema prisional) e negros e latinos possuem menor chance de receberem fiança ou livramento condicional.

Os dados apresentados demonstram a clara existência de um desfavorecimento estrutural em desfavor dos negros.

A discriminação estrutural abrange diversos aspectos da vida social e atinge especialmente o usufruto de direitos básicos. Para Hansdeep, Jaspreet and Prabhjot:

> Laws, societal norms, cultural values, and government structures often mask discrimination that impedes on the social, cultural, economic, or political rights of vulnerable communities. These forms of structural discrimination help to create and enshire social inequality amongst marginalized groups. Because this phenomenon is not immediately apparent, impacted communities may find it difficult to establish that they are being discriminated against, and the public may subsequently reject their claims of discrimination, instead blaming

> the impacted communities for failing to assimilate into the existing social structure[58]. (SINGH *et al*., 2013, p. 110)

Para os autores, cinco são as formas de discriminação estrutural: exclusão social, igualdade formal, supressão de identidade, falha em proteger e normas culturais (SINGH *et al*., 2013).

Em relação ao caráter enraizado da discriminação, Consiglio enfatiza sua gravidade:

> "particolamente perniciosa, perché spinge i gruppi com caratteristiche socialmente e storicamente salienti in posizione di marginalità e favorisce il loro soggiogamento da parte di gruppi dominante oppure li mantiene in posizione subalterna[59]. (CONSIGLIO, 2020, p. 123)

E, de forma ainda mais específica, acerca da gravidade no que toca ao racismo estrutural do sistema penal, ressalta Borges:

> Constantemente afirmamos que, por ser estrutural, o racismo perpassa todas as instituições e relações na sociedade. Mas o sistema criminal ganha contornos mais profundos nesse processo. Mais do que perpassado pelo racismo, o sistema criminal é construído e ressignificado historicamente, reconfigurando e mantendo essa opressão que tem na hierarquia racial um dos pilares de sustentação (BORGES, 2019, p. 33)

Mas, o que justifica a existência de um sistema tão enraizado em práticas discriminatórias contra um determinado grupo? A resposta pode estar na origem e políticas de policiamento.

Goff e Kahn (2010), utilizando de dados do sistema americano de justiça, afirmam que negros tem quatro vezes mais chances de serem objeto de uso da força pela polícia e negros e latinos são abordados e presos em taxas superiores a sua representação populacional, notadamente em infrações relacionadas a drogas.

---

[58] Leis, normas sociais, valores culturais e estruturas governamentais muitas vezes mascaram a discriminação que impede os direitos sociais, culturais, econômicos ou políticos de comunidades vulneráveis. Essas formas de discriminação estrutural ajudam a criar e fomentar a desigualdade social entre os grupos marginalizados. Como esse fenômeno não é imediatamente aparente, as comunidades afetadas podem achar difícil estabelecer que estão sendo discriminadas e, posteriormente, o público pode rejeitar suas reivindicações de discriminação, culpando as comunidades afetadas por não se assimilarem à estrutura social existente. (tradução nossa)

[59] particularmente perniciosa, porque empurra grupos com características social e historicamente salientes para uma posição de marginalidade e favorece a sua subjugação por grupos dominantes ou os mantém numa posição subordinada (tradução nossa)

Este contexto pode gerar um número desproporcional de prisões e condenações criminais que, aliados a discriminação social, geral uma repercussão negativa na população negra, que se vê injustamente atacada pelo sistema de justiça. Surgem, então, políticas veladas de *racial profiling*.

O *racial profiling* pode ser definido como "*any police-initiated action that relies on the race, ethnicity, or national origin and not merely on the behavior of an individual*[60]" (RISSE; ZECKHAUSER, 2004, p. 136). Assim, trata-se de ações policiais que são tomadas em decorrência da cor do indivíduo, independente da conduta por ele cometida.

Pode-se pensar que o *racial profiling se* trata apenas de uma teoria e que não encontra eco de forma deliberada nas forças responsáveis pelo policiamento, mas as evidências indicam o contrário. Marshall Frank foi capitão de polícia no Estado da Flórida, nos Estados Unidos e aposentou-se após 30 (trinta) anos de serviço. Em um artigo do jornal Miami Herald, escreveu:

> Yes, racial profiling exists – not as written policy, *de facto*. There's no training for it, no mention of it in any manual. It's subtle, a product of the gut. … Label me a racist if you wish, but the cold fact remains that African Americans comprise 12 percent of the nation's population but occupy nearly half the state and federal prison cells. African Americans account for 2.165 inmates per 100.000 population, versus 307 for non-Hispanic whites and 823 for Hispanics. Now, you may think that the criminal-justice system is racist, and that the police, courts and correctional systems are in cahoots to pack prisons with black folks just because they are black, but I think not. It signals, unfortunately, that African Americans are responsible for a disproportionate chunk of all serious crime[61]. (FRANK, 1999, 7B)

O que se pode deduzir do artigo escrito pelo policial aposentado é que muitos aceitam as políticas de *racial profiling* como algo justo e benéfico, estando muitas vezes tal pensamento presente intrinsecamente na mentalidade policial. Assim, a utilização de dados frios sobre a criminalidade, sem levar em conta o histórico ou

---

[60] qualquer ação iniciada pela polícia que se baseie na raça, etnia ou nacionalidade e não apenas no comportamento de um indivíduo (tradução nossa)

[61] Sim, o perfil racial existe – não como política escrita, de fato. Não há treinamento para isso, nenhuma menção a isso em nenhum manual. É sutil, um produto da intuição. … Rotule-me de racista, se quiser, mas permanece o fato de que os afro-americanos compreendem 12% da população do país, mas ocupam quase metade das celas das prisões estaduais e federais. Os afro-americanos representam 2.165 presos por 100.000 habitantes, contra 307 para brancos não hispânicos e 823 para hispânicos. Agora, você pode pensar que o sistema de justiça criminal é racista e que a polícia, os tribunais e os sistemas correcionais estão em conluio para lotar as prisões com negros só porque são negros, mas acho que não. Isso sinaliza, infelizmente, que os afro-americanos são responsáveis por uma parcela desproporcional de todos os crimes graves. (tradução nossa)

causas do crime, podem levar a conclusão trazida no texto jornalístico, criando uma cultura racista e de fortalecimento do racismo estrutural acima mencionado.

O *racial profiling* pode claramente ser observado na política do *stop and frisk*, que consiste, basicamente, em regras acerca das condições necessárias para realização de revista policial. "*Stop and frisk policies allowed police officers to conduct a pat-down search (stop and frisk) on anyone that the officer reasonably suspected of being engaged in criminal activity*[62]" (BROWING; ARRIGO, 2020, p. 299). Assim, se atribui ao policial, e ao seu prudente arbítrio, a possibilidade de revistar indivíduos de acordo com a sua conduta suspeita. Como esperado, em decorrência da histórica atuação discriminatória da polícia, jovens de grupos minoritários foram desproporcionalmente atingidos por esta política. Segundo O'NEAL (2020), 85% (oitenta e cinco por cento) das abordagens nos Estados Unidos envolviam homens jovens negros ou latinos.

Outro movimento que impactou negativamente grupos minoritários em tempos recentes foi a Teoria das Janelas Quebradas. Esta consiste na ideia de que deve o Poder Público empenhar-se em punir todo e qualquer tipo de pequenas infrações e contravenções penais, além de reparar danos em prédios e estruturas públicas, com a finalidade de se estabelecer um estado de ordem.

Nas palavras de Kelling e Wilson:

> Patrol officers might be encouraged to go to and from duty stations on public transportation and, while on the bus or subway car, enforce rules about smoking, drinking, disorderly conduct, and the like. The enforcement need involve nothing more than ejecting the offender (the offense, after all, is not one with which a booking officer or a judge wishes to be bothered). Perhaps the random but relentless maintenance of standards on buses would lead to conditions on buses that approximate the level of civility we now take for granted on airplanes. … Above all, we must return to our long-abandoned view that the police ought to protect communities as well as individuals. Our crime statistics and victimization surveys measure individual losses, but they do not measure communal losses. Just as physicians now recognize the importance of fostering health rather than simply treating illness, so the police--and the rest of us--ought to recognize the importance of maintaining, intact, communities without broken windows[63]. (KELLING; WILSON, 1982, p. 38).

[62] As políticas de parar e revistar permitiam que os policiais conduzissem uma busca minuciosa (parar e revistar) em qualquer pessoa que o policial suspeitasse razoavelmente de estar envolvido em atividades criminosas (tradução nossa)

[63] Os patrulheiros podem ser encorajados a ir e voltar dos postos de serviço em transporte público e, enquanto estiverem no ônibus ou no vagão do metrô, aplicar regras sobre fumar, beber, conduta desordeira e coisas do gênero. A necessidade de aplicação envolve nada mais do que expulsar o infrator (a ofensa, afinal, não é aquela com a qual um oficial de registro ou um juiz deseja ser incomodado). Talvez a manutenção aleatória, mas implacável, de padrões nos ônibus leve a condições nos ônibus que se aproximem do nível de civilidade que hoje consideramos natural nos aviões. …

Esta teoria serviu de base para a política de tolerância zero que, nos anos 90, foi implementada na cidade de Nova York, nos Estados Unidos. Em decorrência de sua aplicação, que detinha grande apoio popular, foram presos milhões de indivíduos, em sua maioria jovens de minorias (O'NEAL, 2020).

Assim, como "porta de entrada" para o sistema penal, as práticas de policiamento e as políticas públicas que elas envolvem estão diretamente relacionadas com o panorama atual de discriminação do sistema penal, sendo necessário que haja uma radical mudança em seu desenvolvimento para que o sistema criminal possa se afastar dessa condição de participante nessa complexa engrenagem de discriminação racial.

Seria possível cogitar, contudo, que tais grupos minoritários efetivamente cometem mais crimes?

Para Drakulich e Rodriguez-Whitney (2018), três são as causas que poderiam explicar esta disparidade de contato entre minorias e brancos com a polícia: a) há uma efetiva diferença no cometimento de crimes por tais classes sociais; b) as diferenças são produzidas por vieses nas leis e políticas criminais; c) as disparidades são produtos de vieses individuais dos atores que participam do sistema de justiça.

Em relação à primeira possibilidade, os autores apontam que o conceito de raça não é biológico, mas social. Nesse sentido, considerando que grande parte dos negros residem em bairros com nível socioeconômico inferior aos brancos, esta poderia ser uma das causas.

A segunda levanta a questão das políticas públicas de combate ao crime como as acima exemplificadas. Por focarem em "crimes de rua", normalmente cometidos por indivíduos em situação de desvantagem econômica – em sua maioria negros -, em detrimento de crimes "de colarinho branco" – de maioria branca -, haveria uma seleção legislativa e política que desfavoreceria grupos vulneráveis.

Por fim, a terceira hipótese aventada é a de que haveria viéses discriminatórios por parte dos indivíduos que compõem o sistema de justiça. Nesse caso,

---

Acima de tudo, devemos retornar a nossa visão há muito abandonada de que a polícia deve proteger tanto as comunidades quanto os indivíduos. Nossas estatísticas criminais e pesquisas de vitimização medem perdas individuais, mas não medem perdas comunitárias. Assim como os médicos agora reconhecem a importância de promover a saúde em vez de simplesmente tratar a doença, a polícia - e o resto de nós - deve reconhecer a importância de manter comunidades intactas, sem janelas quebradas.

considerando que se trata de um complexo sistema que envolve diversos órgãos, seria necessário avaliar a questão como um conjunto de fatores que, desde a infância, levam as minorias a uma situação de sujeição ao sistema penal. Nesse sentido:

> Taken together, this body of research assessing the cumulative effects of racial disparities within the various institutions of the justice system indicates that minority youth, young African American males in particular, find themselves more likely to experience police contact and intervention, which, if the result is processing as a juvenile offender, places them at higher risk for custodial sentencing, reduced access to rehabilitative intervention, and increased risk of arrest as adults, at which point they are more likely to be aggressively prosecuted, subjected to pretrial detention, and given longer, more punitive sentences. This induces a disproportionately severe criminal stigma the deleterious effects of which can haunt them for decades after[64] (DRAKULICH; RODRIGUEZ-WHITNEY, 2018, p. 24-25)

Sobre o tema, conclui Zaffaroni (1988) ao afirmar que essa seletividade de cunho racial decorre de uma criminologia positivista que sobrevive até os dias de hoje. Apesar de não ser mais admitida expressamente a figura do “criminoso nato” de Cesare Lombroso, a legitimidade de um sistema penal que prioriza fatores biológicos ainda se faz presente nas faculdades de direito e na formação dos policiais.

Em resumo, as diferentes formas no cometimento de crimes entre indivíduos podem ser melhor explicadas através de suas diferenças socioeconômicas do que por fatores biológicos. Mas o que acontece práticas discriminatórias como as acima exemplificadas ocorrem no âmbito do Poder Judiciário de forma automatizada?

5.5. Justiça Preditiva e Sistema Penal

Dentro da perspectiva de garantia de direitos a grupos vulneráveis – sob o manto protetivo do direito à cidadania – o presente estudo busca analisar a utilização de softwares de inteligência artificial - mais especificamente o seu uso dentro do sistema penal - possíveis práticas de violação contra estas minorias decorrentes do

---

[64] Tomados em conjunto, esse corpo de pesquisa avaliando os efeitos cumulativos das disparidades raciais dentro das várias instituições do sistema de justiça indica que jovens de minorias, jovens afro-americanos em particular, encontram-se mais propensos a experimentar contato e intervenção policial, o que, se o resultado for o processamento de um infrator juvenil, os coloca em maior risco de prisão preventiva, acesso reduzido a intervenção de reabilitação e maior risco de prisão quando adultos, ponto em que eles são mais propensos a serem mais rigorosamente processados , sujeitos a prisão preventiva e condenados a sentenças mais longas e punitivas . Isso induz um estigma criminal desproporcionalmente grave, cujos efeitos deletérios podem assombrá-los por décadas. (tradução nossa)

uso de tais programas e como está sendo feita a sua regulamentação pelos Órgãos responsáveis.

Neste ponto, importa ressaltar que a regulamentação pode se apresentar como uma ferramenta eficaz, tendo em vista a existência de indícios de que a utilização de tais softwares em processos decisórios tem gerado novas formas de vitimização de grupos já marginalizados, como veremos a seguir. E, com a construção de normas rígidas, seria possível mitigar tais danos.

Um termômetro nas mãos de um homem branco é avaliado como um binóculo pelo programa de análise de imagens do Google Vision, enquanto nas mãos de um negro, como uma arma (ROCHA, 2020). A imagem de um homem calvo na cozinha é rotulada como de uma mulher pela inteligência artificial (SALAS, 2017). Cidadãos de determinados países eram selecionados pelo computador em aeroportos para filas de embarque rápido com base na prosperidade e riqueza de sua terra natal (HEAVEN, 2020). Todas essas situações aconteceram e foram causadas pela problemática que o presente estudo se propõe a pesquisar: vieses de preconceito na utilização de inteligência artificial.

Como se pode imaginar, tais vieses ganham maior repercussão quando utilizados, inclusive muitas vezes de forma autônoma e sem supervisão humana, para decidir sobre aspectos da vida de terceiros. Interessante ainda notar, como reflete Cathy O'Neal (2020), que tais decisões são reiteradamente tomadas por máquinas em detrimento das massas, permanecendo as classes privilegiadas com um processamento realizado por humanos. Ou seja: muito embora uma grande quantidade de pessoas seja atingida pelos experimentos da máquina e dos novos programas, apenas um grupo menor é protegido, garantindo que seus direitos não sejam violados. Além disso, os impactos do uso de inteligência artificial não são igualmente distribuídos, permanecendo a desigualdade na medida em que grupos mais favorecidos experimentam melhores resultados que outros em estado de vulnerabilidade (RASO et al., 2018).

Powers e Ganascia (2020) lembram que, assim como a tecnologia nuclear e o estudo do DNA recombinante, novas tecnologias trazem riscos e benefícios, cabendo uma análise social e governamental acerca de seus limites.

Diversos estudos foram realizados no intuito de evitar tratamentos discriminatórios pelo uso de inteligência artificial, inclusive por entidades de caráter global, como o World Economic Forum, sem que fosse possível evitar a ocorrência de

novos casos. O *paper* intitulado "*How to prevent discriminatory outcomes in machine learning*" destacou em sua conclusão: "*There is no one-size-fits-all solution to eliminate the risk of discrimination in machine learning systems, and we recognize that many of our recommendations will require context-specific tailoring*.[65]" (WORLD ECONOMIC FORUM, 2018).

Já uma pesquisa realizada pelo Berkman Klein Center, da *Harvard Law School*, constatou, relacionando diretamente a inteligência artificial aos direitos humanos, a existência de 06 eixos em que se identificou algum tipo de vitimização a grupos específicos: justiça criminal, acesso ao sistema financeiro, saúde, educação, moderação de conteúdo online e recursos humanos (RASO et al., 2018).

E ainda, para Ales Zavrsnik:

> The human rights that may be impacted through the use of automated processing techniques and algorithms are: (1) the right to a fair trial and due process; (2) privacy and data protection; (3) freedom of expression, (4) freedom of assembly and association, (5) the right to an effective remedy, (6) the prohibition of discrimination, (7) social rights and access to public services, and (8) the right to free elections[66]. (ZAVRSNIK, 2020, p. 575)

Resta demonstrado, portanto, que além da constatação do problema causado pelos vieses discriminatórios de *softwares* de inteligência artificial, existe também a dificuldade em se vislumbrar efetivas medidas para sua resolução, decorrendo daí a relevância e necessidade de se estudar a problemática.

Mas para melhor compreendê-la, e regulamentá-la, se faz necessário entender o seu funcionamento.

5.5.1. Modelos de Decisão

Como definido no capítulo 2, em simples palavras, um software de inteligência artificial corresponde a um modelo, um conjunto de regras preestabelecidas que ditam

---

[65] "Não há uma solução única para eliminar o risco de discriminação em sistemas de aprendizado de máquina, e reconhecemos que muitas de nossas recomendações exigirão adaptação específica ao contexto. (tradução nossa)

[66] Os direitos humanos que podem ser afetados pelo uso de técnicas e algoritmos de processamento automatizado são: (1) o direito a um julgamento justo e devido processo legal; (2) privacidade e proteção de dados; (3) liberdade de expressão, (4) liberdade de reunião e associação, (5) direito a um recurso efetivo, (6) proibição de discriminação, (7) direitos sociais e acesso a serviços públicos e (8) direito a eleições livres. (tradução nossa)

à máquina quais ações devem ser tomadas. Um modelo, portanto, nada mais é do que isso: uma fórmula.

O'Neal (2020) lembra que todos possuem modelos em seu subconsciente. Através deles é possível analisar situações e decidir a melhor forma de se comportar. O mesmo acontece com as máquinas.

Contudo, as máquinas adquirem os seus "modelos" através do trabalho de profissionais de tecnologia. Estes, no exercício de sua atividade, estabelecem situações e critérios nos quais o *software* deve comportar-se de determinada forma. Mas, quais são estes critérios? Quem os define? É aqui que o problema começa a surgir.

De início, é possível perceber que, apesar de toda a capacidade de processamento de dados, uma máquina nunca poderá atingir a inteligência humana no que toca à análise das complexas relações sociais e suas nuances. Neste ponto, carece o computador de informações que, até hoje, não são possíveis de serem fornecidas. É o que O'Neal definiu como "pontos cegos". E são exatamente os pontos cegos que abrem margem para que sejam encontrados resultados imprecisos. Esses pontos cegos, muitas vezes, geram a necessidade de utilização de dados acessórios (*proxy*) para que se chegue ao resultado almejado, o que contribui também para a imprecisão da resposta.

O fenômeno da utilização de *proxies* passou a ser notado também após a adoção massiva de um procedimento conhecido como *Blindness* ou *treatment parity*. No intuito de se impedir o uso de informações sensíveis que levasse à decisões discriminatórias, programadores passaram a simplesmente a suprimir os dados controversos. Com tal conduta, os programas passaram a utilizar de proxies que substituíam da mesma forma a característica suprimida, não possuindo o procedimento, portanto, adesão dos pesquisadores (Tschantz, 2022)

A utilização de proxies para substituição de dados ausentes também pode gerar resultados tão discriminatórios quanto o próprio dado excluído. Por exemplo, pode-se relacionar estatisticamente o CEP de um indivíduo com o seu potencial para pagamento de um empréstimo (ONEAL, 2020).

Além desta ausência de dados, *softwares* de inteligência artificial podem sofrer com a inclusão de dados *ruins*. São estes dados ruins, por sua vez, que geram decisões dotadas de *vieses* (*bias*).

Podemos definir os vieses como:

> ... unconscious, automatic tendencies to associate certain traits with members of particular social groups, in ways that lead to some very disturbing errors: we tend to judge members of stigmatized groups more negatively, in a whole host of ways[67] (SAUL, 2013, p. 244).

Da definição apresentada, vemos que o viés ocorre de forma inconsciente e automática, relacionada a traços de determinados grupos sociais. É comum, apesar de tratar-se de tratamento preconceituoso, que todas as pessoas possuam algum tipo de viés quando deparadas com determinadas situações do cotidiano. Contudo, em relação às máquinas, estas recebem tais influências por parte do homem, reproduzindo assim o seu comportamento.

Importante ressaltar que, para Friedman e Nissenbaum (1996), a ocorrência do *bias* exige recorrência. Ou seja, a discriminação injusta, por si só, não configura o viés decisório, a não ser que seja recorrente.

Continuam os autores ao aduzir que os vieses podem ser classificados em três categorias: preexistente, técnico e emergente. O viés preexistente é aquele que consiste na obtenção, por parte do *software*, de preconceitos e tendências já existentes socialmente, independentemente de sua atuação. Podem tais dados terem sido obtidos através de vieses do próprio programador ou de dados coletados de forma aleatória.

O viés técnico advém não de um problema nos dados, mas sim de limitações da própria máquina. Podem ser encontrados em várias situações:

> Sources of technical bias can be found in several aspects of the design process, including limitations of computer tools such as hardware, software, and peripherals; the process of ascribing social meaning to algorithms developed out of context; imperfections in pseudorandom number generation; and the attempt to make human constructs amendable to computers, when we quantify the qualitative, discretize the continuous , or formalize the nonformal[68] (FRIEDMAN; NISSENBAUM, 1996, p. 335)

---

[67] Tendências inconscientes e automáticas de associar certos traços com membros de determinados grupos sociais, de modo que gera alguns erros desconcertantes: nós tendemos a julgar membros de grupos estigmatizados mais negativamente das mais diversas formas (tradução nossa)

[68] Fontes de viés técnico podem ser encontradas em vários aspectos do processo de design, incluindo limitações de ferramentas de computador como hardware, software e periféricos; o processo de atribuir significado social a algoritmos desenvolvidos fora de contexto; imperfeições na geração de números pseudoaleatórios; e a tentativa de tornar construções humanas alteráveis a computadores, quando quantificamos o qualitativo, discretizamos o contínuo ou formalizamos o não formal (tradução nossa)

Por fim, o viés emergente é aquele que ocorre somente após o uso. Assim, após atuação do usuário, os dados tornam-se tendenciosos, necessitando de ajuste para que sejam evitados resultados imprecisos ou incorretos.

Além disso, o viés pode surgir em decorrência de dois fatores: a) os desenvolvedores possuem um viés discriminatório que não tem ciência, passando essa característica de forma inconsciente para o algoritmo; ou b) fatores estáticos utilizados na programação não acompanham os fatores dinâmicos, impossibilitando a análise atual dos fatos e perpetuando ideias preconceituosas preexistentes (Zhang; Han, 2022).

Resta claro que tais vieses tomam maiores contornos quando são inseridos na máquina, que não possuem a capacidade de reflexão própria do ser humano na medida em que, ainda que se depare com vieses pessoais, pode afastá-los e decidir de forma justa.

Estando diante de uma possível prática institucionalizada de discriminação, é necessária a investigação sobre o fato, principalmente para evitar sua ocorrência – se possível – e o seu avanço.

A discriminação estatística, muitas vezes ocorrida pelo uso de ferramentas de análise de risco, pode ser definida como:

> The phenomenon of a decision-maker using observable characteristics of individuals as a proxy for unobservable, but outcome relevant characteristics. The decision-makers can be employers, college admission officers, health care provides, law enforcement officers, etc., depending on the specific situation. The observable characteristics are easily recognizable physical traits which are used in society to broadly categorize demographic group by race, ethnicity, or gender. But, sometimes the group characteristics can also be endogenously chosen, such as club membership or language[69]. (FANG; MORO, 2010, p.1)

Por isto, a concessão de poder decisório à softwares de inteligência artificial que não possuem a capacidade de reflexão humana pode gerar a institucionalização da injustiça, impedindo que possa o usuário do serviço se valer do princípio básico da equidade.

---

[69] O fenômeno de um tomador de decisão usando características observáveis de indivíduos como um proxy para características não observáveis, mas relevantes para o resultado. Os tomadores de decisão podem ser empregadores, oficiais de admissão de faculdades, prestadores de serviços de saúde, policiais etc., dependendo da situação específica. As características observáveis são traços físicos facilmente reconhecíveis que são usados na sociedade para categorizar amplamente o grupo demográfico por raça, etnia ou gênero. Mas, às vezes, as características do grupo também podem ser escolhidas endogenamente, como associação ao clube ou idioma. (tradução nossa)

A discriminação estatística – e por que não dizer algorítmica – em modelos de decisão automatizados institucionaliza um critério de generalização, impedindo uma análise individual de cada caso, ou seja, "*[o]n the basis of a characteristic of some members of a class, we reach conclusions or make decisions about the entire class*[70]" (SCHAUER, 2003, p.4).

O funcionamento do *softwares*, conforme estabelecido, se dá por um sistema de *input* e *output*, sendo o primeiro os dados que foram utilizados no treinamento da máquina, enquanto o segundo diz respeito à resposta ou resultado advindo do problema apresentado. Para Gillespie, "*Algorithms need not be software: in the broadest sense, they are encoded procedures for transforming input data into a desired output, based on specified calculations*[71]" (GILLESPIE, 2014, p. 167).

De acordo com Arowosegbe (2023), o *bias* pode se fazer presente em qualquer destes dois momentos, contudo é preciso notar que "*input bias usually result in output data bias*[72]*"(AROWOSEGBE, 2023, p. 27).* Foi isso que deu origem ao axioma "*garbage in, garbage out*". Portanto, é essencial que o dado coletado possa ser o mais "limpo" possível, evitando que seja levado qualquer tipo de enviesamento por ele carregado.

Para Dilmegani*, "AI bias is an anomaly in the output of machine learning algorithms, due to the prejudiced assumptions made during the algorithm development process or prejudices in the training data*[73]". (DILMEGANI, 2023, online). Sobre tal aspecto, esse *bias* preexistente pode ocorrer em razão de erros inconscientes na programação (cognitivo) ou por falta de maior abrangência dos dados coletados. Este segundo decorre da falta de representatividade nas amostras analisadas, por exemplo.

Dessa forma, considerando que os dados são informações e produtos do mundo em que são coletados, é esperado que estes sejam acompanhados de todos os desvios identificados na sociedade. Logo, os dados de uma sociedade

---

[70] Com base em uma característica de alguns membros de uma classe, chegamos a conclusões ou tomamos decisões sobre toda a classe. (tradução nossa)

[71] Os algoritmos não precisam ser software: no sentido mais amplo, são procedimentos codificados para transformar dados de entrada em uma saída desejada, com base em cálculos especificados. (tradução nossa)

[72] viés de entrada geralmente resulta em viés de dados de saída (tradução nossa)

[73] O viés da IA é uma anomalia na saída dos algoritmos de aprendizado de máquina, devido às suposições preconceituosas feitas durante o processo de desenvolvimento do algoritmo ou preconceitos nos dados de treinamento (tradução nossa). Acessado em: https://research.aimultiple.com/ai-bias/

extremamente racista certamente gerarão *outputs* igualmente discriminatórios. Para se evitar essa contaminação, Gillespie sugere o seguinte processo:

> the information incuded in he database must be rendered into data, formalized so that algorithms can act on it automatically… Recognizing the ways in which data must be "cleaned up" is an important counter to the seeming automaticity of algorithms… algorithms can be understood by looking closely at how the information must be oriented to face them, how it is made algorithm ready[74]" (GILLESPIE, 2014, p. 170-171)

O processo descrito por Gillespie permite visualizar o funcionamento do *software* com o olhar voltado para a categorização da informação, possibilitando que determinados dados possam ser tratados de forma diferente, de acordo com a sua classificação. Tal processo é frequentemente utilizado nos dias atuais. Gillespie (2014) cita como exemplo a exclusão, por ferramentas de busca, de documentos protegidos por direitos *copyright*, sites pornográficos ou vírus.

Nesse sentido, este processo de captação se torna um campo fértil para a proliferação de vieses, conforme aduz Arowosegbe: "

> on a basic level, it is factual to note that data is taken from the world. The data in its original state is unorganized, random and crude. The data thus have to be collected, cleaned and processed for machine-learning purposes. These processes are naturally prone to biased outcomes[75]. (AROWOSEGBE, 2023, p.27)

Obseva-se, então, que esse processo de captação orgânica de informação também pode ser a causa de vieses, exigindo, assim, um cuidado ainda maior no desenvolvimento do algoritmo para que se evite o transporte de posicionamentos discriminatórios de uma sociedade ou do próprio programador.

Há, portanto, uma clara noção de controvérsia acerca dos dados e na forma pela qual são coletados, gerando, conforme Arora *et al* (2023), um desequilíbrio que reforça a marginalização:

---

[74] as informações incluídas no banco de dados devem ser transformadas em dados, formalizadas para que os algoritmos possam agir sobre eles automaticamente... Reconhecer as maneiras pelas quais os dados devem ser "limpos" é um importante contraponto à aparente automaticidade dos algoritmos... os algoritmos podem ser compreendidos observando atentamente como as informações devem ser orientadas ao encontrá-los, como elas deixaram os algoritmos prontos" (tradução nossa)

[75] em um nível básico, é factual notar que os dados são retirados do mundo. Os dados em seu estado original são desorganizados, aleatórios e brutos. Os dados, portanto, precisam ser coletados, limpos e processados para fins de aprendizado de máquina. Esses processos são naturalmente propensos a resultados tendenciosos (tradução nossa)

> As such, we suggest that an imbalance of risk and harm in AI development is reinforcing marginalization. As AI products progress further based on unrepresentative datasets and technological infrastructure that is not available to all, the represented population will gain benefits and advantages over the marginalized, and in the process create even more value and wealth. While in the short term marginalized populations may be provided with temporary resources in the form of low-pay and precarious work, ultimately, this can result in an exacerbation of inequality and marginalization in the longer run where the marginalized end up in an even less equitable position. In short, AI can further marginalize the marginalized, and a more meaningful way of examining and understanding such risks is necessary in order to manage the dynamic of complex, inter-related benefit and harm[76] (ARORA et al., 2023, p. 3)

Explicitada a forma como os dados são coletados e como este processo pode influenciar na sua própria confiabilidade, será demonstrado a seguir como seria possível implementar um mecanismo de proteção pautado em direitos humanos e como tais modelos decisórios geram abrangentes impactos quando utilizados pela Justiça Criminal.

5.5.2. A utilização de modelos de decisão em casos envolvendo matéria penal

Os modelos de inteligência artificial mais largamente utilizados na esfera penal são aqueles relacionados à análise de risco (*risk assessment)*. Como já definidos e mais bem explicitados no capítulo 2, os programas de análise de risco são desenvolvidos para, ponderando informações específicas de cada indivíduo submetido ao sistema de justiça penal, fornecer dados acerca da possibilidade de cometimento de novos crimes.

O funcionamento do *software*, de uma forma geral, parte de uma premissa simples: em uma etapa inicial, na apresentação de um questionário predefinido, no qual constam informações pessoais, como local onde residência, antecedentes criminais, circunstâncias relacionadas ao seu nascimento e sua formação, constando,

---

[76] Como tal, sugerimos que um desequilíbrio de risco e dano no desenvolvimento da IA está reforçando a marginalização. À medida que os produtos de IA progridem ainda mais com base em conjuntos de dados não representativos e infraestrutura tecnológica que não está disponível para todos, a população representada ganhará benefícios e vantagens sobre os marginalizados e, no processo, criará ainda mais valor e riqueza. Embora no curto prazo as populações marginalizadas possam receber recursos temporários na forma de trabalho precário e de baixa remuneração, em última análise, isso pode resultar em uma exacerbação da desigualdade e marginalização no longo prazo, onde os marginalizados acabam em uma posição ainda menos equitativa. Em suma, a IA pode marginalizar ainda mais os marginalizados, e uma maneira mais significativa de examinar e compreender esses riscos é necessária para gerenciar a dinâmica de benefícios e danos complexos e inter-relacionados (tradução nossa)

ainda, dados sobre parentes e amigos e a existência de condenações penais, por exemplo. Tais informações podem, ainda, ser obtidas através de um processo automatizado, com coleta direta em bancos de dados de órgãos públicos ou privados. De posse de tais informações, é fornecido um *score* em que se aponta o fator de risco.

Sobre a utilização de tais *softwares,* é importante pontuar que foram desenvolvidos, a princípio, com o intuito de melhor gerir a população carcerária, garantindo o direito de responder ao processo em liberdade quando houvesse baixo risco de reincidência. Além disso, a busca pela isonomia em tais decisões foi um fator decisivo para o seu desenvolvimento. É importante pontuar, também, que sua utilização não é algo recente, havendo registros de uso de tais programas desde 1928, quando foi empregado pela Corte de Justiça do estado americano de Illinois para auxílio quanto às decisões de soltura (DeMichele *et al*, 2020).

No tocante aos sistemas de *risk assessment*, diz Uchôas:

> ... fundada em análise estatística e tratamento de dados por softwares, visa o prognóstico de risco de violência, auxiliando os juízes a manter encarcerados os indivíduos que, eventualmente, representem alguma espécie de risco social. A classificação influenciará no *quantum* de pena que será atribuído ao réu, bem como a quais benefícios ele fará jus no curso da execução penal. [...] Contudo, ao escolher em quais casos os testes devem ser aplicados, é possível afirmar que, embora os resultados sejam estatísticos, não estão livres das preferências pessoais. Também não há evidências empíricas de que o emprego dessas tecnologias torna as decisões mais racionais (Hart, Michi e Cooke, 2007) o que influencia negativamente a argumentação dos juízes nas decisões envolvendo resultados de testes preditivos. (UCHOAS, 2018, p. 42)

A ideia acerca da utilização de ferramentas de avaliação de risco é exatamente a oposta do fenômeno aqui estudado. O que se imaginou é que, com a fixação de critérios estáticos, que não dependessem do livre – e oculto – arbítrio dos magistrados, haveria um aumento da objetividade e equidade das decisões. Nesse sentido, a máquina seria mais justa por aplicar critérios idênticos. O que não se imaginou é que os dados utilizados pelo *software* poderiam, de alguma forma, apresentar também vieses discriminatórios.

Acerca da parcialidade dos juízes, estudos já mostraram que existem preconceitos relacionados à raça em relação à condenação e decisões de soltura, ainda que mediante a utilização de softwares de análise de risco. Na Flórida, por exemplo, constatou-se que juízes davam sentenças mais severas em relação à réus negros em detrimento dos brancos, mesmo diante de *scores* idênticos (GREEN;

CHEN, 2019). Então, ao estudar a temática dos vieses de preconceito nos modelos de previsão de risco, é importante ter em mente que tal fenômeno não é uma característica própria das máquinas, mas sim um fato suscetível de ocorrer com todos que, de alguma forma, participam de algum processo decisório.

Quanto à importância da discussão sobre o uso de tais programas e o seu benefício dentro do sistema penal, Kleinberg *et al.* (2017) identificaram que juízes prendem muitos em casos que não se constatou um alto risco dos réus, mas em contrapartida soltam casos de efetiva gravidade. Mesmo assim, menos de 10% dos magistrados tendem a acreditar que seu julgamento é superior ao da máquina (HYAT; CHANENSON, 2016).

Em um estudo realizado para avaliar a forma como os *softwares* de análise de risco de fato influenciam os responsáveis pela tomada de decisão, Green e Chen (2019) concluíram que as pessoas, mesmo fazendo uso de tais programas e analisando os seus resultados, apresentavam decisões menos precisas do que o conselho fornecido. Além disso, constataram também que, diante dos dados fornecidos pelo modelo de inteligência artificial, as pessoas tendem a aumentar a análise de risco em detrimento de negros em até 25,9% e reduzi-la quanto aos brancos. Atestaram, assim, que os *softwares* atingem melhor resultados que os humanos, mas salientaram que é importante não confiar completamente em tais sistemas, considerando seu potencial de engessamento de reformas no sistema criminal.

Destarte, se o *software* apresenta um percentual de acerto melhor que o homem, por que então há tanto debate acerca de seu uso dentro do âmbito do sistema de justiça penal?

Um dos importantes contrapontos a levantados diz respeito a uma política de *color-evasion*, ou seja, a ocorrência de um afastamento por parte das instituições acerca de questões relacionadas à raça ou cor, sob a suposta alegação de neutralidade. Para DaViera *et al.* (2021) o problema de tal política supostamente imparcial consiste exatamente no fato de que se ignora diversos aspectos históricos, sociais e políticos que devem permear o debate público quando questões de tratamento de grupos minoritários são tratadas. A adesão a tal política de silêncio somente mantém a situação de vulnerabilização já estabelecida, sem permitir avanços significativos em direção a uma sociedade mais igualitária.

Além disso, importante salientar, como já dito, a incapacidade evolutiva dos *softwares* de inteligência artificial. Por atuarem com base em modelos pré-definidos, a incapacidade de inovação é um dos principais pontos de críticas a tais sistemas. Se for considerado hoje que existe uma estrutura desigual e de opressão às minorias, como já demonstrado, por exemplo, no item 5.4. deste estudo, como se pode buscar um melhoramento social mantendo tais práticas?

Sobre este ponto: "*If past decisions are rooted in bias or prejudice, then the data that express these decisions is contaminated, and decisions (high probability predictions) derived from that data will perpetuate the inequities*[77]." (SUSSKIND,2019, p. 288).

No tópico seguinte, serão avaliados casos práticos acerca de falhas graves supostamente identificadas em tais softwares que podem gerar um imenso debate acerca de sua utilização, mesmo diante de tais dados promissores relacionados a sua precisão.

5.5.2.1. O Caso "COMPASS"

O caso COMPASS tornou-se mundialmente conhecido pela visibilidade do estudo realizado por Larson *et al.* (2016), publicado pela ProPublica em um formato de matéria jornalística. Os dados ensejaram grande debate acadêmico e deram início a uma série de estudos acerca do tema dos vieses algorítmicos.

Como já exposto, a ideia de automação de atividades permeia o ambiente jurídico há décadas, mas somente em tempos recentes surgiram discussões mais contundentes – técnicas e filosóficas – sobre o uso desta tecnologia. Por óbvio, o avanço na capacidade de processamento das máquinas deu novos contornos e trouxe novas preocupações à temática, mas é possível afirmar que com a publicação dos dados pela ProPublica o debate alcançou níveis antes não vistos.

O estudo em análise diz respeito aos dados relacionados ao *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS), *software* de análise de risco que fornece ao juiz que está analisando o caso um *score* relacionado ao risco de reincidência ou fuga do indivíduo.

[77] Se as decisões passadas estão enraizadas em viés ou preconceito, então os dados que expressam essas decisões estão contaminados e as decisões (previsões de alta probabilidade) derivadas desses dados irão perpetuar as desigualdades (tradução nossas)

Acerca do COMPAS:

> Si tratta di un algoritmo prodotto dalla società privata E-quivant che valuta il rischio di recidiva e la pericolosità sociale di un individuo sulla base di vari dati statistici, precedenti giudiziari, un questionario che viene somministrato all'imputato stesso, nonché su di una serie di altri variabili coperte da proprietà intellettuale da parte della società medesima[78]. (SIMONCINI, 2019, p.72)

Em agosto de 2013, Eric Loomis foi condenado a uma pena de 8 anos e 6 meses de prisão por direção de veículo roubado e fuga da polícia. Na decisão, o juiz declarou expressamente o uso do COMPAS e que este tinha indicado o alto risco do condenado para a comunidade (ANGWIN, 2016). Loomis, então, recorreu à Suprema Corte do Estado de Wisconsin, alegando violação do devido processo legal, uma vez que não teria tido acesso aos dados que levaram o programa a indicá-lo como de alto risco. A Corte negou o recurso, validando o uso do *software.* Decidiu o Tribunal nesse sentido:

> We determine that because the circuit court explained that its consideration of the COMPAS risk scores was supported by other independent factors, its use was not determinative in deciding whether Loomis could be supervised safely and effectively in the community. Therefore, the circuit court did not erroneously exercise its discretion. We further conclude that the circuit court's consideration of the read-in charges was not an erroneous exercise of discretion because it employed recognized legal standards[79]. (UNITED STATES, 2013, on-line)

A decisão, que foi questionada junto à Suprema Corte dos Estados Unidos e ainda assim mantida, efetivamente decidiu que como a ferramenta atuou de forma auxiliar, e não exclusiva, junto ao Tribunal, não haveria violação ao devido processo legal.

A partir daí, e analisando outros casos, Larson *et al.* (2016) colheu dados relacionados ao COMPAS e concluiu, em suma, que:

---

[78] Se se trata de um algoritmo produzido pela sociedade privada E-quivant que avalia o risco de reincidência e pericolosidade social de um indivíduo com base em vários dados estatísticos, precedentes judiciais, um questionário preenchido pelo acusado, além de uma série de outras variáveis protegidas pelos direitos de propriedade intelectual por parte da empresa. (tradução nossa)

[79] Determinamos que, como o Tribunal do Circuito explicou que sua consideração sobre as pontuações de risco COMPAS foram apoiadas por outros fatores independentes, seu uso não foi determinante para decidir se Loomis poderia ser supervisionado com segurança e eficácia na comunidade. Portanto, o Tribunal de Circuito não exerceu erroneamente seu poder discricionário. Concluímos ainda que a consideração do Tribunal de Circuito sobre a leitura das acusações não foi um exercício errôneo de discricionariedade porque empregou padrões legais reconhecidos. (tradução nossa)

a) Negros apresentavam um *score* de alto risco de reincidência com maior frequência do que realmente ocorria. Negros condenados que não reincidiram dois anos após a condenação foram quase duas vezes mais severamente classificados que os brancos;
b) Brancos recebiam um *score* de menor risco com mais frequência do que ocorria. Na análise, constatou-se que brancos que reincidiram no período de dois anos após a condenação foram tidos por baixo risco de reincidência em quase duas vezes mais que os negros;
c) Negros possuíam 45% mais chances de receberem uma nota de alto risco;

Os dados apresentados despertaram uma preocupação na comunidade jurídica, uma vez que o *software* se encontra em pleno funcionamento e que estaria sendo endossada pelo Estado uma política claramente discriminatória.

Contudo, o estudo recebeu também diversas críticas quanto à metodologia empregada na análise dos dados.

Flores, Bechtel e Lowenkamp (2016) destacaram que o estudo não foi capaz de demonstrar viés discriminatório do *software*. Ao destacar o já existente panorama de desigualdade no sistema penitenciário, os autores aduzem que o COMPAS apresenta uma análise menos enviesada do que as decisões que levaram ao panorama prisional atual.

Prosseguem os autores indicando erros metodológicos da análise realizada por Larson *et al.*, em que, por exemplo, teriam reduzido os dados a uma análise binária (alto e baixo risco), enquanto que o *software* havia sido programado para uma classificação em três categorias, o que causaria enorme impacto nos resultados.

Finalizam confirmando a higidez do *software e* aduzindo que "*given the higher observed recidivism rates for black defendants, and given the demonstrated validity of the COMPAS, it is nothing short of logical that these defendants evidence higher COMPAS scores*[80]" (FLORES, BECHTEL e LOWENKAMP, 2016, p. 45)

Pertinente observar o argumento trazido pelos autores de que: se comprovado que negros apresentam maiores taxas de reincidência (nisso considerados diversos fatores sociais, econômicos etc.), não é esperado que o *software* apresente tais resultados? Seria coerente exigir que os resultados fossem apresentados de forma

[80] Dadas as taxas de reincidência mais altas observadas para réus negros e dada a validade demonstrada do COMPAS, é lógico que esses réus apresentem pontuações COMPAS mais altas. (tradução nossa)

diversa para fins de política criminal? Na verdade, a ilegalidade consiste em *agravar* a situação dos indivíduos por causa de sua cor, não havendo demonstração de discriminação o fato de que, nos dados analisados, negros cometeram mais crimes.

5.5.2.2. O Caso Catalunha

Um estudo realizado por Tolan *et al.* (2019) analisou e comparou resultados preditivos entre um sistema de inteligência artificial e o instrumento conhecido por *Structured Assessment of Violence Risk in Youth* (SAVRY), que é utilizado para análise de risco de violência no sistema de justiça de menores na Catalunha.

A mencionada análise baseou-se em dois fatores: performance preditiva e justiça. As conclusões puderam demonstrar a ambiguidade no uso de sistemas inteligentes.

O SAVRY é um instrumento de análise de risco que traz um grande envolvimento de profissionais em sua operação, apresentando uma estrutura transparente e interpretável de passos a serem dados durante a sua utilização, que verifica 24 fatores de risco e seis fatores de proteção (TOLAN *et al*, 2019). Ao final, são apresentados os dados e o profissional apresenta a sua avaliação final.

Em relação ao *software* de inteligência artificial, somente foram fornecidos dados demográficos e que diziam respeito ao histórico criminal dos menores, buscando, assim, eliminar possíveis interferências de vieses discriminatórios.

Os resultados demonstraram uma ambiguidade que merece ser mais bem trabalhada pelos desenvolvedores de *softwares* de análise de risco. Para Tolan *et al* (2019), foi possível constatar uma relação inversamente proporcional entre a acurácia e a justiça dos resultados. Segundo o estudo, quanto mais dados possui a inteligência artificial, melhor a previsão de risco, mas também maior a possibilidade de incorrência de decisões injustas ou discriminatórias.

Demonstra-se, portanto, um paradigma que precisa ser sempre analisado quando do desenvolvimento de tais *softwares*, para que sejam implementadas medidas que garantam a justiça da decisão e a acurácia necessária para o aprimoramento do serviço.

5.5.2.3. *Predictive Policing*

A fase investigatória faz parte do sistema penal e também está sendo foco de diversos *softwares* para seu aprimoramento. Nesse contexto, necessário analisar um dos procedimentos que está cada vez mais sendo utilizado e encontra total pertinência com o assunto tratado neste estudo: *predictive policing*.

Predictive policing consiste em uma prática viabilizada pela inteligência artificial capaz de fornecer dados acerca da possibilidade de ocorrência de crimes, utilizando-se de uma grande quantidade de dados disponíveis, normalmente nos bancos de informações de entes públicos, principalmente relacionadas ao cometimento anterior de crimes. Com sua ajuda, a polícia pode estabelecer ações de intervenção ou prevenção em determinadas áreas ou grupos. Assim, "*predictive policing is typically comprised of two elements: a prediction model that uses an algorithm to identify instances of increased crime risk, and an associated prevention strategy to mitigate and/or reduce those risks*[81]" (SAUNDERS, HUNT, HOLLYWOOD, 2016, p. 348).

Portanto, o *predictive policing* consiste em uma análise computacional acerca de dados anteriores relacionados à ocorrência de crimes (como a autoria e o local de ocorrência) que pode auxiliar as forças policiais no patrulhamento e na criação de estratégia para combate à criminalidade. Assim como as ferramentas de análise de risco, o objetivo é tornar a atuação da polícia mais objetiva, evitando práticas discriminatórias.

Um exemplo de tais programas é o *Strategic Subject List* (SSL), que foi utilizado pela polícia da cidade de Chicago, nos Estados Unidos, e que tinha como objetivo prever o risco de ocorrência de crimes com armas (SAUNDERS; HUNT; HOLLYWOOD, 2016). Tal previsão ocorreria com base na análise de diversos dados, como número de prisões anteriores por crimes graves, idade da última prisão, afiliação com gangues etc.

De acordo com DaViera *et al.* (2023), a eficácia do programa nunca foi efetivamente atestada, havendo estudos que apontaram o *software* como melhor ferramenta para análise preditiva de crimes, enquanto outros não apontaram qualquer redução em incidentes com armas de fogo.

---

[81] O policiamento preditivo é normalmente composto por dois elementos: um modelo de previsão que usa um algoritmo para identificar instâncias de maior risco de crime e uma estratégia de prevenção associada para mitigar e/ou reduzir esses riscos (tradução nossa)

Browing e Arrigo (2020) apontam a existência de estudos que comprovariam sucesso na prática do *predictive policing*, mas ao mesmo tempo atestam que tais estudos são bastante escassos, não havendo efetiva demonstração de sua eficácia.

O principal problema com o *predictive policing* diz respeito aqueles grupos que são superexpostos a dados do sistema prisional. Segundo DaViera *et al* (2023), não-brancos são presos em uma taxa maior e a probabilidade de serem presos é superior a dos brancos. Além disso, historicamente os grupos minoritários recebem maior policiamento, são segregados e desprovidos de recursos, o que impacta na coleta dos dados (BROWING; ARRIGO, 2020).

Tal fato gera uma superexposição de grupos minoritários, levando a uma super-representação no código dos programas, impedindo assim que haja uma análise igualitária.

Poder-se-ia partir de uma premissa com base nesses dados de que tais grupos seriam mais propensos ao cometimento de crime. Contudo, segundo Browing e Arringo (2020), as taxas de crimes cometidos por negros não são tão altas quanto os dados sugerem e que são, em alguns casos, iguais a dos brancos.

Sendo mais policiados, os grupos minoritários são presos com mais frequência, mais condenações são aplicadas e as prisões acabam sendo proporcionalmente maiores contra si. Cria-se um *loop* que somente tem a capacidade de aumentar as diferenças entre os grupos dentro do sistema penal, dando início ao que O'neal chama de "ciclo malicioso de feedback" (2020, p.89). A cada prisão, novos dados são gerados e tais dados acabam servindo de suporte para ratificar a suposta eficácia da política implementada.

Dessa forma, por promover uma supervigilância contra determinados grupos, há forte oposição à sua utilização. Para Browing e Arrigo (2020), como o sistema criminal é discriminatório, os dados por ele gerados também serão. Assim, conclusões retiradas com bases nesses dados enviesados não deveriam ser fonte de informação para formulação de políticas públicas de policiamento. Somente com uma verdadeira reforma no sistema - e em toda a sociedade – é que poderia ser falado em uma coleta de dados limpa, sem vieses discriminatórios que poderiam ser utilizados pela máquina para garantir atuação igualitária da polícia.

Para Browing e Arrigo (2020), em conclusão, evidências indicam que o *predictive police* seria apenas uma nova forma de práticas discriminatórias da polícia,

bem como que teria potencial para vulnerabilizar, ainda mais, minorias e comunidades de baixa renda.

Demonstrado, portanto, o funcionamento destes softwares de decisão automatizada, resta claro o grande potencial de abuso e agravamento dos quadros de marginalização em detrimento destes grupos mais vulnerabilizados. Tal agravamento, aliás, já pode ser constatado em diversas esferas, como demonstrado. Assim, o que pode ser feito para se evitar sua ocorrência? É esse questionamento, decorrente da própria questão problema do presente estudo, que será debatido no próximo tópico, com uma análise acerca do desenvolvimento de *softwares* originalmente pautados na proteção dos direitos humanos.

5.6. Algoritmo e Direitos Humanos

Tendo sido demonstrado o funcionamento de casos em matéria penal, inclusive com conclusões que indicam efetivos casos de violação de direitos humanos, resta importante estabelecer como poderiam ser evitadas tais práticas.

No Capítulo 3, onde se discorreu sobre os princípios éticos que devem ser implementados para que haja um funcionamento saudável de *softwares* de inteligência artificial, restou estabelecido que os direitos humanos são os melhores parâmetros para serem utilizados pelos desenvolvedores, garantindo uma preocupação e direcionamento das decisões à proteção das garantias do indivíduo desde a sua concepção. Contudo, resta estabelecer como poderiam efetivamente implementar tais critérios.

Há, contudo, certa dificuldade na escolha deste critério, uma vez que, muito embora tenha sido estabelecida sua universalidade pela ONU na Declaração Universal dos Direitos Humanos, ainda existe controvérsia acerca da extensão ou alcance de determinados direitos, como, por exemplo, o tratamento destinado às mulheres em determinados países da África e Oriente Médio.

Neste sentido, Hars (2024) estabelece cinco principais problemas relacionados à implementação de tais critérios em ferramentas de IA que serão explicitados a seguir.

O primeiro diz respeito a *o que* transferir. Assim, considerando que há uma variedade de direitos humanos, previstos em diversos diplomas e que existe, ainda, certa discussão acerca de seu alcance e amplitude, torna-se difícil indicar, para a

máquina, qual norma deve ser seguida ou, ainda, qual teria prevalência sobre outra em caso de conflito.

O segundo ponto é o *como*. Considerando que a legislação internacional de direitos humanos possui características de comandos programáticos, com uma construção voltada para o fomento de compromissos dos países aderentes, as normas acabam se aperfeiçoando no âmbito do direito local, havendo grande possibilidade de existência de conflitos – ainda que aparentes - entre as legislações aplicadas. Além disso, analisando apenas tratados internacionais de linguagem aberta, o próprio *software* terá a necessidade de tornar aquela linguagem "binária", o que poderá implicar em uma conclusão diversa daquela prevista pelo legislador.

Em *terceiro,* o autor suscitou a discussão sobre se *seria possível incorporar os instrumentos de direito internacional na programação do software*. Para Hars (2024), o direito internacional possui complexas relações e institutos que, atualmente, são cuidadosamente analisados pelas Cortes e advogados, mas podem surgir como uma dificuldade de implementação para uma abordagem voltada aos direitos humanos na inteligência artificial. Cita, então, como exemplo, a reserva realizada por alguns países quando da assinatura de tratados.

Segundo a Convenção de Viena Sobre o Direito dos Tratados (CVDT), a reserva significa um:

> unilateral statement, however phrased or named, made by a State, when signing, ratifying, accepting, approving or acceding to a treaty, whereby it purports to exclude or to modify the legal effect of certain provisions of the treaty in their application to that State[82]"(ONU, online)

Assim, analisando, por exemplo, A Convenção Sobre os Direitos das Crianças, é possível notar que há ressalva por parte de diversos países muçulmanos relacionada à liberdade de pensamento e religião das crianças. O Afeganistão assim se reservou, ao estabelecer que:

> The Government of the Republic of Afghanistan reserves the right to express, upon ratifying the Convention, reservations on all provisions of the Convention

[82] declaração unilateral, qualquer que seja a sua formulação ou denominação, feita por um Estado, ao assinar, ratificar, aceitar, aprovar ou aderir a um tratado, pela qual pretende excluir ou modificar o efeito jurídico de certas disposições do tratado na sua aplicação a esse Estado (tradução nossa). Disponível em: https://legal.un.org/ilc/texts/instruments/english/conventions/1_1_1969.pdf

> that are incompatible with the laws of Islamic Shari'a and the local legislation in effect[83] (ONU, online)

Diante de tal cenário de ressaltas relacionadas aos direitos estabelecidos, é possível perceber que a incorporação do direito aplicável passa a ser ato extremamente complexo para os sistemas automatizados.

O quarto ponto de dificuldade indicado seriam as *diferenças regionais*. Além do já discutido acerca das diversas interpretações e a própria extensão dos direitos humanos nos mais diversos territórios, é importante ter em mente que há organizações regionais responsáveis pela aplicação do direito no seu âmbito jurisdicional. Cortes como o Conselho da Europa ou a Organização dos Estados Americanos (OEA) produzem constantemente decisões relacionadas aos tratados regionais, formando um corpo de normas com incidência específica na respectiva região. Tal fato aumenta ainda as diferenças nas decisões entre países e blocos de regiões diversas, imponto uma maior necessidade de direcionamento e *expertise* à máquina.

Por fim, o último entrave seria o próprio *desenvolvimento* constante dos direitos humanos. Hars (2024) destaca que o desenvolvimento dos direitos humanos é algo relativamente novo, decorrendo dos séculos 17 e 18, exemplificando grandes e recentes avanços como a abolição da escravidão e a gradual diminuição da discriminação contra mulheres. Dessa forma, considerando estes avanços ainda constantes, poderia surgir uma dificuldade para que o algoritmo se adaptasse diante da ausência de dados pretéritos sobre como se comportar.

Inobstante, apontados os problemas, é necessário trabalhar em sua resolução, considerando que o fenômeno da inteligência artificial é algo inevitável.

Quanto aos dois primeiros argumentos, o *o que* e o *como*, é possível vislumbrar uma abordagem fragmentada, utilizando-se do texto de tratados consolidados como fonte primária. Sugere Hars:

> One of the solutions is to start small, select an international treaty which has a judicial mechanism such as the European Convention on Human Rights and extensive judicial practice such as that of the European Court of Human

[83] O Governo da República do Afeganistão reserva-se o direito de expressar, após ratificar a Convenção, reservas sobre todas as disposições da Convenção que sejam incompatíveis com as leis da Sharia Islâmica e a legislação local em vigor. (tradução nova) Disponível em: https://treaties.un.org/Pages/ViewDetails.aspx?src=TREATY&mtdsg_no=IV-11&chapter=4&clang=_en

> Rights, and use deep learning to predict human rights violations[84]. (HARS, 2024, p. 130)

A partir da inclusão de tratados-chave, seria possível ampliar a abrangência, incluindo-se do direito regional e, posteriormente, realizando a junção destes subgrupos com os demais países ou regiões.

Quanto ao problema dos mecanismos típicos de direito internacional, Hars (2024) indica que a capacidade da máquina, se bem trabalhada, poderia superar a maioria destas dificuldades, desde que seguindo uma abordagem pedagógica que facilite uma utilização em escala dos termos e institutos. Contudo, destaca a dificuldade em se obter respostas para alguns conflitos mesmo se tais decisões fossem tomadas exclusivamente por humanos:

> It is also unclear if competing norms or those that can collide, for instance the freedom of speech and the freedom of religion or personal self-determination, and the rights to life could be resolve, even if it has some judicial practice[85] (HARS, 2024, p.131).

As dificuldades encontradas diante dos múltiplos sistemas regionais podem ser solucionadas através de um sistema de prevenção daquelas Cortes mais próximas da questão, possibilitando que aqueles com melhor condições de instruir os casos possam ser a fonte do direito primária.

Por fim, quanto a constante evolução dos direitos humanos, é imprescindível contar com a participação humana no constante aprimoramento da máquina. Ao contrário dos homens, o *software* de inteligência artificial não evolui, apenas reproduz comportamentos de acordo com seu "treinamento". Dessa forma, os homens tem papel essencial para que seja estabelecido uma rotina de constante renovação do arcabouço jurídico-normativo do algoritmo, possibilitando que este acompanhe a evolução social.

---

[84] Uma das soluções é começar pequeno, selecionar um tratado internacional que tenha um mecanismo judicial como a Convenção Europeia dos Direitos Humanos e uma prática judicial abrangente como a do Tribunal Europeu dos Direitos Humanos, e usar a aprendizagem profunda para prever violações dos direitos humanos. (tradução nossa)

[85] Também não é claro se normas concorrentes ou que podem colidir, por exemplo, a liberdade de expressão e a liberdade de religião ou a autodeterminação pessoal, e os direitos à vida, pudessem ser resolvidos, mesmo que houvesse alguma prática judicial (tradução nossa).

Exploradas nuances teóricas sobre a efetivação de um algoritmo fundado em direitos humanos, seria possível vislumbrar a existência de algum construído nesses padrões?

Em primeiro lugar, é preciso estabelecer que já constam decisões que afastam a incidência de algoritmos pela violação aos direitos humanos, o que consubstancia certa prevalência de tal parâmetro com fundamento em direito internacional.

Um exemplo que pode ser destacado foi o do caso relacionado ao System Risk Indication (SyRI) utilizado pelo Governo Holandês, que tinha como finalidade identificar fraudes no sistema de assistência social. Segundo Rachovitsa e Johann (2022), o algoritmo utilizava-se do banco de dados estatal e de suas agências para fornecimento de informações como aquelas relacionadas aos impostos, seguro saúde, endereço entre outros. Dessa forma, gerava perfis de risco que indicavam possíveis casos suspeitos de irregularidades ou fraudes.

O uso do *software* foi, então, contestado, sob o fundamento de violação ao direito à privacidade, conforme previsto no art. 8 da Convenção Europeia dos Direitos Humanos.

Submetido à análise do Poder Judiciário, a Corte Distrital de Haia assim decidiu:

> Considering the lack of respect for the principle of transparency in conjunction with inadequate safeguards, the Court thus found a violation of Article 8 of the ECHR and declared that the SyRi legislation 'ha[s] no binding effect with respect to [the admissible claimants] and on the individuals whose interests these parties promote[86] (RACHOVITSA; JOHANN, 2022, p.7)

Esse julgamento, ainda Segundo Rachovitsa e Johann (2022), serviu como marco na proteção e promoção de um bem estar digital, promovendo o início das iniciativas de proteção dos cidadãos em face do uso inadequado de algoritmos de inteligência artificial que violassem direitos humanos.

Além de demonstrar a possibilidade de controle judicial sobre algoritmos que violam direitos humanos, já é possível identificar praticas de obediência a tais princípios quando do desenvolvimento do *software*.

---

[86] Considerando a falta de respeito pelo princípio da transparência em conjunto com salvaguardas inadequadas, o Tribunal concluiu, assim, pela violação do artigo 8.º da CEDH e declarou que a legislação SyRi não tem efeito vinculativo relativamente aos [requerentes admissíveis] e aos indivíduos cujos interesses estas partes promovem. (tradução nossa)

O Public Safety Assessment (PSA) foi um *software* de inteligência artificial desenvolvido com a finalidade de se analisar o nível de confiabilidade dos réus de processos criminais no tocante à reincidência ou cumprimento de obrigações impostas pela Corte Judicial.

Segundo DeMichele *et al* (2020), para seu desenvolvimento, utilizou-se dados de mais de 1.5 milhões de casos de mais de 300 jurisdições americanas, e optaram os desenvolvedores por deixar de lado fatores demográficos relacionados à raça e gênero, assim como as condições econômicas variáveis, como a estabilidade residencial, educação e emprego, assim agindo para evitar que se estabeleçam *proxies* ou *bias* em detrimento dos pobres ou minorias.

Observa-se, portanto, uma preocupação voltada para o respeito dos padrões éticos durante o próprio desenvolvimento do *software*, o que viabiliza um funcionamento hígido e, provavelmente, menos suscetível à violações de direitos humanos.

Tendo analisados os dados dos casos submetidos ao PSA, DeMichele *et al* (2020) puderam observar que os *scores* fornecidos pela *software* foram, em geral, similares para todas as raças. Utilizando-se de padrões estabelecidos por entidades americanas de referência (como por exemplo a *American Educational Research Association, American Psychological Association* e a *Society for Industrial and Organizational Psychology*) , os autores foram capazes de estabelecer que o *software* foi capaz de atuar dentro dos padrões esperados para a justiça criminal, não tendo encontrado diferença significante nas previsões realizadas para negros e brancos.

O estudo realizado conseguiu demonstrar como a dedicação desde o desenvolvimento do algoritmo ao cumprimento de princípios éticos e boas escolhas no tocante ao funcionamento do sistema podem gerar bons resultados e viabilizar o uso de inteligência artificial na justiça criminal.

Demonstrado, portanto, que existe a possibilidade de efetiva implementação dos mecanismos de controle responsáveis pela proteção dos direitos humanos, há a necessidade de tornar impositiva tal observância aos desenvolvedores. Se não existe uma obrigação jurídica que efetivamente imponha às empresas este compromisso, as chances de que as violações continuarem a existir são grandes, dado o histórico observado. Tal relação, entre a regulamentação da atividade e o *compliance* por parte das corporações será mais bem explorada no capítulo seguinte.

6. PROPOSTAS DE RESOLUÇÃO DOS PROBLEMAS E PRINCÍPIOS GERAIS PARA REGULAMENTAÇÃO

Decorridas as análises dos capítulos precedentes, pode-se estabelecer uma posição em que é possível estipular critérios e princípios gerais nos quais devem se pautar as legislações que se propõem a regulamentar o tema. Contudo, mesmo sendo possível, seria adequado? E qual seria a forma ideal de se proceder o estabelecimento de tais regras?

Neste capítulo serão abordadas questões relacionadas à teoria da regulamentação, analisando a necessidade de se estabelecer um regramento pelo Estado para observação por parte dos envolvidos no processo de desenvolvimento de softwares de inteligência artificial e, caso opte pela fixação de regras, quais critérios deveriam ser buscados pela legislação.

Uma vez estabelecidos os critérios quanto à regulamentação, será analisada a matéria a ser regulamentada, ou seja, o que busca o ente regulador ao estabelecer os limites e normas de desenvolvimento da inteligência artificial no Poder Judiciário.

6.1. O Estado Regulador

Com o fim da segunda guerra mundial, estabeleceu-se em grande parte dos países um regime de participação ativa do Estado na busca pela garantia de direitos de seus cidadãos, o conhecido *Welfare State*. Para Yeung (2010), o consenso social de que caberia ao Estado garantir o planejamento e estabilidade do mercado e garantia do bem-estar, suprindo as necessidades básicas de todos, fez com que muitos países passassem a controlar recursos e serviços.

Continua a autora ao afirmar que tal panorama começou a mudar nos anos 80, quando muitos Estados industrializados iniciaram um programa de privatizações e, por consequência, um programa de regulamentação dos serviços terceirizados. Tal processo trouxe três principais mudanças: a fragmentação estatal, com o fornecimento de serviços públicos através de diversos órgãos e agentes; a transferência ao setor privado (ou não-estatal) no fornecimento de serviços essenciais; e a reconfiguração da missão como regulador, ao invés do efetivo fornecedor dos serviços e bens.

Observa-se, portanto, que a transição para um Estado regulatório ocorreu em meio ao reconhecimento de que seu papel de garantidor do bem-estar social não estaria funcionando a contento, sendo necessária uma nova organização. Passa-se, então, ao que Karen Yeung chama de "*governs at a distance*" (YEUNG, 2010, p. 3).

Um regime regulatório pode ser definido como "*a system of control which may comprise many actors, but within which it is possible to identify standards of some kind, ways of detecting deviation from the standards, and mechanisms for correcting such deviations*[87]" (SCOTT, 2010, p. 2).

Por outro lado, a regulamentação em si pode ser definida de diversas formas, a depender do sentido que se busca empregar. Para Baldwin, Cave e Lodge (2012), a regulamentação pode ser entendida como um conjunto de comandos, onde um conjunto de regras são aplicadas a um determinado grupo ou atividade. Pode ser entendida também como um estado de deliberada influência, em um sendo mais amplo, dizendo respeito a todas as ações estatais voltadas para influenciar empresas ou comportamentos sociais. E, por fim, pode-se também definir o conceito como todas as formas de influência social ou econômica.

Para Baldwin, Cave e Lodge (2012), diversos são os fundamentos pelo qual a regulamentação poderia ser exigida, como por exemplo: falhas de mercado, monopólios, inadequação de informações, continuidade do serviço, comportamento anticompetitivo ou precificação predatória, utilidade pública, desproporcional poder de barganha, entre outros. Assim, percebe-se uma variedade de motivos pelos quais a regulamentação pode se fazer necessária.

6.1.1. Teorias da Regulamentação

As teorias da regulamentação buscam trazer as justificativas que levam os entes a fixar regras para o desenvolvimento de certas atividades e sob quais argumentos elas deveriam se estabelecer. Mais do que somente apontar possíveis causas, as teorias buscam explicar teoricamente – e de forma abrangente – a necessidade de regulamentação. Nesse sentido, torna-se importante analisar tal aspecto para que se possa identificar sob qual prisma se justificaria uma

[87] Um sistema de controle que pode compreender vários atores, mas dentro do qual é possível identificar padrões, formas de detectar desvios dos padrões e mecanismos para corrigir tais desvios (tradução nossa)

regulamentação da inteligência artificial pelos países, notadamente considerando as manifestações no sentido de que uma autorregulamentação do setor poderia ser suficiente.

A doutrina clássica distingue as teorias da regulamentação em duas principais escolas: escola econômica da regulamentação e escola do interesse público. Por certo que existem diversas outras classificações, como costuma ocorrer em casos semelhantes, contudo tal divisão foi adotada de forma a melhor ilustrar a dicotomia apresentada entre o interesse público e o privado que se aplica perfeitamente ao estudo em tela.

Para Salomão Filho (2008), a escola econômica da regulamentação traz a noção de que a regulamentação teria como único objetivo a substituição ou correção do mercado, afastando qualquer motivação relacionada ao interesse público. Assim, a regulamentação teria a função precipuamente comercial, sendo bastante utilizada na fundamentação legislativa relacionada aos monopólios, por exemplo. Nesse tipo de mercado, aponta Ogus (2004), a regulamentação teria como principal função promover uma espécie de substituto à competição, simulando a concorrência ocorrida em mercados de naturezas diversas.

Em contrapartida, tratando sobre a escola do interesse público, define o autor:

> The public interest justifications for social regulation, which deals with such matters as health and safety, environmental protection, and consumer protection, tend to centre on two types of market failure. First, individuals in an existing, or potential, contractual relationship with firms supplying goods or services often have inadequate information concerning the quality offered by suppliers; in consequence, the unregulated market may fail to meet their preferences. Secondly, even if this information problem does not exist, market transactions may have spillover effects (or externalities) which adversely affect individuals who are not involved in the transactions[88] (OGUS, 2004, p.4)

Assim, o conceito de escola do interesse público é amplo e está ligado a diversas formas em que o consumidor (ou cidadão) encontra-se ameaçado em seus

[88] As justificativas de interesse público para a regulação social, que trata de questões como saúde e segurança, proteção ambiental e proteção do consumidor, tendem a centrar-se em dois tipos de falhas de mercado. Primeiramente, os indivíduos que mantêm uma relação contratual existente ou potencial com empresas fornecedoras de bens ou serviços dispõem frequentemente de informações inadequadas relativamente à qualidade oferecida pelos fornecedores; em consequência, o mercado não regulamentado pode não conseguir satisfazer as suas exigências. Em segundo lugar, mesmo que este problema de informação não exista, as transações de mercado podem ter efeitos de repercussão (ou externalidades) que afetam negativamente os indivíduos que não estão envolvidos nas transações. (tradução nossa)

direitos, seja pela não prestação do serviço, seja pela sua execução sem as devidas cautelas.

Ogus (2004) aponta a regulamentação da informação como um dos tipos de regulamentação social. E, como já visto, a ausência de clareza quanto às informações prestadas (ou utilizadas), é um dos grandes problemas relacionados à utilização de algoritmos. Para o autor, há direito do cidadão em obter informações precisas quando há utilização de dados pessoais (como registros médicos, por exemplo) e quando houver envolvimento de tomadas de decisões que impactem em políticas públicas, como corolário do direito democrático de participação. Há, portanto, clara incidência de ambas as hipóteses do estudo em tela.

Como os *softwares* de inteligência artificial têm se utilizado de dados pessoais e privados dos usuários sem que eles saibam, e, muitas das vezes para tomada de decisões políticas, haveria a necessidade, por ambos os motivos, de se estabelecer uma regulamentação eficaz da atividade, do ponto de vista da escola do interesse público.

É importante observar que o conceito de regulamentação e sua utilização, segundo a doutrina clássica, sempre se relacionou apenas com questões econômicas, notadamente evidenciando o conceito de falha de mercado. Assim, trata-se somente de uma tentativa de se controlar a atividade econômica em virtude da constatação de eventual necessidade social. Contudo, Prosser (2010) traz ao tema a questão social, apontando também a necessidade de se regulamentar determinada matéria para proteção de direitos humanos ou para desenvolver a solidariedade social, como por exemplo regramentos de proteção ambiental.

Nesse sentido, é necessário observar que: "*social objectives, moreover, are sometimes furthered by regulating even where this involves overruling the preferences of market players and acting paternalistically*[89]" (BALDWIN; CAVE; LODGE, 2012, p. 23). É neste contexto que se insere o objeto de estudo da presente tese. Percebe-se, claramente, que não haveria, a princípio, interesse da indústria em se impor limites no desenvolvimento de novos produtos. Contudo, passa a regulamentação a reger tais preferências no intuito de proteger os cidadãos contra violação de direitos fundamentais.

[89] Além disso, os objetivos sociais são por vezes promovidos através da regulamentação, mesmo quando isso envolve anular as preferências dos envolvidos e agir de forma paternalista. (tradução nossa)

6.2. Autorregulamentação

As análises acerca da regulamentação normalmente encontram-se em uma balança que compara os níveis de liberdade e controle. A autorregulamentação, nesta escala de forças, tende ao alto nível de liberdade, eis que o Estado renuncia a todo o controle para delegar o regramento da atividade em estudo para as próprias empresas.

A autorregulamentação pode ser observada quando "*a group of firms or individuals exerts control over its own membership and their behavior*[90]" (BALDWIN; CAVE; LODGE, 2012, p.137). No mesmo sentido, pode ser entendida como "*any system of regulation in which the regulatory target – either at the individual-firm level or sometimes through an industry association that represents targets – imposes commands and consequences upon itself*[91]". (COGLIANESE; MENDELSON, 2010, p. 3).

Considerando que há uma clara possibilidade de conflitos de interesses no caso, por qual motivo iria-se favorecer um sistema de autorregulamentação? Para Baldwin, Cave e Lodge (2012) a principal vantagem reside na expertise das empresas e no ganho de eficiência.

No tocante à expertise, deve-se considerar que as empresas possuem, de forma geral, melhores condições de analisar as circunstâncias dos fatos a serem regulamentados, exatamente por terem como atividade principal o seu desenvolvimento e comercialização. Nesse sentido, ao delegar a regulamentação aos entes públicos, haveria uma grande possibilidade de desconexão entre o que seria possível e viável entre o estabelecido na regra. Trata-se do que Baldwin, Cave e Lodge (2012) chamam de eficácia regulatória, ou seja, deve-se propor uma legislação que possa efetivamente ser atendida pelos envolvidos, sem que haja um estabelecimento de obrigações inalcançáveis.

Sobre esse ponto, destacam ainda Coglianese e Mendelson:

> Conventional means or performance regulation usually requires information about the risks created by certain products or modes of production. Regulators

[90] Um grupo de empresas ou indivíduos exerce controle sobre seus próprios membros e seu comportamento (tradução nossa)

[91] qualquer sistema de regulação em que o alvo da regulamentação – seja no nível da empresa individual ou às vezes através de uma associação industrial que os representa– impõe comandos e consequências sobre si mesmo (tradução nossa)

> need to know the magnitude of potential harm and the probability of that harm occurring. Especially in newly developing markets, such as with nanotechnology at present, regulators are likely to find themselves at a significant information disadvantage compared to the industries that they oversee[92] (COGLIANESE; MENDELSON, 2010, p.3)

Assim, os autores chamam atenção para um ponto sensível relacionado ao presente estudo. A tecnologia relacionada à inteligência artificial é bastante nova e alvo de frequentes mudanças e evoluções. Assim, destacam que reguladores teriam maiores dificuldades em estabelecer regras acerca do tema em comparação com as próprias empresas.

Quanto à eficiência, Baldwin, Cave e Lodge (2012) apontam que as empresas possuem melhores condições de exercer controle, notadamente pela informalidade que se estabeleceria através da autorregulamentação voluntária. Destacam, ainda, que tal controle delegaria os custos da atividade aos entes privados, eximindo o erário público de arcar com tais valores.

Virgínia Dignum (2020) ressalta dois importantes pontos a serem considerados em favor da autorregulamentação. O primeiro é que esta pode impactar no desenvolvimento de novas tecnologias e o segundo é que, diante da enorme variedade de áreas abrangidas pela inteligência artificial, seria difícil obter uma legislação adequada e abrangente para todas as áreas. A referida autora cita como exemplo:

> Consider the case in which legislation will restrict the use of data and demand explanation of all results achieved by an AI system. These requirements probably mean that many of the current approaches, based on neural networks and deep learning, are not able to meet these demands. This can be seen as a limitation on the use of AI and be approached with complains and a refusal to comply, claiming economic losses and a delay on development[93] (DIGNUM, 2020, p. 224).

---

[92] Os meios convencionais ou a regulação do desempenho geralmente exigem informações sobre os riscos criados por determinados produtos ou modos de produção. Os reguladores precisam saber a magnitude do dano potencial e a probabilidade desse dano ocorrer. Especialmente nos mercados em desenvolvimento, como acontece atualmente com a nanotecnologia, os reguladores provavelmente se encontram numa desvantagem significativa em termos de informação, em comparação com as indústrias que supervisionam. (tradução nossa)

[93] Consideremos o caso em que a legislação restringirá a utilização de dados e exigirá explicação de todos os resultados alcançados por um sistema de IA. Estes requisitos provavelmente significam que muitas das abordagens atuais, baseadas em redes neurais e *deep learning*, não são capazes de satisfazer estas exigências. Isto pode ser visto como uma limitação ao uso da IA e dar ensejo a reclamações e recusa de cumprimento, alegando perdas econômicas e um atraso no desenvolvimento. (tradução nossa)

Por outro lado, Ogus (2004) destaca que as críticas tradicionais se pautam, basicamente, no alto grau de monopólio no controle da atividade, bem como no baixo grau de responsabilização ou controle público. De fato, é possível vislumbrar diversos cenários em que uma regulamentação estabelecida por empresas para cumprimento por elas proteja insuficientemente a sociedade contra eventuais danos à direitos humanos. Considerando que as empresas têm como principal objetivo o lucro, delegar a tais entes esta responsabilidade pode gerar além de uma desconfiança pública, um panorama de proteção ineficaz.

O'neal exemplifica casos em que houve graves violações em virtude da ausência de regulamentação:

> Na ausência de regulamentações de saúde ou segurança, as minas de carvão eram armadilhas mortais. Apenas em 1907, 3.242 mineiros morreram. Trabalhadores em frigoríficos trabalhavam de doze a quinze horas por dia em condições insalubres e muitas vezes enviavam produtos intoxicados. A Armour and Co. distribuía latas de carne podre às toneladas para as tropas do exército, usando uma camada de ácido bórico para mascarar o fedor...Claramente, o livre mercado não conseguia controlar seus excessos. Então depois que jornalistas como Ida Tarbell e Upton Sinclair expuseram esses e outros problemas, o governo interveio. Estabeleceu protocolos de segurança e inspeções sanitárias para alimentos... (O'NEAL, 2020, p.190)

Assim, diversos são os argumentos que depõem contra a autorregulamentação e, estando diante de uma questão que afeta a liberdade do indivíduo, cautelas maiores devem ser empregadas.

Em uma análise acerca dos pontos trazidos pelos teólogos favoráveis e contrários à autorregulamentação, pode-se resumir que:

> The key consideration may be whether the expertise and efficiency gains to be achieved by self-regulation do out-balance any weakness in mandate definition, accountability, and fairness that will remain after appropriate steps have been taken to ward off criticisms on these fronts[94] (BALDWIN; CAVE; LODGE, 2012, p. 145-146)

Contudo, é necessário observar que a matéria cuja regulamentação se está estudando afeta de forma grave diversos aspectos da vida do cidadão. A possibilidade de atraso nos avanços científicos deve ser analisada em confronto com a garantia dos

[94] A principal consideração pode ser se os ganhos de experiência e eficiência a serem alcançados pela auto-regulação compensam qualquer fraqueza na definição do mandato, na responsabilização e justiça que permanecerão depois de terem sido tomadas medidas apropriadas para evitar críticas nestas frentes. (tradução nossa)

direitos individuais, que não podem ser mitigados em favor de interesses empresariais. Há possibilidade, inclusive, como visto no caso do *Cambridge Analitica*, de interferências no processo democrático dos países. Então, parece haver uma preponderância de tais garantias sobre qualquer argumento relacionado a ganhos com expertise, eficiência ou avanço científico. A legislação deve ter como prioridade a proteção do cidadão e de seus direitos mais caros, restando relegada a análise econômica a um segundo plano.

6.3. Regulamentação e Democracia

Outro aspecto que fala em favor da necessidade de regulamentação diz respeito ao fator democrático da decisão.

Ponto comum em diversas Constituições mundiais (art. 1º, parágrafo único da Constituição Federal do Brasil e art. 1 da Constituição Italiana, por exemplo), a outorga de Poder ao povo, como destinatário e possuidor deste, faz refletir acerca da necessidade de autorização expressa para que possam os s*oftwares* de inteligência artificial operarem dentro do Poder Judiciário.

De acordo com a divisão clássica de Aristóteles, posteriormente aprimorada por Montesquieu, o Poder Estatal é dividido em três, recaindo, de um modo geral, sobre o Executivo a administração da coisa pública, o Legislativo a criação de leis e o Judiciário o papel de julgador e de resolução dos conflitos.

Atualmente, de uma forma geral, os países têm adotado como regra a positivação do direito, que nada mais é, segundo Streck (2017), que o estabelecimento do arcabouço jurídico por uma autoridade legitimada. Neste sentido, ao juiz cabe o cotejo dos fatos apresentados com ordenamento jurídico vigente, em um exercício de adequação.

Assim, a perspectiva de positivação do direito vai ao encontro da legitimação da própria lei, estando a vontade do povo devidamente respeitada pelos instrumentos jurídicos previstos nas Constituições nacionais através dos instrumentos de participação democrática.

Além disso, importante mencionar que deve haver correlação entre o aspecto social e formal da norma. Para que haja, portanto, validade jurídica, a norma deve encontrar eco nas vontades advindas de sua fonte de poder: o povo.

A aferição desta vontade pode se dar de diversas formas, seja através de instrumentos de democracia participativa direta (como o referendo ou a consulta pública) ou indiretamente, com a escolha de representantes que possam defender os seus interesses. Somente assim, é possível haver uma legitimação dos desejos sociais tornados lei.

A partir daí, diante do estabelecimento de critérios e regras que atendam aos anseios sociais, com a sua consequente positivação na norma, a regra passa a ser válida e passível de aplicação forçada pelo Poder Judiciário.

Entretanto, para Bezerra Neto, o agente da decisão, ou seja, aquele que a profere é "o humano, o ser que tem consciência, capacidade de compreensão, interpretação, decisão, que vive em sociedade e que, portanto, está inserido em determinada tradição historicamente construída" (Bezerra Neto, 2018, posição 468).

O autor, portanto, deixa claro que o agente da decisão judicial é um ser humano, ressaltando seus aspectos subjetivos como consciência, capacidade de compreensão e sua vivência social. Ainda que não tenha sido aventada a possibilidade de decisão automatizada em sua teoria da decisão judicial, os elementos por ele trazidos deixa clara a necessidade de certa subjetividade pelo julgador e a sensibilidade de se perceber o contexto social em que ocorreu o caso concreto.

Reforça ainda:

> O que importa considerar acerca do agente da decisão é que ele seja racional no sentido mais geral, ou seja, no sentido do 'logos' grego, significando que, a partir do seu acervo intelecto-cultural, bem como do seu horizonte interpretativo, é capaz de empreender o raciocínio (intelectivo e discursivo) e formular proposições. (Bezerra Neto, 2018, posição 488).

Ainda nesta linha, mas tratando acerca da própria valoração da prova, aduz Paulo de Barros Carvalho:

> Ato psicológico de valorar, segundo o qual atribuímos a objetos, aqui considerados em toda a sua plenitude semântica, qualidades positivas ou negativas. E o que nos dá acesso ao reino dos valores é a intuição emocional, não sensível nem a intelectual (Carvalho, 2013, p. 176)

Considerando as citações acima, será possível afirmar que as máquinas podem, então, ser capazes de substituir o homem na tomada de decisões judiciais? E mais: estão legalmente e socialmente habilitadas para tanto?

Acerca do tema, importante mencionar a percepção de Byung-Chul Han (2022) acerca do regime de informação. Para ele, em um regime de informação, são os dados e o seu processamento por algoritmos que determinam os processos sociais. Nesse ponto, eles buscam calcular tudo, o que é e o que virá a ocorrer (em um processo chamado dataísmo, como sugere o autor). Trata-se de uma transição entre a sociedade da comunicação a sociedade da informação, nas palavras de Giacomelli (2019).

Ainda para Han (2022), a rede digital processa o envio de informações de forma diferente do que sempre ocorreu. Não há uma confluência de ideias para formação de uma esfera pública de discussão, mas sim a sua produção em espaços privados para outros espaços privados. Dessa forma, inexiste qualquer ação comunicativa entre as pessoas, o que impede o próprio debate democrático.

Para Hannah Arendt, a democracia como práxis discursiva consiste na observância da opinião do outro, na medida em que "formo uma opinião ao observar determinada questão a partir de diferentes pontos de vista, ao imaginar os pontos de vista dos ausentes e, assim, correpresentá-los." (ARENDT, 2011, p. 342)

Esta também é a conclusão de Habermas acerca de uma ação comunicativa:

> "o conceito de ação comunicativa compele a se observar tanto como falantes quanto como ouvintes os agentes que se referem a algo do mundo objetivo, social ou subjetivo e, ao fazer isso, elevam mutuamente as reivindicações de validade que podem ser aceitas ou questionadas" (HABERMAS, 2020, p. 588)

Tendo em vista esta característica, é possível se falar, como aduz o autor, em uma "crise de escuta atenta" (HAN, 2022, p. 35), em que não se considera a opinião do outro, mas somente aquelas informações que circularam em uma "bolha" personalizada de informações.

Mas qual seria a relevância de tal construção teórica para o tema em análise?

É importante mencionar que a atualidade marcha para uma percepção de que a racionalidade digital seria superior a racionalidade humana/comunicativa. Assim, caminha-se para uma sociedade sem política, em que os dados determinariam, supostamente através de decisões melhores, os rumos da sociedade.

Sobre tal fato, Pentland prediz:

> Já dentro de poucos anos teremos à disposição dados abrangentes praticamente sobre o comportamento da humanidade inteira – e cada vez mais. [...] E assim que tivermos desenvolvido uma visualização precisa do modelo da vida humana, poderemos esperar compreender e dirigir nossa sociedade moderna de um modo melhor ajustado à nossa rede complexa de humano e tecnologia (PENTLAND, 2015, p. 190)

Essa perspectiva leva a um esvaziamento social e político do homem, levando, como sugere Foucault, a própria morte da espécie:

> O ser humano é uma invenção cuja recente data a arqueologia de nosso pensamento mostra facilmente. E talvez seu fim esteja próximo. [...] Então pode se apostar que o ser humano desapareceria, como um rosto de areia na beira do mar (FOUCAULT, 2000, p.536)

Será, portanto, interesse da sociedade que sejam todos completamente dirigidos por algoritmos?

O estabelecimento de agente decisório diverso daquele estipulado na Constituição de um país, sem que haja a correspondente autorização popular – através dos instrumentos democráticos estabelecidos – parece usurpar do povo o seu direito de ser julgado pelo juízo natural da causa, considerando este como aquele previamente escolhido para conhecê-la.

Neste sentido, ao submeter um caso à análise judicial/social de forma automatizada sem prévia autorização legal, estar-se-á violando diretamente o direito de ser julgado pela autoridade competente estabelecida na Constituição Federal.

Conforme já estabelecido no capítulo 3, em relação aos *softwares* de inteligência artificial, por serem capazes de decidir, é necessário esperar destes que se comportem eticamente e, ainda de forma mais assertiva, conforme a lei que deriva da vontade popular.

Neste caso, como a lei não autoriza explicitamente que robôs exerçam a jurisdição, e aqui está-se diante de uma interpretação restritiva de forma a proteger a prerrogativa do juiz natural, cria-se um paradoxo em que o próprio algoritmo estaria violando a lei ao decidir.

Assim, como forma de se garantir a prevalência dos princípios democráticos e da garantia do juiz natural, deve a legislação prever expressamente a possibilidade de decisão judicial automatizada, ainda que de forma parcial.

6.4. Regulamentação Transnacional

As discussões sobre regulamentações transnacionais ganharam maiores relevâncias diante de um contexto de globalização e do surgimento de fatos que ultrapassaram os limites geográficos dos países, como do caso em análise, onde as tecnologias desenvolvidas frequentemente possuem um alcance global.

Segundo Calpado (2008), com o fim da Segunda Guerra Mundial e a criação da Organização das Nações Unidas (ONU), houve uma mudança de panorama no tocante às normas internacionais. Diante das atrocidades da guerra, o direito internacional passou a focar mais nos indivíduos do que no próprio Estado. Aduz a autora:

> It gradually lost the features of the classical era, placing greater emphasis on individuals, peoples, human beings as a whole, humanity, and the future generations. State sovereignty has been redefined by developments in the field of the safeguard of human rights, peoples' law, the 'human' environment, the common heritage of mankind, the cultural heritage, sustainable development and international trade[95]. (CALPADO, 2008, p. 1)

Além disso, Arts e Kerwer (2007) destacam que desde 1990 houve um aumento visível de regulação relacionada à matérias extrafronteiras, tendo esse regime de "regulamentação global" alterado a forma com que os países lidam com o direito internacional. Trata-se do que se denomina "verticalização" do poder, com a criação, pela sociedade internacional, de regras e estruturas que se aproximam de verdadeiras cortes internacionais autônomas.

Assim, esse "enfraquecimento" da soberania estatal acaba por fortalecer uma estrutura de poder internacional pautada em "processos multilaterais de decisões" (CAPALDO, 2008, p. 3) que tem por objetivo proteger os direitos fundamentais e também os interesses protegidos pelos regramentos internacionais.

Podem ser citados como exemplos diversas áreas, como os próprios direitos humanos, a democracia, economia, comércio, saúde, meio ambiente e o objeto do

95 Perdeu-se gradualmente as características da era clássica, colocando maior ênfase nos indivíduos, povos, seres humanos como um todo, humanidade e as gerações futuras. A soberania do Estado foi redefinida por desenvolvimento no campo da salvaguarda dos direitos humanos, direito dos povos, ambiente 'humano', patrimônio comum da humanidade, patrimônio cultural, desenvolvimento sustentável e comércio internacional (tradução nossa)

presente estudo, a inteligência artificial, que se estabeleceu como uma prática que pode auxiliar ou afetar vários dos direitos acima elencados.

A Carta das Nações Unidas foi o documento que maximizou o alcance das regulamentações internacionais. Apesar de estar embasada na própria soberania dos Estados, ela contempla mecanismos e idealiza a proteção dos direitos individuais dos povos. Além disso, a comunidade internacional, que carece de uma estrutura organizada, passou a contar com uma estrutura, ainda que atípica, para resolução de suas questões que ultrapassassem as fronteiras, em um sistema de clara governança mundial.

Sobre o tema, aduz Capaldo que:

> "World Governance is legitimized by pronouncements of the International Court of Justice, where it affirms its authority and power of control on states' organs and on those international organs endowed with the world's decision-making power[96]". (CALPADO, 2008, p. 7)

Além disso, Mathias Albert aponta quatro fatores que demonstram o crescimento desse processo de evolução do sistema legal da sociedade mundial:

> 1. The continuing Evolution of international law, including the emergence of more and more elements of supranational law;
> 2. The emergence of 'new' legal arrangements mostly in the realm of a so called 'transnational law';
> 3. The increasing internationalization of national (and subnational) legal systems;
> 4. The increasing 'legalization' of various fields of social relations, but particularly also of international political relations (ALBERT, 2007, p. 192)

Restou-se, portanto, estabelecido um caráter dúplice também no direito internacional: normas regulamentando a relação entre os Estados e a outra os valores eleitos para proteção internacional pela comunidade mundial.

Contudo, inobstante o crescimento da quantidade destas leis e de seus mecanismos de aplicação, é importante mencionar que a sua formação decorre de uma dinâmica diferenciada em relação ao processo legislativo normalmente empreendido dentro de cada Estado e pode assumir a forma de tratados ou de regras voluntárias (*standards*).

---

[96] A Governança Mundial é legitimada pelos pronunciamentos da Corte Internacional de Justiça, onde é afirmada a sua autoridade e poder de controle sobre os órgãos dos Estados e sobre os órgãos internacionais dotados do poder de decisão mundial. (tradução nossa)

Os tratados são "*an international agreement concluded between States in written form and governed by international law, whether embodied in a single instrument or in two or more related instruments and whatever its particular designation*[97] (ONU, online)

As regras voluntárias, por outro lado, se ocupam de estabelecer condições ou padrões de comportamento que são aceitas pelos países, mediante um processo de adesão, como o próprio nome já diz, espontânea.

Mas, por qual motivo os países iriam aderir a regras voluntárias? Arts e Kerwer (2007) apontam que regras voluntárias facilitam a coordenação. Nesse sentido, seria interesse dos anuentes que regras fossem estabelecidas para evitar conflitos relacionados ao próprio exercício da atividade, como por em questões relacionadas ao comércio. Os autores citam, como exemplo, a padronização relacionada aos *plugs* de tomadas e o interesse de todos os envolvidos que haja um padrão previamente estabelecido.

Contudo, no atual panorama político mundial, marcado pela extrema conflitualidade entre as vertentes ideológicas opostas, parece difícil crer em uma construção harmônica de regras que atendam os interesses de ambos. No entanto, a experiência recente pode demonstrar um caminho para o consenso.

Em relação à regulamentação da inteligência artificial, a grande maioria dos diplomas legais existentes possuem abrangência nacional, não havendo uma regulamentação supranacional com mecanismos eficientes de controle e supervisão das falhas identificadas. Tal panorama pode ser compreendido pela enorme diversidade cultural, que impede, por exemplo, o reconhecimento de direitos humanos universais e irrestritos em todo o planeta.

Contudo, a experiência recente com a *Recommendation on the Ethics of Artificial Intelligence*, da Organização das Nações Unidas para Educação, a Ciência e a Cultura (UNESCO), pode demonstrar como seria possível a construção de tais regulamentações voluntárias.

Natorski (2024) destaca a dificuldade em se obter um acordo entre os estados com orientação liberal e aqueles com um viés de resguardo da soberania local. Para

[97] Um acordo internacional celebrado entre Estados por escrito e regido pelo direito internacional, quer esteja consagrado num único instrumento, quer em dois ou mais instrumentos conexos, qualquer que seja a sua designação específica (tradução nossa). Disponível em: https://legal.un.org/ilc/texts/instruments/english/conventions/1_1_1969.pdf

os primeiros, haveria uma priorização da necessidade de proteção dos direitos humanos universais, igualdade de gênero e uma abordagem multilateral. Os segundos, ao contrário, enfatizariam o resguardo do estado soberano e de suas especificidades culturais na implementação da inteligência artificial.

Nesta perspectiva, diante de cenário de improvável consenso, como tal normativo foi aprovado?

Nas palavras de Natorski:

> During the examined negotiations, no state unequivocally shifted its inicial positions or joined a group of countries justifying different approaches. On the contraty, all states followed their general positions during the negotiation until the last hour of deliberation, and despite the persistent diversity of positions, they achieved a compromise. However, instead of changing the preferences, the persuasion in negotiations of amendments focused on convincing that the negotiated text would still fit into the preferences of other partners[98] (NATORSKI, 2024, p.1093)

Observa-se, portanto, que o texto apresentou uma escrita com possibilidades de interpretação flexível, possibilitando uma fusão entre um sistema universal de valores e a possibilidade dos Estados estabelecerem seus próprios regulamentos, preservando suas características culturais e soberania,

Neste sentido, destaca Natorski (2024) que o documento, em harmonia com as prioridades liberais, mencionou 63 vezes a necessidade de respeito aos direitos humanos e 25 vezes o respeito pela lei internacional. Destacou a lista de critérios não-discriminatórios e a definição de valores do respeito, proteção e promoção dos direitos humanos, liberdades individuais, dignidade da pessoa humana, diversidade e igualdade.

Em contrapartida, o texto estipulou que os Estados são os principais sujeitos responsáveis pela implementação do documento. Neste sentido, destacou o autor que dos 80 parágrafos em que se estabeleceram as áreas de atuação, em apenas dois não restou explicitada a responsabilidade do Estado-membro no que toca a sua regulamentação.

---

98 Durante as negociações examinadas, nenhum Estado mudou inequivocamente suas posições iniciais ou se juntou a um grupo de países justificando diferentes abordagens. Pelo contrário, todos os Estados seguiram suas posições gerais durante a negociação até a última hora de deliberação e, apesar da diversidade persistente de posições, eles alcançaram um compromisso. No entanto, em vez de mudar as preferências, a persuasão nas negociações de emendas se concentrou em convencer que o texto negociado ainda se encaixaria nas preferências de outros parceiros (tradução nossa)

Resume o autor que "*the draft text submitted to negotiations merged the principle of universalism from the liberal order of worth with particularism from the sovereigntist positions*[99]" (NATORSKI, 2024, p.1103).

A solução encontrada – e aprovada por todos os Estados que compõem a UNESCO -, então, pode demonstrar que, apesar de difícil, a aprovação de uma regulamentação mundial em inteligência artificial não é impossível.

Ainda que não se trate de um texto que contem mecanismos de execução forçada, é importante ter em mente a influência exercida por tais normas de *standards* no direito interno de cada país. Para Calpado (2007), há uma crescente confusão entre o direito interno e internacional, com limites turvos entre ambos. Para ela:

> … states do pay great attention to the effects of international treaties on the domestic legal order. The need to cordinate treaties with possibly conflicting internal norms has become urgent as a result of the enormous expansion (not only quantitative) of treaties, by their tendency to govern relations between individuals of various nature (e. g., commercial, social, or economic) and by the need to protect the rights of private persons[100] (CALPADO, 2008, p. 174-175).

Para ilustrar a influência das obrigações internacionalmente assumidas pelos Estados, importante a decisão tomada pela Corte Internacional de Justiça no caso LaGrand, onde se estabeleceu que "*The Court observes in this regard that it can determine the existence of a violation of a international obligation. If necessary, it can also hold that a domestic law has been the cause of this violation*[101]" (CORTE INTERNACIONAL DE JUSTIÇA, 2001, *online*). Observa-se, aqui, que houve punição pela não-adesão, de lei interna, aos princípios estabelecidos em regramento internacional, com clara demonstração de supremacia do direito internacional voltado a proteção dos direitos humanos.

Dessa forma, restou demonstrado que o estabelecimento de regras, ainda que principiológicas, mas adotadas de forma consensual entre os países em sede de

---

99 O projeto de texto submetido a negociações fundiu o princípio do universalismo da ordem liberal do valor com o particularismo das posições soberanistas (tradução nossa).

100 … os estados prestam grande atenção aos efeitos dos tratados internacionais na ordem jurídica doméstica. A necessidade de coordenar tratados com normas internas possivelmente conflitantes tornou-se urgente como resultado da enorme expansão (não apenas quantitativa) de tratados, por sua tendência a governar relações entre indivíduos de várias naturezas (por exemplo, comercial, social ou econômica) e pela necessidade de proteger os direitos de pessoas privadas (tradução nossa).

101 O Tribunal observa a este respeito que pode determinar a existência de uma violação de uma obrigação internacional. Se necessário, pode também considerar que uma lei interna foi a causa desta violação. (tradução nossa)

direito internacional, poderá acelerar e impor aos Estados um padrão mínimo de proteção aos direitos dos vulneráveis no desenvolvimento e fiscalização de so*ftwares* de inteligência artificial, o que valida e reforça a tese de que a regulamentação é uma ferramenta extremamente necessária para ocorrência desta proteção.

6.5. Os Eixos Fundamentais para o Desenvolvimento e Uso da Inteligência Artificial

Explicitadas diversas questões sobre o tema analisado no presente estudo, é possível se estabelecer a necessidade de parâmetros acerca do desenvolvimento e uso de inteligência artificial. Estes, como também demonstrado, decorrem diretamente da capacidade e necessidade de regulamentação para que possam ter força impositiva.

Assim, apresenta-se a proposta de uma divisão tríplice de Eixos Fundamentais para o Desenvolvimento de Inteligência Artificial que juntos podem proporcionar um elevado grau de segurança quanto à utilização de tais *softwares* no momento da tomada de decisões.

Os Eixos englobam as diretrizes e orientações necessárias e busca sistematizar, de forma didática, como se estabelecer um controle de algoritmos preciso, possibilitando que, uma vez aplicado, seja obtido o fim último na utilização de tais programas: eficiência aliada à justiça.

Divididos em três categorias, os Eixos Diretivos demonstram as medidas e condutas a serem tomadas em momentos muitas vezes diversos, permitindo uma atuação e controle eficaz, conforme mais bem explicitado a seguir.

6.5.1. Eixo de Conteúdo

O Eixo de Conteúdo traz o núcleo protetivo material que deve existir em todo e qualquer *software* que use inteligência artificial na tomada de decisões envolvendo seres humanos. Há, neste ponto, um aparente consenso entre os mais diversos diplomas que buscam normatizar aspectos da inteligência artificial em estabelecer os direitos humanos como diretriz básica no seu desenvolvimento, a exemplo do *Asilomar Principles, The Barcelona Declaration, The Montreal Declaration e The Ethical Guidelines from the Japanese Society for Artificial Intelligence*.

Yeung, Howes e Pogrebna se manifestam no mesmo sentido:

> We believe that international human rights standards offer the most promising set of ethical standards for AI systems, as several civil society organizations have suggested. As an international governance framework, human rights law is intended to establish global standards (norms) and mechanisms of accountability that specify that the ways in which individuals are entitled to be treated[102] (YEUNG, HOWES; POGREBNA, 2020, p. 80-81)

Assim, deve ser estabelecida o que se convencionou em chamar de *human-centered approach*, ou seja, uma abordagem em que o ser humano seja sempre colocado no centro de todo o processo, desde o desenvolvimento até o resultado. A centralização, portanto, deve ocorrer tanto em relação ao resultado, buscando-se sempre o bem-estar do homem, quanto em relação à própria responsabilidade, já que não se poderia, a princípio, estabelecer o *software* como responsável por eventual ilícito.

Na mesma toada, observa-se que o uso massivo dos dados em diversas esferas do cotidiano despersonaliza o usuário. Assim, "*i corpi diventano dati e la rappresentazione sociale del soggetto è sempre più affidata ad algoritmi che elaborano le informazioni raccolte e ai profili che su questa base vengono costruiti*[103]" (GIACOMELLI, 2019, p. 273). Essa categorização tende a gerar as desigualdades aqui demonstradas e cabe aos desenvolvedores de tais *softwares* a criação de ferramentas que possibilitem a "humanização" do processo.

Com o objetivo de humanização, Green e Chen (2019) apontam o estabelecimento de um framework chamado *algorithm-in-the-loop (AITL)*, em que o foco seria na interação entre o homem e o algoritmo para que seja melhorada a decisão humana, sem que o foco seja no aprimoramento do próprio código. Explicam:

> Instead of improving computation by using humans to handle algorithmic blind sports (such as analyzing unstructured data), AITL systems improve human decisions by using computation to handle cognitive blind sports (such as finding patterns in large, complex datasets. This framework centes human-algorithm interactions as the locus of study and **prioritizes the human's**

[102] Acreditamos que as normas internacionais de direitos humanos oferecem o conjunto mais promissor de normas éticas para os sistemas de IA, como sugeriram várias organizações da sociedade civil. Como quadro de governa internacional, os direitos humanos pretendem estabelecer padrões globais (normas) e mecanismos de responsabilização que especifiquem de que formas os indivíduos têm direito a ser tratados (tradução nossa)

[103] os corpos tornam-se dados e a representação social do sujeito é cada vez mais confiada aos algoritmos que processam a informação recolhida e aos perfis que se constroem a partir desta base (tradução nossa)

> **decision over the algorithm's as the most important outcome**.[104] (GREEN; CHEN, 2019, p. 97-98, grifo nosso)

Assim, vê-se que a proposta busca priorizar a decisão humana e não o algoritmo. Há um retorno ao homem como centro decisório sem que haja uma excessiva necessidade de se melhorar o sistema com o foco em sua autonomia. Como exemplos de implementação da mecânica proposta temos situações em que o *software* apresenta ao usuário todos os dados decorrentes de sua análise de forma estruturada, possibilitando que o indivíduo analise e realize a tomada de decisão de forma mais bem informada.

Todas essas medidas são voltadas para a humanização do processo. Seja para resguardar os direitos daqueles submetidos à decisão automatizada, seja para trazer o homem de volta ao centro do processo decisório. Mas, de tudo que foi demonstrado, entende-se ainda não ser suficiente tal estabelecimento genérico de proteção aos direitos humanos. É necessário que os valores também sejam expressamente definidos.

Como já demonstrado no Capítulo 3 (Ética e Inteligência Artificial), o estabelecimento de princípios éticos também se tornou uma frequente na grande maioria dos documentos que tratam sobre inteligência artificial. E, como demonstrado naquele capítulo, mais especificamente no ponto 3.5, entende-se que os princípios ali elencados são obrigatórios para que haja uma mínima proteção aos direitos dos indivíduos, inclusive com a sua necessária positivação na legislação pertinente.

Estipula-se necessário, desta forma, que todo e qualquer algoritmo que vise à tomada de decisões, deve se submeter aos princípios da beneficência, não-maleficência, autonomia, justiça e explicabilidade.

6.5.2. Eixo de Controle

O Eixo de Controle diz respeito às medidas e direcionamentos que devem ser tomados no intuito de se *corrigir* as falhas advindas da coleta de dados pela máquina.

---

[104] Em vez de melhorar a computação usando humanos para lidar com os pontos cegos dos algoritmos (como a análise de dados não estruturados), os sistemas AITL melhoram as decisões humanas usando a computação para lidar com pontos cegos cognitivos (como encontrar padrões em conjuntos de dados grandes e complexos). Esta estrutura centra-se nas interações humano-algoritmo como o *locus* de estudo e prioriza a decisão humana sobre a do algoritmo como o resultado mais importante. (tradução nossa)

Como já demonstrado com mais detalhes no Capítulo 05, durante a utilização de algoritmos de modelos decisórios, pode ser detectado o fenômeno de enviesamento (ou *bias*) das decisões, o que gera um cenário de violação do direito ao tratamento igualitário do cidadão submetido à decisão.

Considerando que tais vieses decorrem de diversos fatores, na obtenção dos dados ou até durante o próprio uso, por exemplo, é necessário que algumas medidas sejam empregadas para que se evite a sistematização de um tratamento discriminatório e é disso que se ocupam as medidas incluídas neste Eixo.

O primeiro ponto a ser observado dentro do 2º Eixo é o da análise cautelar dos dados, com base no *Princípio da Precaução.* A evolução tecnológica é muito mais rápida do que a capacidade humana de regulamentar suas relações e corrigir os seus erros. Para evitar – ou pelo menos mitigar - a ocorrência de violações, Andrea Simoncini (2020) sugere a adoção de um mecanismo desenvolvido pelo direito ambiental: o princípio da precaução. Para o autor, a atual situação relacionada aos avanços tecnológicos se assemelha a do direito ambiental, quando foi necessária a elaboração de estratégias prévias para impedir violação de direitos individuais e coletivos diante de uma incerteza sobre as consequências de determinada prática.

Aduz o autor:

> Protremmo sintetizzare così il princípio: la condizione di incerteza a riguardo dei possibili effetti negativi dell'impiego di uma tecnologia (inclusa l'inteligenza artificiale) non può essere utilizzata come uma ragione legittima per non regolare e limitare tale sviluppo[105] (SIMONCINI, 2020. P. 62)

Prossegue o autor afirmando o discurso, por parte das *big techs*, de que a imposição de qualquer limite ou regulamentação aos setores de difusão da tecnologia poderia ferir a liberdade de pesquisa e, por via de consequência, ao próprio desenvolvimento, o que não passaria de um mito. Tal construção parece ignorar a necessidade de resguardo dos direitos e garantias individuais – aí amparado o direito a não-discriminação – que deve estar em patamar superior ao desenvolvimento *de per si*.

Dessa forma, a prevalência de uma cultura de precaução quanto ao desenvolvimento de novas tecnologias nos leva a pensar sempre em priorizar aquilo

[105] Poderíamos resumir o princípio da seguinte forma: a condição de incerteza quanto aos possíveis efeitos negativos do uso de uma tecnologia (incluindo a inteligência artificial) não pode ser usada como uma razão legítima para não regulamentar e limitar tal desenvolvimento. (tradução nossa)

que é mais caro para a sociedade. O avanço tecnológico não pode ser um fim em si mesmo às custas de lesões a direitos do povo. Assim, tal fator, observando sempre a prevalência dos direitos individuais diante de uma dúvida razoável acerca da implementação de determinada tecnologia, produz uma garantia de resguardo e proteção social em face de ameaças desconhecidas – mas potencialmente lesivas – no uso de determinados programas.

Se alicerçam também neste Eixo as discussões acerca da *Diversidade*. Gebru (2020) destaca que, ao analisar o pensamento científico do século XIX e grandes avanços tecnológicos, é possível perceber que a ausência de representatividade entre os responsáveis pela tecnologia tem o potencial de gerar desequilíbrio de poder no mundo além de trazer consequências negativas para aqueles não representados no seu desenvolvimento.

Assim, o que se percebe é que, a ausência de diversidade no desenvolvimento do algoritmo, seja sob o ponto de vista do programador, seja sob o ponto de vista dos próprios dados coletados, tem grave repercussão na própria acuidade do programa, infringindo a supressão ou diminuição de direitos em determinados grupos que não foram contemplados.

Sobre esse ponto torna-se adequada a análise do caso *Ewert* v. Canadá, julgado pela Suprema Corte canadense. Jeffrey Ewert foi condenado por crimes de homicídio e agressão sexual, sendo-lhe determinada a custódia em presídios de segurança média e máxima. Ao solicitar avaliação psicológica para redução do nível de segurança e para ser autorizado a demandar liberdade condicional, teve seu pedido negado pelo *Correctional Service of Canada* (CSC). O condenado, então, desafiou a decisão do CSC, sob fundamento de que a decisão tomada com auxílio de algoritmo não levou em conta sua origem indígena e que, por isso, não seria válida.

Sobre este ponto, decidiu a Suprema Corte do Canadá:

> *In continuing to rely on the impugned tools without ensuring that they are valid when applied to Indigenous offenders, the CSC breached its obligation under s. 24 (1) of the CCRA to take all reasonable steps to ensure that any information about an offender that it uses is as accurate as possible*[106] *(SUPREME COURT OF CANADA, 1998, on-line)*

[106] Ao continuar a confiar nas ferramentas impugnadas sem garantir que sejam válidas quando aplicadas a infratores indígenas, o CSC violou a sua obrigação nos termos do s. 24 (1) do CCRA de tomar todas as medidas razoáveis para garantir que qualquer informação sobre um infrator que ela utiliza seja tão precisa quanto possível (tradução nossa)

Para Giacomelli (2019), os juízes observaram que a validade de uma decisão automatizada, notadamente quando de seu uso na esfera penal, deve levar em conta as diferenças culturais daqueles a ela submetidos, sob pena de violação de direitos e tratamento discriminatório. É direito do réu, portanto, de ter analisadas todas as circunstâncias de seu caso, garantindo-se o seu direito à ampla defesa. Trata-se de claro uso do princípio da individualização da pena (e das próprias decisões judiciais) diante de diferentes características do indivíduo.

Ainda neste sentido, Virgínia Dignum reforça a questão relacionando o desenvolvimento de um *software* de forma participativa e diversa à obtenção de um produto mais adequado e responsável:

> Responsible IA is about participation. It is necessary to understand how different people work with and live with AI technologies across cultures in order to develop frameworks for responsible IA. In fact, AI does not stand in itself, but must be understood as a part of sociotechnical relations with all its diversity[107] (DIGNUM, 2020, p. 220)

Assim, para a autora, os times de desenvolvimento devem ser compostos por uma equipe multidisciplinar e diversa, incluindo cientistas sociais, filósofos, e estudiosos de temas relacionados a gênero, etnias, diferenças culturais etc., sendo este, portanto, um ponto crucial para obtenção de dados "limpos" e afastados de eventual viés advindo do programador.

Outra medida que pode ser empenhada dentro do Eixo de Controle diz respeito ao estabelecimento do *equilíbrio estatístico*. Trata-se, pois, de um campo de estudo em que se busca como encontrar justiça e equidade nos algoritmos. DeMichele *et al.* (2020) sugerem três medidas que podem atuar neste espectro: equilíbrio na taxa de erro, paridade preditiva e calibragem.

Para os autores, o equilíbrio na taxa de erro é obtido quando as taxas de falso-positivo e falso-negativo são iguais entre todos os grupos em determinada questão (por exemplo: quando as negros e brancos são incorretamente etiquetados como de alto risco em uma mesma proporção). Quanto à paridade preditiva, esta é alcançada quando há uma paridade em relação ao índice de reincidência entre todos os grupos. E, por fim, a calibragem é obtida quando um instrumento determina as mesmas

[107] AI responsável é sobre participação. É necessário compreender como diferentes pessoas trabalham e convivem com as tecnologias de IA em todas as culturas, a fim de desenvolver estruturas para uma IA responsável. Na verdade, a IA não é independente, mas deve ser entendida como parte das relações sociotécnicas com toda a sua diversidade (tradução nossa)

probabilidades de reincidência independente do grupo a que pertença o indivíduo analisado.

Nesse sentido, a medida parece se aproximar das discussões tratadas de forma mais específica nos Estados Unidos e na União Europeia acerca de ações afirmativas na programação dos *softwares*.

Ações afirmativas são medidas de "discriminação positiva", em que há a adoção de medidas específicas que visam à promoção da igualdade material. Podem ser definidas como:

> um conjunto de políticas públicas para proteger minorias e grupos que, em uma determinada sociedade, tenham sido discriminados no passado. A ação afirmativa visa remover barreiras, formais e informais, que impeçam o acesso de certos grupos ao mercado de trabalho, universidades e posições de liderança. Em termos práticos, as ações afirmativas incentivam as organizações a agir positivamente a fim de favorecer pessoas de segmentos sociais discriminados a terem oportunidade de ascender a postos de comando. (CAMPOS OLIVEN, 2007, p. 30)

É importante mencionar que tais medidas estão previstas expressamente no Art. 1, parágrafo 4º da Convenção sobre a Eliminação de todas as formas de Discriminação Racial da ONU.

As ações desta natureza, quando relacionadas à inteligência artificial, deram origem à expressão "ação afirmativa algorítmica" quando ela se insere dentro da própria programação do sistema no intuito de promover um tratamento igualitário. A medida, embora se apresente amparada pelo intuito de se garantir isonomia, não tem sido completamente aceita, como será mostrado a seguir.

Sobre a necessidade de inclusão explícita de valores e como poderiam atuar afirmativamente:

> Mas precisamos impor valores humanos nos sistemas, mesmo a custo da eficiência. Por exemplo, um modelo pode ser programado para garantir que vários grupos éticos ou níveis de renda estejam representados dentro de grupos de eleitores ou consumidores. Ou poderia destacar casos nos quais pessoas de certas localidades pagam o dobro da média por certos serviços. (O'NEAL, 2020, p. 192)

No mesmo sentido se posiciona Giacomelli:

> Una proposta di riabilitazione della tutela antidiscriminatoria di fronte alle sfide dell'intelligenza artificiale potrebbe allora essere quella di sviluppare un

> ‘algorithmic affirmative action’, ovverosia un piano di pratiche e azioni proattive[108] (GIACOMELLI, 2019, p. 296)

Para que se possa analisar o cabimento de uma ação afirmativa, foram sugeridas algumas abordagens em que seria possível estabelecer uma relação de desigualdade e repará-la.

A primeira delas é a de justiça cega (*blindness*, como abordado no tópico 5.5.1). Nesta abordagem, se busca a retirada dos dados de todas as variáveis sensíveis protegidas, como raça, sexo, cor, religião etc. Tal abordagem se mostra eficaz em diversas situações, como no exemplo citado por Bent (2020) em que testes de orquestras sinfônicas são feitos com as cortinas abaixadas (e em muitos casos com os candidatos descalços – evitando assim que se escute o barulho de saltos altos no palco). Atesta o autor que tal medida aumentou substancialmente as chances de mulheres serem selecionadas para as orquestras.

O problema com essa abordagem é exatamente a utilização dos proxies, tema também já trabalhado. Em relação às máquinas, outros dados são utilizados e acabam tornando sem eficácia a mera supressão das informações sensíveis. A medida, portanto, é normalmente ineficaz.

A segunda propositura diz respeito à justiça de grupo *(group fairness*). Nessa categoria, a medida de justiça e equidade seria analisada através da comparação de resultados obtidos após análise do *software* entre dois grupos diferentes com base em características sensíveis (grupos de brancos e de negros, por exemplo).

Para Bent (2020), em uma análise mais restritiva, esta proposta exigiria que os resultados da predição fossem iguais para todos os grupos. Nesse sentido, seria exigido um forçoso ajuste pelo *software*, para que se obtivessem exatamente as mesmas taxas entre os grupos de características sensíveis, o que poderia impactar em uma grande falta de precisão.

Em um posicionamento mais moderado, seria possível analisar os resultados entre grupos, mas ponderando determinadas características e limites, aceitando-se certa variação. Bent (2020) propõe o uso da *four-fifths rule*, em que seria exigido do algoritmo que as taxas analisadas não tivessem uma diferença de quatro quintos entre

---

[108] Uma proposta para a reabilitação da tutela antidiscriminação face aos desafios da inteligência artificial poderia então ser o desenvolvimento de uma 'ação afirmativa algorítmica', ou seja, um plano de práticas e ações proativas. (tradução nossa)

si. Esta seria apenas uma sugestão quanto a um suposto balanceamento nos grupos, sem exigir a rigidez de necessidade de um resultado exatamente igual.

Fazendo um contraponto à justiça de grupo, teóricos defendem a justiça individual (*individual fairness*), argumentando que ao invés de se focar nas comparações entre grupos, deveriam ser empreendidas análises relacionadas aos indivíduos, garantindo que aqueles que apresentem características idênticas sejam classificados da mesma forma.

Em todos esses casos de balanceamento, ocorre, como mencionado, uma perda na precisão do resultado. Para Bent (2020), trata-se do chamado preço da justiça (*price of fairness*). Há que se tomar em consideração que tais ajustes também possuem as mesmas características em alguns outros tipos de ações afirmativas, notadamente ao se conceder, por exemplo, vagas em universidades aqueles que obtiveram uma colocação inferior. É inegável que há uma alteração no resultado esperado, mas este ocorre dentro do previsto e no intuito de se obter um tratamento discriminatório sancionado pelo Estado.

É importante mencionar que a aplicação de ajustes numéricos como forma de ações afirmativas pode encontrar resistência nas Cortes de Justiça. Já constam decisões na justiça americana que parecem afastar a possibilidade deste tipo de conduta, com base na doutrina *disparate treatment* e proteção de tratamento igualitário.

A Universidade de Michigan estabeleceu um programa em que concedia a indivíduos de minorias sub-representadas um bônus de 20 (vinte) pontos durante o processo de seleção. Para a Suprema Corte Americana (Gratz v. Bollinger), a política da instituição violou a doutrina da proteção igualitária uma vez que a utilização do fator raça e a concessão de um critério puramente matemático imporia a todos uma diferença de tratamento não amparada na Constituição.

Assim, quanto a tais medidas, ainda parece ser necessário discussões quanto à sua aplicação e validade perante os sistemas jurídicos.

Ressalte-se que as medidas propostas figuram em um rol exemplificativo e carentes de ampla discussão. As propostas a serem efetivadas no Eixo de Controle devem ser adequadas ao tipo de *software* desenvolvido, devendo ter sempre em mente uma análise voltada para o equilíbrio das relações e garantia de isonomia de tratamento entre todos os que serão submetidos às decisões automatizadas, sem violar qualquer norma de proteção contra discriminação. Trata-se, portanto, de um

campo que se encontra em plena discussão e estudo, no que se pode aguardar evolução teórica em um curto período de tempo.

6.5.3. Eixo de Supervisão

O Eixo de Supervisão figura na proposta como uma garantia do cidadão de que os *softwares* atenderam os requisitos do Eixo de Conteúdo, passaram por eventual medida corretiva do Eixo de Controle e encontram-se aptos a serem utilizados.

Serve, ainda, para cessar eventual agressão a direitos que surgirem após início da operação do algoritmo, impedindo a propagação de danos e suspendendo as suas atividades.

Neste Eixo, portanto, encontram-se medidas de fiscalização e acompanhamento que devem existir desde o desenvolvimento dos programas até a sua efetiva utilização, garantindo o bom uso da tecnologia.

Diversas são as medidas que podem ser implementadas sob o espectro de proteção do Eixo de Supervisão.

A base de funcionamento de todo o sistema de supervisão é a *transparência*. A transparência é quem possibilita que haja uma fiscalização efetiva por parte dos órgãos dotados desta responsabilidade, além de possibilitar também a própria responsabilização daquele que violou o sistema de proteção de direitos humanos.

É facilmente dedutível que o simples fornecimento de informações não é suficiente para impedir o cometimento de ilícitos, uma vez que é necessário o estabelecimento de um sistema de sanções, mas é através desta ferramenta que se pode viabilizar o efetivo controle das empresas envolvidas no processo de desenvolvimento de tecnologia dotada de inteligência artificial. É necessário, ainda, que seja documentada a cadeia de eventos que deram causa a determinado resultado, possibilitando a rastreabilidade das decisões e processos.

Diakopoulos (2020) destaca que não é possível verificar a questão como uma simples dicotomia entre o transparente e não-transparente. Para o autor, existem diversas questões que precisam ser analisadas quando se trata de transparência. Sobre tal temática, estabelece o autor pontos a serem notados: “*relevant factors include the type, scope, and reliability of information made available; the recipients of*

transparency information and how they plan to use it; and the relationship between the disclosing entity and the recipient[109]" (DIAKOPOULOS, 2020, p.199).

Assim, as informações que são fornecidas podem ser diferentes de acordo com o sujeito a quem são destinadas. Além disso, deve ser observada desde o momento do desenvolvimento do *software* uma preocupação com a disponibilização de informação, ou *usable transparency*. Nesse sentido, "*transparency information can be formated in a number of different modalities such as in structured databases or documents, in written texts (perhaps even using natural language generation), or via visual and interactive interfaces*[110]" (DIAKOPOULOS, 2020, p. 204).

Trata-se de medida que visa a facilitar a supervisão, com o fornecimento de informação através de um sistema que, de fato, permita ser analisado o processamento e funcionamento do *software*, ou como se convencionou chamar: *explainable AI.*

Para melhor compreensão, *explainable AI* pode ser definida como: *Explainability is the capacity to express why an AI system reached a particular decision, recommendation, or prediction. Developing this capability requires understanding how the AI model operates and the types of data used to train it*[111]. (GRENNAN *et al.*, 2022, on-line)

Outro processo sugerido por Diakopoulos (2020) como forma de facilitar a entrega de dados é o da pirâmide de informação, em que o conhecimento é aprofundado na medida em que o agente envolvido se aprofunda no processo fiscalizatório.

Há que ser entendido, de fato, que existem problemas relacionados com o exercício pleno da transparência, como a manipulação de resultados, violação de privacidade, (des) vantagens competitivas e problemas relacionados com direito de proteção à propriedade intelectual. Contudo, é exatamente no intuito de se evitar abusos diante da transparência exigida pela fiscalização que entra em cena talvez a

---

[109] os fatores relevantes incluem o tipo, o escopo e a confiabilidade das informações disponibilizadas; os destinatários da informação sobre transparência e como planejam utilizá-la; e a relação entre a entidade divulgadora e o destinatário (tradução nossa)

[110] as informações de transparência podem ser formatadas em diversas modalidades, como em bancos de dados ou documentos estruturados, em textos escritos (talvez até usando geração de linguagem natural) ou por meio de interfaces visuais e interativas. (tradução nossa)

[111] Explicabilidade é a capacidade de expressar porque um sistema de IA tomou uma determinada decisão, recomendação ou previsão. O desenvolvimento desta capacidade requer a compreensão de como o modelo de IA funciona e os tipos de dados usados para treiná-lo. (tradução nossa)

mais importante das medidas relacionadas ao desenvolvimento de *softwares* de inteligência artificial: a regulamentação.

A regulamentação, como já explorado no início do capítulo, se afigura como ferramenta eficaz – e necessária – para gerir todo o processo de desenvolvimento e controle de ferramentas inteligentes. Somente através da regulamentação, é possível conferir um conjunto de normas seguras para o cidadão e ao mesmo tempo balizas claras de desenvolvimento de *softwares* para as empresas, evitando, assim, perdas decorrentes da criação de produtos que posteriormente venham a ser contestados.

A regulamentação, então, deve estabelecer todos os parâmetros necessários para o desenvolvimento, uso e supervisão de *softwares* inteligentes, estipulando deveres legais, como a obediência aos princípios e metodologias indicadas no Eixo de Conteúdo, eventuais ajustes para garantia de isonomia, conforme discorrido no Eixo de Ajuste, e, por fim, mecanismos de fiscalização previstos neste tópico do Eixo de Supervisão.

Além disso, é importante que seja estabelecida uma *supervisão por autoridade competente e independente*. Somente através deste tipo de gestão é que seria possível garantir um efetivo controle acerca a obediência dos critérios estabelecidos na regulamentação e também a efetivação das sanções ali estipuladas.

É necessário que sejam fixados parâmetros claros de autoridade, com a indicação de uma autoridade responsável pela fiscalização, ferramentas de controle e ainda uma gama de sanções possivelmente aplicáveis. A princípio, o caminho trilhado pela legislação europeia, de avaliação de risco, aparenta encontrar melhor amparo entre os estudiosos, servindo, então, como uma boa baliza no momento de desenvolvimento de produtos.

Há ainda outros dois mecanismos que também garantem um maior controle quanto à introdução de um novo produto no mercado a ser implementado pela autoridade de controle: a certificação e a validação.

A *certificação* surgiria como um procedimento prévio de verificação por parte de entes independentes, que garantissem, ao menos naquela fase inicial, a obediência do desenvolvedor aos princípios e normas estabelecidos na legislação. Somente após ultrapassados os testes e garantida a sua regularidade é que poderia o produto ser, enfim, comercializado.

A *validação*, por outro lado, seria um procedimento posterior. Consistiria em um processo utilizado para analisar o comportamento do *software* e se ele está

acompanhando as intenções do desenvolvedor e as permissões do Estado. Logo, serviria o método como ferramenta posterior de acompanhamento dos programas, que poderia ocorrer de forma recorrente ou mediante indícios de violação de direitos.

Os dois procedimentos de checagem e controle seriam previamente descritos na legislação, com procedimentos próprios, possibilitando a preparação por parte dos desenvolvedores para preenchimento de todos os requisitos ali estabelecidos.

Acredita-se, portanto, que a efetiva observação e respeito aos Eixos Fundamentais possa garantir a utilização da inteligência artificial livre de graves danos ou repercussões aos direitos humanos, possibilitando, naqueles casos em que houver qualquer indício de violação, a efetiva ação do poder fiscalizatório.

CONSIDERAÇÕES FINAIS

É inegável o avanço da inteligência artificial nas mais diversas áreas. De simples sugestões de vídeos relacionados aos gostos dos usuários até dispositivos militares, *softwares* dotados de capacidades “humanas” estão sendo desenvolvidos a cada dia e é necessário que haja um acompanhamento próximo por parte de todos.

Como diversas outras tecnologias inovadoras, a inteligência artificial é dotada de riscos e recompensas, cabendo a todos, em um esforço cooperativo, atuar para um uso seguro para a sociedade.

Contudo, alguns riscos merecem maior atenção.

O cenário de busca por igualdade e celeridade dentro do Poder Judiciário encontrou na inteligência artificial terreno fértil para sua incessante procura por melhorias. A automação de processos mecânicos e a adoção de ferramentas auxiliares de decisão são claros exemplos de utilização de tais programas. Contudo, é exatamente a partir de seu uso que se pode extrair a problemática da pesquisa: o estudo de seus vieses discriminatórios e como podem ser evitados.

Este problema, conforme demonstrado no capítulo introdutório, e mais bem explorado no decorrer do texto, pode trazer diversos problemas graves, envolvendo inclusive a supressão ilegal da liberdade dos indivíduos submetidos à jurisdição penal.

Ao observar a questão problema (como impedir que a automação da atividade judicante por *software* de inteligência artificial viole direitos humanos de grupos vulneráveis, impondo-lhes regime injustificadamente mais gravoso na seara penal?), entende-se que a hipótese de pesquisa foi confirmada.

Restou demonstrado que a criação de algoritmos pautados na defesa dos Direitos Humanos conduz diretamente ao fim de atos discriminatórios em decorrência da utilização de *softwares* de inteligência artificial no Poder Judiciário. E que, somente através de uma regulamentação efetiva é que se pode evitar a ocorrência de novos fenômenos desta natureza.

Tal conclusão decorre da leitura analítica de todo o estudo produzido, como evidenciado a seguir.

O primeiro capítulo do desenvolvimento do texto tratou sobre a inteligência artificial. Foi apresentado o desenvolvimento histórico da ferramenta e seus principais conceitos. Em seguida, apresentou-se a relação da tecnologia com o direito no intuito de se demonstrar a íntima relação entre as áreas e como podem surgir questões

importantes que merecem ser estudadas e observadas. Neste cenário, apresentou-se ainda diversos casos de decisões automatizadas e a possibilidade de se vislumbrar o potencial benéfico (e/ou lesivo) da tecnologia, com a apresentação das ressalvas comumente apontadas.

No capítulo seguinte foram abordados temas relacionados às questões éticas de *softwares* criados para tomada de decisões. Foi necessária a abordagem teórica para que, estabelecidos conceitos básicos sobre a ética, fosse discutida a necessidade de tais digressões e sob qual enfoque a matéria deveria ser observada em se tratando de decisão automatizada.

O fato é que, como agentes responsáveis pela tomada de decisões, os humanos esperam que as máquinas ajam eticamente. Por esta razão, a discussão se torna extremamente importante. O estabelecimento de padrões éticos se torna necessário para avaliação de adequação do *software* e se este preenche os requisitos necessários para efetivo uso.

Além disso, o capítulo demonstrou um apanhado de princípios comumente apontados como essenciais para o desenvolvimento de *softwares* de inteligência artificial. A centralização no homem e o respeito aos direitos humanos surgem como pilares fundamentais para proteção contra comportamentos lesivos dos robôs. Além disso, diversos outros princípios ali explicitados demonstraram a necessidade de dedicação à ética das decisões automatizadas.

Demonstrou-se, então, que o processo decisório automatizado deve respeitar necessariamente os valores éticos da sociedade no qual é inserido e que estes devem permear todo o desenvolvimento, evitando a inserção de dados ou processos enviesados.

No capítulo seguinte, intitulado Regulamentação da Inteligência Artificial: um estudo comparado, dedicou-se o estudo a uma análise comparativa acerca da regulamentação até então aprovada nos mais importantes *players* do mercado: a China, Estados Unidos, Canadá e Europa. Neste ponto, realizou-se também um contraponto com a situação do Brasil, tendo em vista a necessidade de contribuição com o debate nacional sobre o tema.

O estudo do direito comparado se fez necessário por dois principais motivos: o alcance mundial do tema exige uma análise global acerca da postura adotada pelos principais responsáveis pelo desenvolvimento dos *softwares,* obtendo-se assim um importante panorama sobre como se posicionam cada um dos entes; e a necessidade

de aperfeiçoamento legislativo, considerando que ainda há diversos Estados que ainda não se posicionaram pela regulamentação, ou ainda encontram-se na fase de discussões legislativas. Assim, um apanhado acerca das legislações mundiais é um instrumento eficaz para que se possa investigar as melhores experiências normativas para eventual reprodução.

Posteriormente, sob o título de Grupos Vulneráveis, Justiça Preditiva e Sistema Penal, repousa o capítulo dedicado ao estudo do fenômeno da vulnerabilização social e o surgimento de novas formas de agressão às classes marginalizadas pela utilização de *softwares* inteligentes.

Um apanhado acerca da teoria dos direitos humanos foi necessário para que se pudesse validar a cidadania como um direito extensível a todos e do qual decorre a necessidade de se discriminar positivamente aqueles que precisam de proteção estatal. Neste sentido, o estudo das minorias foi necessário para que se evidenciasse a forma com que determinados grupos sofrem desigualmente violações por parte de instrumentos de Estado, como por exemplo as práticas de *racial profilling*.

Aprofundando a temática em estudo, foi explicitado o funcionamento da inteligência artificial dentro do Poder Judiciário, notadamente em *softwares* responsáveis pela análise de risco, e como tais programas podem vulnerabilizar ainda mais determinados grupos já historicamente marginalizados. Tal análise contou ainda com a demonstração de três casos/práticas específicos sobre a temática e comprovação de como pode ser danosa a utilização sem critérios da tomada de decisão automatizada.

O último capítulo do desenvolvimento, por fim, apresentou aspectos relacionados à Teoria da Regulamentação e de como uma efetiva legislação é necessária para que o problema apresentado no estudo possa ser resolvido. As práticas de autorregulamentação não se mostraram eficazes no combate aos danos causados pelas decisões automatizadas, sendo necessária, portanto, a intervenção do Estado na matéria. Destacou-se, ainda, a regulamentação como uma forma de legitimar o uso de tais *softwares* através de um processo democrático, exigindo dos desenvolvedores o respeito aos critérios eleitos pelo povo para que possam, eventualmente, ser submetidos ao processamento por qualquer tipo de inteligência artificial.

Neste capítulo também foi apresentada a proposta do autor acerca do tema, confirmando a hipótese apresentada e sugerindo a adoção de determinados critérios

e práticas em um sistema segmentado, os Eixos Fundamentais para o Desenvolvimento e Uso da Inteligência Artificial.

A proposta divide um conjunto de prática em três eixos fundamentais: o Eixo de Conteúdo, o Eixo de Controle e o Eixo de Supervisão. O Eixo de Conteúdo seria responsável pelas práticas de controle de processamento da matéria a ser submetida à máquina, com uma abordagem centrada no homem e o estabelecimento de princípios éticos para o seu funcionamento. O Eixo de Controle seria responsável por eventuais ajustes no funcionamento do algoritmo com o objetivo de garantir um tratamento igualitário para todos, evitando que vieses sociais (ou do programador) possam ser perpetuados pela máquina. Por fim, o Eixo de Supervisão possibilita o controle por parte do Poder Público quanto ao funcionamento do *software*, garantindo uma fiscalização eficaz e a possibilidade de punição quando verificado o descumprimento das regras estabelecidas.

Assim, pelo que foi apresentado, é de se observar que houve a confirmação da hipótese de pesquisa, uma vez que se demonstrou a possibilidade de se desenvolver *softwares* eticamente direcionados para proteção dos direitos humanos, mas que tal direcionamento deve ocorrer de forma compulsória por parte dos Estados, não sendo eficaz o estabelecimento de parâmetros gerais ou critérios de autorregulamentação. A necessidade de um regramento forte e específico faz com que possa o desenvolvedor trabalhar em cima de critérios claros e aptos a promover a proteção social e sua necessidade se mostra também comprovada pela própria adesão em massa dos grandes Estados (e União Europeia) a um sistema de regulamentação exaustiva.

Não se pode esperar que a inteligência artificial consiga sozinha mudar séculos de um padrão de desigualdade em desfavor de grupos minoritários, mas é possível que ela seja transformada em mais uma ferramenta nesta luta que deve ser combatida por todos.

O que se precisa ponderar é se a resolução destes problemas históricos será totalmente debitada em ferramentas de inteligência artificial que realizam o processamento de demandas de acordo com os dados que lhe são previamente fornecidos pelo próprio homem. Obviamente, existe razão para se buscar uma solução para a questão da desigualdade social, que desagua no próprio aumento da criminalidade junto a grupos vulneráveis, mas tais medidas não devem ser

depositadas exclusivamente em *softwares* que, como estabelecido, somente reproduzem o que está ocorrendo na sociedade.

A transformação social precisa ser profunda e amparada por políticas de desenvolvimento social e redução da desigualdade, não cabendo exclusivamente aos *softwares* inteligentes desempenhar o papel que deve ser do Estado.

Assim, acredita-se que a regulamentação efetiva do assunto poderá estabelecer condições para o pleno funcionamento das máquinas, possibilitando que, ao revés, sejam inseridas políticas mais claras e transparentes de tratamento isonômico.

Espera-se, portanto, que o presente estudo, com a sistematização bibliográfica realizada, e em especial com a sugestão de divisão das medidas práticas do sistema de Eixos Fundamentais para o Desenvolvimento e Uso da Inteligência Artificial possa engrandecer o debate acadêmico necessário e atual sobre tal tecnologia, fornecendo um suporte teórico para proteção das minorias marginalizadas, evitando a perpetuação de violações graves aos seus direitos fundamentais.

REFERÊNCIAS

ABBAGNANO, N. **Dicionário de filosofia**. São Paulo: Martins Fontes,1998.

ALBERT, M.. Beyond legalization. Reading the increase, variation and differentiation of legal and law-like arrangements in international relations through world society *in* **Law and Legalization in Transnational Relations**, ed. BRÜTSCH, C.; LEHMKUHL, D..Oxon: Routledge, 2007, p.185-201.

ALGORITHMS in the Criminal Justice System: Pre-Trial Risk Assessment Tools. Washington: EPIC, [20--]. Disponível em: https://epic.org/algorithmic- transparency/crim-justice/. Acesso em: 18 set. 2022.

ANGWIN, J. *et al*. Machine Bias. **Propublica**, [*s.l.*], 23 maio 2016. Disponívelem: https://www.propublica.org/article/machine-bias-risk-assessments-in- criminal-sentencing. Acesso em: 19 jan. 2022.

ARENDT, H. **As origens do totalitarismo**. Trad. Roberto Raposo. SãoPaulo: Companhia das Letras, 1989.

ARENDT, H. **Verdade e política**. São Paulo: Perspectiva, 2011

ARISTÓTELES. **Arte Retórica e Arte Poética.** Difusão Européia do Livro,1959.

ARISTÓTELES. **Ética a Nicômaco**. 4 ed. São Paulo: Nova Cultural, 1991.ASIMOV, I. **I, Robot**. New York: Bantam Spectra, 2008.

Arowosegbe, J. O. "**Data bias, intelligent systems and criminal justice outcomes**." *Int. J. Law Inf. Technol.* 31 (2023): 22-45.

ARTS, B; KERWER, D. Beyond legalization? How global standards work *in* **Law and Legalization in Transnational Relations**, ed. BRÜTSCH, C.; LEHMKUHL, D..Oxon: Routledge, 2007, p. 144-165.

ATIBA GOFF, P.; BARSAMIAN KAHN, K. Racial Bias in Policing: Why We Know Less Than We Should. **Social Issues and Policy Review**, v. 6,n. 1, p. 177-210, mar. 2012. Disponível em: https://doi.org/10.1111/j.1751-2409.2011.01039.x. Acesso em: 1 dez. 2023.

ARORA, A.; BARRETT, M; LEE, E., OSBORN, E., and PRINCE, K. **Risk and the future of AI: Algorithmic bias, data colonialism, and marginalization**. Inf. Organ. 33, 3, Sep 2023. Disponível em: https://doi.org/10.1016/j.infoandorg.2023.100478. Acessado em: 08 set 2024.

AWAD, E. et al. The Moral Machine experiment. **Nature**, v. 563, n. 7729,p. 59-64, 24 out. 2018. Disponível em: https://doi.org/10.1038/s41586-018-0637-6. Acesso em: 1 dez. 2023.

BALDWIN, R., CAVE, M.; LOGDE, M. *Understanding Regulation: Theory, Strategy, and Practice*, 2ª ed. Disponível em: https://doi.org/10.1093/acprof:osobl/9780199576081.001.0001. Acessado em: 23 fev. 2024.

BARAK-EREZ, D. Legislation as Transplantation In: LUPO, N.; SCAFFARDI, L.(eds). **Comparative law in legislative drafting**. The Hague: Eleven International Publishing, 2014.

BASTOS, M. S. Da Inclusão das Minorias e dos Grupos Vulneráveis: Uma vertente Eficaz e Necessária para a Continuidade da Ordem Jurídica Constitucional. **Revista Brasileira de Direito Constitucional**, n. 18, p. 39-69,2011.

BAUMAN, Z. **Comunidade:** a busca por segurança no mundo atual. Rio deJaneiro: Renovar, 2003.

BENG, R. V.; SLAGTER, R. Autonomous compliance: standing on the shoulders of RegTech! [*s.l.: s.n.*], **Compact**, 2017. Disponível em: https://www.compact.nl/en/articles/autonomous-compliance/. Acesso em: 13 ago. 2022.

BENSOUSSAN, A.; CHAMPION, R. **Droit de La Robotique**. Courbevoie:Primnext, 2013.

BENT, J. R., Is Algorithmic Affirmative Action Legal? (April 16, 2019). **Georgetown Law Journal**, Forthcoming, Stetson University College of Law Research Paper No. 2019-6.

BEZERRA NETO, B. A. O que define um julgamento e quais são os limites do juiz? São Paulo: Noeses, 2018, edição Kindle.

BOEING, D. H. A.; ROSA, A. M. **Ensinando um robô a julgar: pragmática, discricionariedade, heurística e vieses no uso de aprendizado de máquina no Judiciário**. Florianópolis: EMais Academia, 2020.

BOBBIO, N. **O Positivismo Jurídico:** Lições de filosofia do direito.São Paulo: Ícone, 1995.

BOBBIO, N. **A era dos direitos.** Rio de Janeiro: Elsevier, 2004.

BROWNING, M.; ARRIGO, B. Stop and Risk: Policing, Data, and theDigital Age of Discrimination. **American Journal of Criminal Justice**, 7 ago. 2020. Disponível em: https://doi.org/10.1007/s12103-020-09557-x. Acesso em:1 dez. 2023.

BRASIL. Levantamento Nacional de Informações Penitenciárias – **INFOPEN**. Brasília: Ministério da Justiça e Segurança Pública/Departamento Penitenciário Nacional, 2022. Disponível em: https://www.gov.br/depen/pt-br/servicos/sisdepen/sisdepen. Acesso em: 07 jun. 2022.

BRASIL. Supremo Tribunal Federal. Inteligência artificial vai agilizar a tramitação de processos no STF. **Notícias STF**, Brasília, DF, 30 maio 2018. Disponível em: https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=388443&ori=1. Acesso em: 17 set. 2022.

BREHM, K. *et al*. **O futuro da IA no Sistema Judiciário Brasileiro:** mapeamento, integração e governança da IA. [*S.l.*]: Where the word connect, [20--]. Disponível em: https://itsrio.org/wp- content/uploads/2020/07/TRADUC%CC%A7A%CC%83O-The-Future-of-AI-in- the-Brazilian-Judicial-System.pdf. Acesso em: 17 set. 2022.

BUCHANAN, B. G.; HEADRICK, T. E. Some Speculation about Artificial Intelligence and Legal Reasoning. **Stanford Law Review**, v. 23, n. 1,
p. 40, nov. 1970. Disponível em: https://doi.org/10.2307/1227753. Acesso em: 1dez. 2023.

BULFINCH, T. **O livro de ouro da Mitologia**: história de deuses e heróis. 26.ed. Rio de Janeiro: Ediouro, 2002.

BUREAU OF JUSTICE STATISTICS. **Key statistic**: total correctionalpopulation. [S.l.]: BJS, [2017].

CALPADO, G. Z. **The Pillars of Global Law**. Hampshire: Ashgate Publishing Limited, 2008.

OLIVEN, A. C. Ações afirmativas, relações raciais e política de cotas nas universidades: Uma comparação entre os Estados Unidos e o Brasil. **Educação**, *[S. l.]*, v. 30, n. 1, 2007. Disponível em: https://revistaseletronicas.pucrs.br/faced/article/view/539. Acesso em: 5 nov. 2024.

CANADIAN GOVERNMENT. **Artificial Intelligence and Data Act (AIDA)**. Disponível em: https://www.parl.ca/legisinfo/en/bill/44-1/c-27. Acesso em: 21 jul. 2024.

CARRILLO CASTILLO, L.. El concepto kantiano de ciudadanía. **Estudios deFilosofí**a, n. 42, p. 103-121, 15 jul. 2010. Disponível em:
https://doi.org/10.17533/udea.ef.11595. Acesso em: 1 dez. 2023.

CARSON, E. A. **Jail inmates in 2021: Statistical tables**. Washington, DC: Bureau of Justice Statistics, 2023. Disponível em: https://bjs.ojp.gov/library/publications/jail-inmates-2021-statistical-tables. Acesso em: 21 jul. 2024.

CARVALHO, P. B. **Direito tributário, linguagem e método**. São Paulo: Noeses, 2013.

CHAPPELL, T. **Ethics and Experience:** Life Beyond Moral Theory. New York:Routledge, 2009.

CHARPA, U. Synthetic Biology and the Golem of Prague: PhilosophicalReflections on a Suggestive Metaphor. **Perspectives in Biology and Medicine**, v. 55, n. 4, p. 554-570, 2012. Disponível em:
https://doi.org/10.1353/pbm.2012.0036. Acesso em: 1 dez. 2023.

CHEN, S. China's Court AI reaches Every corner of justice system, advising judges and streamling punishment. **South China Morning Post**. Disponível em:
https://amp.scmp.com/news/china/science/article/3185140/chinas-court-ai- reaches-every-corner-justice-system-advising. Acesso em: 23 nov. 2022.

CHRISLEY, R. A Humem-centered Approach to AI Ethics: A Perspective from Cognitive Science. *In*: DUBBER, M. D. ; PASQUALE, F ; DAS, S. **The Oxford Handbook of Ethics of AI**. Oxford : Oxford University Press, 2020. P. 463-474.

CLARKE, P. B. **Ser ciudadano: conciencia y práxis**. Madrid: Sequitur, 2010.

COELHO, A. Z. As 7 tendências para o uso de inteligência artificial no direito em 2018. [Canadá]: **Thomson Reuters**, [2019]. Disponível em: https://www.thomsonreuters.com.br/content/dam/openweb/documents/pdf/Brazil /whitepaper/As_7_Tend%C3%AAncias_para_o_uso_da_Inteligencia_Artificial_ EM_2018.pdf. Acesso em: 08 jul. 2021.

COMPARATO, F. K.. **A afirmação histórica dos direitoshumanos**. 10. ed. São Paulo: Saraiva, 2015.

COMMISSION EUROPEENNE POUR L'EFFICACITE DE LA JUSTICE. **Charte éthique européenne d'utilisation de l'intelligence artificielle dans lessystèmes judiciaires et leur environnement**. Estrasburgo: Conseil de l'Europe, 2018. Disponível em: https://rm.coe.int/charte-ethique-fr-pour- publication-4-decembre-2018/16808f699b. Acesso em: 9 nov. 2020.

CONSELHO NACIONAL DE JUSTIÇA. **Inteligência artificial no poderjudiciário brasileiro**. Brasília, DF: CNJ, 2023.

CONSELHO NACIONAL DE JUSTIÇA. **Relatório Justiça em Números 2023**.Brasília, DF: CNJ, 2023. Disponível em: https://www.cnj.jus.br/wp-content/uploads/2023/09/sumario-executivo-justica-em-numeros-200923.pdf. Acesso em: 24 out. 2023.

COGLIANESE, C.; MENDELSON, E. **Meta-Regulation and Self-Regulation**. Disponível em: <https://ssrn.com/abstract=2002755>. Acesso em: 22 out. 2024.

CONSIGLIO, E. **Che cosa à la discriminazione?** Un'introduzione teorica aldiritto antidiscriminatorio. Torino: G. Giappichelli Editore, 2020.

CORTE INTERNACIONAL DE JUSTIÇA. **LaGrand Case**. Corte Internacional de Justiça, 27 jun. 2001. Disponível em: https://www.icj-cij.org/sites/default/files/case-related/104/104-20010627-JUD-01-00-EN.pdf. Acesso em: 23 de agosto de 2024.

CREEMERS, R.; TONER, H; WEBSTER, G.. Translation: Internet InformationService Algorithmic Recommendation Management Provisions – Effective March 1, 2022. **Digichina**, Stanford, 10 Janeiro 2022. Disponível em: https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/. Acesso em: 22 nov. 2022.

CUMMINGS, Mary L. Automation and Accountability in Decision Support System Interface Design. **The Journal of Technology Studies**, v. 32, n. 1, 1 set. 2006. Disponível em: https://doi.org/10.21061/jots.v32i1.a.4. Acesso em: 1 dez. 2023.

DAVIERA, A. L. et al. Risk, race, and predictive policing: A critical race theory analysis of the strategic subject list. **American Journal of CommunityPsychology**, 17 abr. 2023. Disponível em: https://doi.org/10.1002/ajcp.12671.Acesso em: 1 dez. 2023.

DELANEY, D. Legal Geography I: constitutivities, complexities and contingencies. **Progress in Human Geography**, New York, v. 39, n. 1, p. 96- 102, 2015. Disponível em: http://eds.a.ebscohost.com/eds/pdfviewer/pdfviewer?vid=0&sid=4161aabc- 023d-4266-b1b9-0eb3c7fd3f01%40sdc-v-sessmgr01. Acesso em: 12 out. 2022.

DEMICHELE, M. et al. The Public Safety Assessment: A Re-Validationand Assessment of Predictive Utility and Differential Prediction by Race and Gender in Kentucky. **SSRN Electronic Journal**, 2018. Disponível em: https://doi.org/10.2139/ssrn.3168452. Acesso em: 1 dez. 2023.

DIAKOPOULOS, N. Transparency. *In*: DUBBER, M. D.; PASQUALE, F; DAS, S. **The Oxford Handbook of Ethics of AI**. Oxford: Oxford University Press, 2020.

DIGNUM, V. Responsibility and Artificial Intelligence. *In*: DUBBER, M. D. ; PASQUALE, F ; DAS, S. **The Oxford Handbook of Ethics of AI**. Oxford : Oxford University Press, 2020. P. 216-231.

DILMEGANI, C. Bias in AI: What it is, Types, Examples & 6 Ways to Fix it in 2024. Ago 2023. Disponível em: https://research.aimultiple.com/ai-bias/. Acesso em 10 setembro de 2024.

DOUZINAS, C.. **O Fim dos Direitos Humanos**. São Leopoldo: Unisinos,2009.

DJEFFAL, C.. Sustainable AI Development (SAID): On the Road to MoreAccess to Justice. **SSRN Electronic Journal**, 2018. Disponível em: https://doi.org/10.2139/ssrn.3298980. Acesso em: 1 dez. 2023.

ELDRIDGE, S. Moral Virtue. *In*: Encyclopaedia Britannica (*on-line*). Disponível em: https://www.britannica.com/topic/moral-virtue. Acesso em: 20 jul. 2024.

ENCICLOPEDIA Treccani. Legge di Moore. Disponível em: https://www.treccani.it/enciclopedia/legge-di-moore_%28Enciclopedia-della-Scienza-e-della-Tecnica%29/. Acesso em: 20 jul. 2024.

EUROPEAN PARLIAMENT. Motion for a European Parliament Resolution.[*S.l.*]: **European Parliament**, 2017. Disponível em: https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html. Acesso em: 18 set. 2019.

ETZIONI, O. How to Regulate Artificial Intelligence. **The New York Times**, [NewYork], 1 set. 2017. Disponível em: https://www.nytimes.com/2017/09/01/opinion/artificial-intelligence-regulations- rules.html. Acesso em: 18 set. 2019.

FANG, Hanming; MORO, Andrea. **Theories of Statistical Discrimination andAffirmative Action:** A Survey. Cambridge, MA: National Bureau of Economic Research, 2010. Disponível em: https://doi.org/10.3386/w15860. Acesso em: 1dez. 2023.

FELIPE, B. F. da C.; PERROTA, R. P. C. Inteligência artificial no direito: umarealidade a ser desbravada. **Revista de Direito, Governança e Novas Tecnologias**, Salvador, v. 4, n. 1, p. 1-16, 2018. Disponível em: https://egov.ufsc.br/portal/conteudo/intelig%C3%AAncia-artificial-no-direito-%E2%80%93-uma-realidade-ser-desbravada. Acesso em: 13 ago. 2022.

FLORES, A. W.; BECHTEL, K.; LOWERKAMP, C. T. False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's SoftwareUsed Across the Country to Predict Future Criminals. And It's Biased Against Blacks". **Federal Probation**. 2016, Washington: vol. 80, n. 2, p. 38-46. Disponível em: https://www.uscourts.gov/sites/default/files/usct10024-fedprobation-sept2016_0.pdf Acesso em: 01 jun. 2023.

FLORIDI, L. **Etica dell'intelligenza Artificiale:** Sviluppi, opportunità, sfide.Milão: Raffaello Cortina Editore, 2022.

FOUCAULT, M. As palavras e as coisas: uma arqueologia das ciências humanas. Martins Fontes: São Paulo, 2000.

FRANK, M. Racial Profiling: better safe than sorry. **Miami Herald**. 19/10/1999.Disponível em: www.newspaper.com/image/618127697. Acesso em: 10 jul. 2023.

GABBAT, A. **IBM computer Watson wins Jeopardy clash**. The Guardian, [Kings Place], 17 fev. 2011. Disponível em: https://www.theguardian.com/technology/2011/feb/17/ibm-computer-watson-wins-jeopardy. Acesso em: 04 jun. 2022.

GEBRU, T. Race and Gender. *In*: DUBBER, M. D. ; PASQUALE, F ; DAS, S. **The Oxford Handbook of Ethics of AI**. Oxford : Oxford University Press, 2020. P. 253-269.

GIACOMELLI, L. Big brother is "gendering" you. Il diritto antidiscriminatorio alla prova dell'intelligenza artificiale: quale tutela per il corpo digitale?. **BioLaw Journal - Rivista di BioDiritto**, v. 17, n. 2, p. 269–297, 17 jul. 2019.

GILLESPILES, T . The Relevance of Algorithms, *in* **Media Technologies: Essays on Communication, Materiality, and Society**. Ed. Gillespie, T.; BOCZKOWSKI, P. J; FOOT, K. A. Cambridge: The MIT Press, 2014, p. 167-194.

GOMES, J. B. B. **Ação afirmativa & Princípio Constitucional da Igualdade:**o direito como instrumento de transformação social. A experiência dos EUA. Rio de Janeiro: Renovar, 2001.

GREEN, B.; CHEN, Y. Disparate Interactions: Na Algorithm-in-the-Loop Analysisof Fairness in Risk Assessments. In **Proceedings of the Conference on Fairness, Accountability, and Transparency** (FAT* '19). Association forComputing Machinery, New York, NY, USA, 90–99.2019 https://doi.org/10.1145/3287560.3287563

GRENNAN, L. *et al*. Why businesses need explainable AI—and how to deliver it? Disponível em: https://www.mckinsey.com/capabilities/quantumblack/our-insights/why-businesses-need-explainable-ai-and-how-to-deliver-it#/. Acesso em: 05 de julho de 2024.

HABERMAS, J. **Era das Transições**. Tradução de Flávio BenoSiebeneichler. Rio de Janeiro: Tempo Brasileiro, 2003

HABERMAS, J. Estudos preliminares e complementares. In: **Fracticidade e validade**. São Paulo: Unesp, 2020.

HAMILTON, M. Risk-need Assessment: Constitutional and Ethical Challenges. **American Criminal Law Review**, Washington, v. 52, 2015.

HAN, B. **Infocracia: digitalização e a crise da democracia**. Petrópolis: Vozes, 2022.

HARRIS, D. A. U.S. experiences with racial and ethnic profiling: history, current issues, and the future. **Critical Criminology**, v. 14, n. 3, p. 213-239, set.2006. Disponível em: https://doi.org/10.1007/s10612-006-9011-3. Acesso em: 1 dez. 2023.

HÁRS, A.. Conceptual Difficulties in the Transformation of Human Rights to the Realm of Artificial Intelligence. **Acta Humana**. 12. 123-135. Disponível em: https://folyoirat.ludovika.hu/index.php/actahumana/article/view/7275/5943. Acesso em: 29 ago. 2024.

HEAVEN, W. D. The UK is dropping an immigration algorithm that critics say isracist. **MIT Technology Review**, [*s.l.*], 5 ago. 2020. Disponível em: https://www.technologyreview.com/2020/08/05/1006034/the-uk-is-dropping-an-immigration-algorithm-that-critics-say-is-racist/. Acesso em: 13 jan. 2021.

HENDERSON, W. D. **A Blueprint for Change**. Bloomington: Legal StudiesResearch Paper Series, 2013.

HOOFT, S. van. **Understanding Virtue Ethics.** Chesham: Acumen PublishingLimited, 2006.

HYATT, J.; CHANENSON, S. L., The Use of Risk Assessment at Sentencing: Implications for Research and Policy (December 2016). **VillanovaLaw/Public Policy Research Paper** No. 2017-1040. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2961288 Acesso em 30 nov. 2023.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Cor ou raça**. Disponível em: https://educa.ibge.gov.br/jovens/conheca-o-brasil/populacao/18319-cor-ou-raca.html. Acesso em: 21 jul. 2024.

ISIN, E. F.; PETER N.. **Routledge Handbook of Global CitizenshipStudies**. London New York: Routledge, 2014.

IZDEBSKI, K. *et al*. **alGOVrithms** – State of Play. EPanstwo Foundation. 2019.Disponível em: https://epf.org.pl/en/projects/algorithms/. Acesso em 29 ago. 2023.

KAHN, K. B.; DAVIES, P. G. What Influences Shooter Bias? The Effects of Suspect Race, Neighborhood, and Clothing on Decisions to Shoot. **Journal of Social Issues**, v. 73, n. 4, p. 723-743, dez. 2017. Disponívelem: https://doi.org/10.1111/josi.12245. Acesso em: 1 dez. 2023.

KANT, I.. **A fundamentação da Metafísica dos Costumes**. Lisboa,Portugal: Edições 70, 2007.

Kelling, G. L.; WILSON, J. Q. The Atlantic Monthly; March 1982; **BrokenWindows**; Volume 249, No. 3; pages 29-38.

KLEINBERG, J. et al. Human Decisions and Machine Predictions*. **TheQuarterly Journal of Economics**, 26 ago. 2017. Disponível em: https://doi.org/10.1093/qje/qjx032. Acesso em: 1 dez. 2023.

KLEMME, H. F. Direito à justificação – dever de justificação: reflexões sobre um modus de fundamentação dos direitos humanos. ***Trans/Form/Ação*,**Marília, v.35, n. 2, p. 187-198, maio/ago. 2012.

KRAMER, X. E; GELDER, E.; THEMELI, E. e-Justice in the Netherlands: the Rocky Road to Digitised Justice. In: WELLER, M; WENDLAND, M. (eds.). **Digital Single Market: Bausteine eines Rechts in der Digitalen Welt**. Tübingen: Mohr Siebeck 2018. p. 209-235. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3167543. Acesso em: 13 ago. 2022.

KUIPERS, B. Perspectives on Ethics of AI. *In*: DUBBER, M. D.; PASQUALE, F ;DAS, S. **The Oxford Handbook of Ethics** of AI. Oxford: Oxford University Press, 2020. P. 421-441.

LARSON, J, MATTU, S.;KIRCHNER, L.; ANGWIN, J. How WeAnalyzed the COMPAS Recidivism Algorithm. **ProPublica** (blog). Disponível em: https://www.propublica.org/article/how-we-analyzed-the-compas- recidivism-algorithm. Acesso em: 13 jul 2023.

LOPES, L.; VIEIRA, R. **Processamento de linguagem natural e o tratamento computacional de linguagens científicas**. Porto Alegre: Linguagens Especializadas in Corpora, 2010.

LUGER, G. F. **Artificial intelligence: structures and strategies for complexproblem solving.** Essex: Addison-Wesley, 2005.

MANYIKA, J. *et al*. A future that works: automation, employment and productivity. [*S.l.*]: **McKinsey & Company**, 2017. Disponível em: https://www.mckinsey.com/mgi/overview/2017-in-review/automation-and-the-future-of-work/a-future-that-works-automation-employment-and-productivity. Acesso em: 18 set. 2021.

MARSHALL, T. H. **Cidadania, Classe Social e Status**. Riode Janeiro: Zahar Editores, 1967.

MARTIN, N. B. Algumas Reflexiones sobre la Infomática Jurídica Decisional. *In*:BAEZ, N. L. X. (org.) *et al*. **O Impacto das novas tecnologias nos direitos fundamentais**. Joaçaba: Editora Unoesc, 2018.

MARX, K. **Crítica do Programa de Gotha**. Trad. Rubens Enderle. São Paulo:Boitempo, 2012.

MAZARIO, J. M. C. **Las Naciones Unidas y la Protección de las Minorías Religiosas:** de la tolerancia a la interculturalidad. Tiran Monografías. España,Universidad de Sevilla Pablo D'Olavide, 1997.

MAZZUOLI, V. DE O. **Curso de direitos humanos**. 6. ed. São Paulo:MÉTODO, 2019.

MEIJER, A.; WESSELS, M. Predictive Policing: Review of Benefits andDrawbacks. **International Journal of Public Administration**, v. 42, n. 12, p. 1031-1039, 12 fev. 2019. Disponível em: https://doi.org/10.1080/01900692.2019.1575664. Acesso em: 1 dez. 2023.

MILL, J. S. **Utilitarianism.** Ontario: Batoche Books, 2001.

MINISTÈRE DE LA JUSTICE. **L'open data des décisions de justice**. [*S.l.*]: MJ, [2017]. Disponível em: http://www.justice.gouv.fr/publication/open_data_rapport.pdf. Acesso em: 19 jul.2022.

MONAHAN, J.; SKEEM, J. L. Risk assessment in criminal sentencing. **AnnualReview of Clinical Psychology**, [*s.l.*], v. 12, p. 489-513, 2016. Disponível em: https://gspp.berkeley.edu/assets/uploads/research/pdf/09-2015_Risk_Assessment_in_Criminal_Sentencing.pdf. Acesso em: 13 ago. 2022.

MINISTÉRIO DA JUSTIÇA E SEGURANÇA PÚBLICA. **Infopen: Levantamento Nacional de Informações Penitenciárias**. Disponível em: https://dados.mj.gov.br/dataset/infopen-levantamento-nacional-de-informacoes-penitenciarias/resource/54cdab5b-b241-4dcc-83af-43cba0250ef3. Acesso em: 21 jul. 2024.

MOOR, J. **Four Kinds of Ethical Robots**. Disponível em: https://philpapers.org/rec/MOOFKO. Acesso em: 10 abr. 2022.

NAKAD-WESTSTRATE, H. W. R. *et al*. Digitally produced judgements in modern courts proceedins. **International Journal for Digital Society**, Londres, v. 6, n. 4, 2015. Disponível em: https://scholarlypublications.universiteitleiden.nl/handle/1887/42379. Acesso em: 13 ago. 2022.

NAÇÕES UNIDAS. **Treaty Series: Multilateral Treaties of the United Nations – Convention on the Rights of the Child**. Disponível em: https://treaties.un.org/Pages/ViewDetails.aspx?src=TREATY&mtdsg_no=IV-11&chapter=4&clang=_en. Acesso em: 29 ago. 2024.

NAÇÕES UNIDAS. **Convenção de Viena sobre o Direito dos Tratados.** Disponível em: https://legal.un.org/ilc/texts/instruments/english/conventions/1_1_1969.pdf. Acesso em: 29 ago 2024.

NATORSKI, M. Compromise in multilateral negotiations and the global regulation of artificial intelligence. **Democratization**, v. 31, n. 5, p. 1091–1116, 2024.

NETTO, I. **TRT-PR cria robô capaz de economizar milhares de horas detrabalho humano**. Curitiba: TRT, 2021. Disponível em: https://www.trt9.jus.br/portal/noticias.xhtml?id=7055109. Acesso em: 17 set.2022.

NILSSON, N. J. **The quest for artificial intelligence**: a history of ideas and achievements. Cambridge: Cambridge University Press, 2010.

OGUS, A. **Regulation – Legal Form and Economic Theory**. Hart Publishing, 2004.

OLIVEIRA, S. R.; COSTA, R. S. Pode a Máquina Julgar? Considerações sobreo Uso de Inteligência Artificial no 25 Processo de Decisão Judicial. **Revista deArgumentação e Hermeneutica Jurídica**, Florianópolis, v. 4, n. 2, 2018.
Disponível em: https://www.indexlaw.org/index.php/HermeneuticaJuridica/article/view/. Acessoem: 12 out. 2022.

ONEAL, C.. **Algoritmos de destruição em massa:** como o *big data* aumenta adesigualdade e ameaça a democracia. Trad. Rafael Abraham. 1. Ed. Santo André: Editora Rua do Sabão, 2020.

ONU - Organização das Nações Unidas. **Declaração Universal dos DireitosHumanos da ONU**. Disponível em: https://www.ohchr.org/EN/UDHR/Pages/Language.aspx?LangID=por. Acesso em: 29 fev. 2022.

ONU – Organização das Nações Unidas**. Human Rights Due Diligence - Interpretive Guide**. Sep 2021. Disponível em: https://www.undp.org/sites/g/files/zskgke326/files/2022-10/HRDD%20Interpretive%20Guide_ENG_Sep%202021.pdf. Acesso em: 20 jul. 2024.

ONU – Organização das Nações Unidas. ***General comment No. 20: Non-discrimination in economic, social and cultural rights (art. 2, para. 2, of the International Covenant on Economic, Social and Cultural Rights)***. Disponível em: https://www.refworld.org/docid/4a60961f2.html. Acesso em: 20 jul. 2024.

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

PASQUALE, F. **The Black Box Society: the secret algorithms that control money and information**. Cambridge: Harvard University Press, 2015.

PETLAND, A. **Social Physics**. Penguin Books: New York, 2015.

PIETROPAOLI; SIMONCINI, A. **Ernie: l'algoritmo "comunista" e il futuro che (non) vogliamo per l'IA**. Disponível em: <https://www.agendadigitale.eu/cultura-digitale/ernie-lalgoritmo-comunista-e-il-futuro-che-vogliamo-per-lia/>. Acesso em: 8 mar. 2023.

PINKSTONE, J. AI-powered JUDGE created in Estonia will settle small courtclaims of up to £6,000 to free up professionals to work on bigger and more important cases. **Mailonline**, [s.l.], 26 mar. 2019. Disponível em: https://www.dailymail.co.uk/sciencetech/article-6851525/Estonia-creating-AI- powered-JUDGE.html. Acesso em: 04 set. 2022.

PIOVESAN, F.. **Direitos Humanos: desafios da ordem internacional contemporânea.** Cadernos de Direito Constitucional, [*s.l.*], p. 5-26, 2006.

POISSON, S.. **Recherches Sur la Probabilité des Jugements enMatière Criminelle et en Matière Civile:** Précédées des Règles Générales duCalcul des Probabilités. [S. l.]: Creative Media Partners, LLC, 2018. ISBN 9780270821215.

PRESCOTT, R.; MARIANO, R. Victor, a IA do STF, reduziu tempo de tarefade 44 minutos para cinco segundos. **Convergência Digital**, [s.l.], 17 out. 2019. Disponível em: https://www.convergenciadigital.com.br/cgi/cgilua.exe/sys/start.htm?UserActiveTemplate=site&infoid=52015&sid=3. Acesso em: 19 maio 2020.

PROSSER, T. The Regulatory Enterprise: Government, Regulation, and Legitimacy. *in* **Oxford Academic.** Disponível em: https://doi.org/10.1093/acprof:oso/9780199579839.001.0001 . Acesso em 22 fev. 2024.

RACHOVITSA, A; JOHANN, N. The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch *SyRI* Case. **Human Rights Law Review**, . Disponível em: https://doi.org/10.1093/hrlr/ngac010. Acesso em: 21 nov 2023.

RAMOS, A. C. **Curso de Direitos Humanos**. 7. ed. SãoPaulo: Saraiva, 2020.

RASO, F. *et al*. **Artificial intelligence & human rights**: **opportunities & risks.** [*S.l.*]: Berkman Klein Center for Internet & Society Research Publication, 2018.Disponível em: http://nrs.harvard.edu/urn-3:HUL.InstRepos:38021439. Acessoem: 19 jan. 2021.

RISSLAND, E. L. Artificial Intelligence and Law: Stepping Stones to a Model of Legal Reasoning. **The Yale Law Journal**, v. 99, n. 8, p. 1957, jun. 1990. Disponível em: https://doi.org/10.2307/796679. Acesso em: 1 dez. 2023.

ROCHA, I. Ativistas denunciam Inteligência Artificial racista, após afro- americano ser preso por erro no programa. **Notícia Preta**, [*s.l.*], 26 jul. 2020. Disponível em: https://noticiapreta.com.br/ativistas-alertam-sobre-inteligencia-artificial-racista/. Acesso em 13 jan. 2021.

RUSSELL, S.; NORVIG, P. **Inteligência artificial**. Rio de Janeiro: Elsevier,2013.

SACCO, R. **Introduzione al Diritto Comparato.** Torino: Unione Tipografico-Editrice Torinense, 1992.

SALAS, J. Se está na cozinha, é uma mulher: como os algoritmos reforçam preconceitos. **El País**, [*s.l.*], 23 set. 2017. Disponível em: https://brasil.elpais.com/brasil/2017/09/19/ciencia/1505818015_847097.html.Acesso em: 13 jan. 2021.

SALOMÃO FILHO, C. **Regulação da atividade econômica: (princípios e fundamentos jurídicos)**. Imprenta: São Paulo, Malheiros, 2008.

SANCTIS, F. M. de. **Inteligência Artificial e Direito**. São Paulo: Almedina, 2020.

SAUL, J. Scepticism and Implicit Bias. **Disputatio**, v. 5, n. 37, p. 243-263, 1 nov. 2013. Disponível em: https://doi.org/10.2478/disp-2013-0019. Acesso em: 1 dez. 2023.

SAUNDERS, J.; HUNT, P.; HOLLYWOOD, J. S. Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot. **Journal of Experimental Criminology**, v. 12, n. 3, p. 347-371, 12 ago. 2016. Disponível em: https://doi.org/10.1007/s11292-016-9272-0. Acesso em: 1dez. 2023.

SAUVÉ, J.-M. **Intervention de Jean-Marc Sauvé à l'occasion du colloque organisé à l'occasion du bicentenaire de l'Ordre des avocats au Conseild'État et à la Cour de cassation le 12 février 2018**. Paris: Conseil d'État, 2018. Disponível em: https://www.conseil-etat.fr/actualites/discours-et- interventions/la-justice-predictive#_ftn2. Acesso em: 21 out. 2022.

SCARCIGLIA, R. **Metodi e Comparazione Giuridica**. Wolters Kluwer: Milano,2018.

SCHAUER, F. **Profiles, Probabilities, and Stereotypes**. Cambridge: HarvardUniversity Press, 2003.

SCOTT, C. D. Standard-Setting in Regulatory Regimes. **SSRN Electronic Journal**, 2009. Disponível em: https://doi.org/10.1093/oxfordhb/9780199560219.003.0006. Acesso em: 23 fev. 2024.

SECRETARIA NACIONAL DE POLÍTICAS PENAIS. **Relatório de Informações Penais** – 15º Ciclo SISDEPEN, 2º Semestre de 2023. Brasília, 2024. Disponível em: https://www.gov.br/senappen/pt-br/servicos/sisdepen/relatorios/relipen/relipen-2-semestre-de-2023.pdf. Acesso em 20 jul. 2024.

SEGUI, E. **Minorias e grupos vulneráveis:** uma abordagem jurídica. Rio deJaneiro: Forense, 2002.

Shafer-Landau, R. **Ethical Theory**: An Anthology. Second Edition.London, Wiley-Blackwell, 2013.

SIMONCINI, A. L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà. **BioLaw Journal – Rivista di BioDiritto**, Trento, n. 1, p- 63-89,2019.

SIMONCINI, A. Diritto costituzionale e decisioni algoritmiche. In: DORIGO, s.(org.). **Il Ragionamento giuridico nell'era dell'inteligenza artificiale**. Pisa:Pacini Editore, 2020.

SINGH, H.; SINGH, J.; SINGH, P. A System Approach to Identifying Structural Discrimination throught the Lens of Hate Crimes. **Asian American Law Journal**, vol. 20, no. 1. 2013, p. 107-138.

SMITS, J. M. Comparative Law and its influence on national legal sistems. In:REIMANN, M.; ZIMMERMANN, R. (org.). **The Oxford Handbook of Comparative Law**. Oxford: Oxford University Press, 2019.

SOURDIN, T. Judge v. Robot? Artificial Intelligence and Judicial Decision- Making. **University of New South Wales Law Journal**, Sidney, v. 41, n. 4, p.114-1133, 2018.

SOUZA, R. de. Batemos um papo com o robô advogado que já venceu 160 mil contestações. **TecMundo**, [*s.l.*], 28 jun. 2016. Disponível em: https://www.tecmundo.com.br/inteligencia-artificial/106644-batemos-papo-robo- advogado-venceu-160-mil-contestacoes.html. Acesso em: 13 ago. 2022.

STRASSER, F. Como um computador venceu o melhor jogador de xadrez domundo. **BBC News**, [Londres], 12 maio 2017. Disponível em: https://www.bbc.com/news/av/world-us-canada-39888639/how-a-computer- beat-the-best-chess-player-in-the-world. Acesso em: 07 jun. 2022.

STRECK, L. L. **Dicionário de Hermeneutica: Quarenta temas fundamentais da Teoria do Direito à luz da Crítica Hermenêutica do Direito**. Belo Horizonte: Letramento, 2017.

SUPREME COURT OF CANADA. **Report [1998] S SCR 217, 25506**. 2013. Disponível em: https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/17133/index.do. Acesso em: 21 jul. 2024.

SUSSKIND, R. Expert Systems In Law: A Jurisprudential Approach To Artificial Intelligence And Legal Reasoning. **The Modern Law Review**, v. 49, n. 2, p. 168-194, mar. 1986. Disponível em: https://doi.org/10.1111/j.1468-2230.1986.tb01683.x. Acesso em: 1 dez. 2023.

SUSSKIND, R. **Online courts and the future of justice**. Oxford: OxfordUniversity Press, 2019.

The White House Office of Science and Technology Policy (OSTP). ***AI Bill of Rights***. Disponível em: https://www.whitehouse.gov/ostp/ai-bill-of-rights/. Acesso em: 20 jul. 2024.

TSCHANTZ, M. C. **What is Proxy Discrimination?** In Proceedings of the 2022ACM Conference on Fairness, Accountability, and Transparency (FAccT '22). 2022, Association for Computing Machinery, New York, NY, USA, 1993–2003.

TOLAN, Songül et al. **Why Machine Learning May Lead to Unfairness**. In: ICAIL'19: SEVENTEENTH INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW, Montreal QC Canada**.** ICAIL '19: Seventeenth International Conference on Artificial Intelligence and Law. New York, NY,USA: ACM, 2019. ISBN 9781450367547. Disponível em: https://doi.org/10.1145/3322640.3326705. Acesso em: 1 dez. 2023.

TONETTO, M. C.. A dignidade da humanidade e os deveresem Kant. **Revista Filos**. Autora, Curitiba, v. 24, n. 34, p. 265-285, 2012.

TORONTO DECLARATION. **The Toronto Declaration on Protecting the Rights to Equality and Non-Discrimination in Machine Learning Systems**. Disponível em: https://www.torontodeclaration.org/declaration-text/english/. Acesso em: 21 jul. 2024.

TURING, A. Computing Machinery and Intelligence. **Mind, v. LIX**, n. 236, p. 433-460, 1 out. 1950. Disponível em: https://doi.org/10.1093/mind/lix.236.433. Acesso em: 1 dez. 2023.

TZAFESTAS, S. G. **Roboethics: A Navigating Overview**. Londres: Springer,2016.

UCHÔAS, B. R. Inovações Tecnológicas aplicadas ao Direito: hiperracionalidade ou irracionalidade? *In*: REIS, I. (org.). **Diálogos sobreretórica e argumentação**. Curitiba: lteridade Editora, 2018. v. 4.

UNIÃO EUROPEIA. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A32016R0679. Acesso em: 21 jul. 2024.

União Europeia. European Union Artificial Intelligence Act. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf. Acesso em: 21 jul. 2024.

UNITED NATIONS. **Human Rights**. [*s.l.*]: United Nations, [20--]. Disponível em: https://www.un.org/en/global-issues/human-rights. Acesso em: 13 abr. 2022.

UNITED STATES 116th Congress. ***H.R. 6216: National Artificial Intelligence Iniciative Act of 2020***. 2020. Disponível em: https://www.congress.gov/bill/116th-congress/house-bill/6216. Acesso em: 20 jul. 2024.

UNITED STATES 117th Congress. ***H.R. 6580: Algorithmic Accountability Act of 2022***. 2022. Disponível em: https://www.congress.gov/bill/117th-congress/house-bill/6580/cosponsors. Acesso em: 21 jul. 2024.

UNITED STATES CENSUS BUREAU. **QuickFacts: United States**. Disponível em: https://www.census.gov/quickfacts/fact/table/US/IPE120221. Acesso em: 21 jul. 2024.

UNITED STATES. **State v. Loomis**. 2016 WI 34. Supreme Court of Wisconsin, 2016. Disponível em: https://cases.justia.com/wisconsin/supreme-court/2016-2015ap000157-cr.pdf?ts=1468415026. Acesso em: 21 jul. 2024.

UNIVERSITÈ DE MONTREAL. **Montreal Declaration for a Responsible Development of Artificial Intelligence**. Disponível em: https://montrealdeclaration-responsibleai.com/the-declaration/. Acesso em: 21 jul. 2024.

VASAK, K. A 30-year struggle; the sustained efforts to give force of law to the Universal Declaration of Human Rights. In: **The UNESCO Courrier: a window open on the world**. XXX, p. 28-29. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000048063 Acesso em: 01 dez. 2023.

WARWICK, K. **Artificial Intelligence**: the basics. Oxon: Routledge, 2012.

WATSON, A. **Legal Transplants**. Edimburgh: Scottish Academic Press, 1974.

WEAVER, J. F. Regulation of artificial intelligence in the Unites States. In: BARFIELD, W.; PAGALLO, U. **Research Handbook on the Law of Artificial Intelligence**. Cheltenham: Edward Elgar Publishing Limited, 2018. p. 155-212.

WHAT is e-Residency. Estônia: Republic of Estonia e-Residency, [201-]. Disponível em: https://learn.e-resident.gov.ee/hc/en-us/articles/360000711978-What-is-e-Residency. Acesso em: 21 out. 2022.

WINTERS, Ben. Algorithms in the Criminal Justice System: Pre-Trial Risk Assessment Tools. Washington: **EPIC.** Disponível em: https://epic.org/algorithmic-transparency/crim-justice/. Acesso em: 20 jul. 2022.

WORLD ECONOMIC FORUM. **How to prevent discriminatory outcomes in machine learning**. Cologny, CH: World Economic Forum, 2018. Disponível em: http://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf. Acesso em: 13 jan. 2021.

XU, J. *et al*. Foundations and applications of information systemsdynamics. **Engineering**, Disponível em: https://doi.org/10.1016/j.eng.2022.04.018. Acesso em: 1 dez. 2023.

YEUNG, K. ; HOWES, A. ; POGREBNA, G. AI Governance by Human Rights-Centered Design, Deliberation, and Oversight. *In*: DUBBER, M. D. ; PASQUALE, F ; DAS, S. **The Oxford Handbook of Ethics of AI**. Oxford : Oxford University Press, 2020. P. 77-106.

YEUNG, K. The Regulatory State. **The Oxford Handbook of Regulation**, p. 63–84. Disponível em: https://doi.org/10.1093/oxfordhb/9780199560219.003.0004. Acesso em 2 set. 2023.

ZAFFARONI, R. E. **Criminología:** aproximación desde um margen. Bogotá:Themis, 1988.

ZAVRŠNIK, A. Criminal justice, artificial intelligence systems, and humanrights. **ERA Forum**, v. 20, n. 4, p. 567-583, 20 fev. 2020. Disponível em: https://doi.org/10.1007/s12027-020-00602-0. Acesso em: 1 dez. 2023.

ZHANG, J.; HAN, Y. Algorithms Have Built Racial Bias in Legal System-Accept or Not? In: 2021 INTERNATIONAL CONFERENCE ON SOCIALDEVELOPMENT AND MEDIA COMMUNICATION (SDMC 2021), 2021, Sanya, China. **2021 International Conference on Social Development and Media Communication (SDMC 2021).** Paris, France: Atlantis Press, 2022. Disponível em: https://doi.org/10.2991/assehr.k.220105.224. Acesso em: 1 dez. 2023.