



**UNIVERSIDADE DE BRASÍLIA - UnB
FACULDADE DE DIREITO
CURSO DE GRADUAÇÃO EM DIREITO**

LUANA FERNANDES MONTEIRO

**IMPACTOS DOS VIESES DA INTELIGÊNCIA
ARTIFICIAL EM MATÉRIA PENAL.**

**BRASÍLIA - DF
2022**

LUANA FERNANDES
MONTEIRO

**IMPACTOS DOS VIESES DA INTELIGÊNCIA
ARTIFICIAL EM MATÉRIA PENAL.**

Monografia apresentada como requisito parcial para a obtenção do grau de Bacharela em Direito pela Faculdade de Direito da Universidade de Brasília.

Orientadora: Professora Doutora Cláudia Rosane Roesler

BRASÍLIA - DF
2022

LUANA FERNANDES
MONTEIRO

IMPACTOS DOS VIESES DA INTELIGÊNCIA ARTIFICIAL EM MATÉRIA PENAL.

Monografia aprovada como requisito parcial para a obtenção do grau de Bacharela em Direito pela Faculdade de Direito da Universidade de Brasília, pela banca examinadora composta por:

Professora Doutora Cláudia Rosane Roesler
(Orientadora – Presidente)

Professora Tainá Aguiar Junquillo
(Membro da Banca Examinadora)

Professor Frank Ned Santa Cruz de Oliveira
(Membro da Banca Examinadora)

AGRADECIMENTOS

Agradeço à Vida por todo o aprendizado e crescimento que me proporcionou até aqui.

Agradeço aos meus Pais, Ricardo e Fátima, pelo amor, educação e direcionamentos constantes.

Agradeço a todos os Mestres que me auxiliam a aprender, me conhecer e ser uma pessoa melhor a cada dia.

Agradeço aos Amigos que me acompanham na Jornada.

Agradeço à Universidade de Brasília, instituição que me acolheu e me permitiu permanecer em Brasília ao me admitir em seu quadro funcional. O ingresso na instituição fez nascer em mim o interesse pela vida acadêmica. Em síntese, ao pensar na UnB sinto profunda gratidão em meu coração, pois foi nesta Casa que minha vida passou por grandes mudanças e crescimento.

Agradeço a todos os Colegas da Procuradoria Federal junto à Universidade de Brasília, companheiros de trabalho, pelo conhecimento compartilhado e apoio.

Agradeço à minha Orientadora, Cláudia Rosane Roesler, e ao Grupo de Pesquisa Retórica, Argumentação e Juridicidades pelo acolhimento, ensinamentos e orientação na vida acadêmica.

Agradeço à Tainá Aguiar Junquilha que durante seu doutorado auxiliou minhas pesquisas de PIBIC e que muito contribuiu para o meu crescimento e é referência em minha jornada acadêmica.

RESUMO

O presente estudo teve como objetivo contribuir com o debate sobre os impactos da inteligência artificial em matéria penal. Nesse sentido, busca-se compreender o conceito de Inteligência Artificial, os vieses e os impactos, possibilidades e regulação da Inteligência Artificial em matéria penal. Em pesquisa exploratória e qualitativa, foi realizado levantamento bibliográfico, necessário para identificar o estágio e reflexões atuais sobre o tema para, de forma indutiva, chegar a reflexões e conclusões sobre o tema. Conclui-se pela impossibilidade de utilização da máquina em decisões judiciais em matéria penal em razão da não neutralidade da Inteligência Artificial e das dificuldades de identificação e correção de vieses. Em casos em que a máquina sirva como auxiliar da decisão, é necessário refletir sobre como ela será utilizada, para proteção da dignidade da pessoa humana e princípios constitucionais, como o devido processo legal, mitigando vieses. É urgente a necessidade de regulação sobre o tema após amplo debate com especialistas, principalmente, por ser necessário ao legislador e à sociedade civil melhor compreensão sobre o que é e como a inteligência artificial funciona.

Palavras-chave: Inteligência Artificial. Vieses decisórios. Direito Penal. Direito Processual Penal.

ABSTRACT

The present study aimed to contribute to the debate on the impacts of artificial intelligence in criminal matters. In this sense, we seek to understand the concept of Artificial Intelligence, the biases and impacts, possibilities and regulation of Artificial Intelligence in criminal matters. In exploratory and qualitative research, a bibliographic survey was carried out, necessary to identify the stage and current reflections on the subject to, inductively, reach reflections and conclusions on the subject. It concludes by the impossibility of using the machine in judicial decisions in criminal matters due to the non-neutrality of Artificial Intelligence and the difficulties of identifying and correcting biases. In cases where the machine serves as an aid to the decision, it is necessary to reflect on how it will be used, to protect the dignity of the human person and constitutional principles, such as due process of law, mitigating biases. There is an urgent need for regulation on the subject after extensive debate with experts, mainly because it is necessary for the legislator and civil society to have a better understanding of what artificial intelligence is and how it works.

Keywords: Artificial Intelligence. Decision Bias. Criminal Law. Criminal Procedure.

SUMÁRIO

1. INTRODUÇÃO	9
2. COMPREENDENDO A INTELIGÊNCIA ARTIFICIAL	10
2.1. HISTÓRIA E CONCEITOS DA INTELIGÊNCIA ARTIFICIAL	10
2.1.1. <i>Machine Learning e Deep Learning Neural Networks</i>	11
2.2. APLICAÇÃO GERAL E AO DIREITO	13
2.2.1. Inteligência Artificial fraca e forte	13
2.2.2. Funcionamento geral e aplicação do <i>Deep Learning</i>	13
2.2.3. Inteligência Artificial aplicada ao Direito	15
2.3. APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL NO JUDICIÁRIO BRASILEIRO	17
3. COMPREENDENDO OS VIESES	21
3.1. O VIÉS NAS DECISÕES HUMANAS	21
3.1.1. Vieses em decisões judiciais	23
3.1.2. Vieses no sistema prisional brasileiro	24
3.2. A PRESENÇA DOS VIESES EM APLICAÇÕES DE INTELIGÊNCIA ARTIFICIAL	28
3.3. VIESES NA INTELIGÊNCIA ARTIFICIAL APLICADA AO DIREITO: O SISTEMA COMPAS	31
4. IMPACTOS, POSSIBILIDADES E REGULAÇÃO NO USO DA INTELIGÊNCIA ARTIFICIAL EM MATÉRIA PENAL	33
4.1. IMPACTOS DA INTELIGÊNCIA ARTIFICIAL APLICADA À MATÉRIA PENAL	33
4.2. POSSIBILIDADES DE APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL EM MATÉRIA PENAL.....	43
4.3. REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL.....	47
5. CONSIDERAÇÕES FINAIS	56
REFERÊNCIAS BIBLIOGRÁFICAS	57

LISTA DE FIGURAS

Figura 1 – Projetos de Inteligência Artificial do Judiciário Brasileiro.....	19
Figura 2 – Distribuição de presos por raça no sistema prisional.....	25
Figura 3 – Frequências dos assuntos dos processos segundo órgão julgador	46
Figura 4 – Frequências dos resultados dos processos segundo relator.....	47

LISTA DE TABELAS

Tabela 1 – Projetos de Inteligência Artificial do Poder Judiciário Brasileiro.....	18
--	----

1. INTRODUÇÃO

A Inteligência Artificial (IA) está presente no dia a dia das pessoas aplicada em diferentes áreas do conhecimento. É considerada uma tecnologia disruptiva, ou seja, que provocou e ainda provocará grandes mudanças na sociedade, a ponto de já indicar que se vive hoje diante de uma mudança de paradigma e em um período de transição. Estamos em plena Quarta Revolução Industrial¹.

Trata-se de uma revolução digital em que sensores se tornam mais poderosos e mais baratos e com a IA é possível maior interação entre ambientes físicos e digitais. A atual revolução se propaga de forma ampla e rápida, em uma velocidade que supera as mudanças históricas que se tem conhecimento, e tende a provocar mudanças profundas em todo o mundo. A Inteligência Artificial progride absurdamente e toda essa mudança afetará a forma como a humanidade vive. (SCHWAB, 2016, p. 23)

Porém, a velocidade de seu crescimento pode se tornar diretamente proporcional ao aumento das desigualdades sociais que pode provocar. Muitas são as reflexões e desafios que surgem nesse contexto de mudanças tão radicais. O Direito não está de fora dessa revolução, sendo, como todas as áreas do conhecimento, impactado e alterado pela nova tecnologia.

E o risco é que o processo pode dar azo a diversas violações como ocorreu em algumas áreas nas quais a aplicação da IA já apresentou resultados com erros chamados de vieses. Na análise de risco de empréstimos a IA atribuía alto risco a negros e hispânicos, o reconhecimento facial do Google que “identificou pessoas negras como gorilas”, o LinkedIn que apresentava preferência de contatos do sexo masculino e o aplicativo SimSimi que “foi suspenso no Brasil por reproduzir ameaças, palavrões e conversas de teor sexual”. (DIERLE; MARQUES, 2018, p. 7)

O campo do Direito parece não estar atento a essas questões e é necessário estabelecer regras para a IA neste campo (DIERLE; MARQUES, 2018, p. 7). A relevância do tema, principalmente diante dos desafios éticos que estão sendo abordados e discutidos em toda a sociedade oferecem a dimensão dos perigos que podem surgir de uma aplicação da IA sem reflexões, limites ou qualquer regulação. Os avanços e impactos desta nova tecnologia puderam e podem melhorar a vida de

¹ BBC. O que é a 4ª revolução industrial - e como ela deve afetar nossas vidas. Disponível em <https://www.bbc.com/portuguese/geral-37658309>. Acesso em 17 de abril de 2022.

muitas pessoas, mas também podem perpetuar e aprofundar desigualdades sociais e preconceitos, o que é relevante e preocupante se esta tecnologia for aplicada ao Direito. Estão sendo desenvolvidas aplicações de IA para o Direito que são utilizadas desde a confecção de um contrato à decisão judicial. Em alguns países, a IA já utilizada até mesmo em decisões criminais.

Por isso, este trabalho buscará analisar os impactos da inteligência artificial em matéria penal. Será necessário compreender o que é a inteligência artificial, sua história e conceitos, sua aplicação geral e ao Direito e a aplicação da IA no Judiciário brasileiro. Importante também compreender os vieses, desde o viés na decisão humana, os vieses em aplicações de IA até os vieses em IA aplicada ao Direito. Por último, será realizada uma análise dos impactos, possibilidades e regulação da IA em matéria penal.

Para atingir o objetivo acima descrito, em pesquisa exploratória e qualitativa, foi realizado levantamento bibliográfico, necessário para identificar o estágio e reflexões atuais sobre o tema. Em bases de dados, utilizou-se palavras-chave relativas ao tema para pesquisas de artigos científicos, relatórios e teses para, de forma indutiva, chegar a reflexões sobre o tema objeto da pesquisa².

2. COMPREENDENDO A INTELIGÊNCIA ARTIFICIAL

2.1. HISTÓRIA E CONCEITOS DA INTELIGÊNCIA ARTIFICIAL

A palavra algoritmo surgiu no século IX e denominava o processo de cálculo de numerais, e, portanto, em essência faz parte da vida humana, já que remete a ações que obedecem a padrões, como tricotar, por exemplo, como transformar algo em uma lei universal, um padrão exato, instruções precisas. (KAUFMAN, 2018, p. 26)

Existem diversas definições de IA, e em razão do uso do termo “inteligência” que pode dificultar a compreensão sobre o assunto, sendo necessário adotar um conceito mais generalista: “IA é a ciência e a engenharia de criar máquinas que tenham funções exercidas pelo cérebro dos animais”.

² Esta pesquisa foi iniciada em janeiro de 2022 e finalizada em abril de 2022. Porém, alguns artigos foram obtidos em pesquisa anterior, realizada no PIBIC 2020/2021, cujo tema era “Vieses nas decisões tomadas pela Inteligência Artificial aplicada ao Judiciário” e cujos resultados foram apresentados no 27º Congresso de Iniciação Científica da UnB e do 18º Congresso de Iniciação Científica do DF.

(KAUFMAN, 2021, p. 75)

Em 1950, Alan Turing criou um teste³ para identificar se a máquina conseguia se comportar como um humano, ou seja, quando a máquina estabelecia comunicação com o ser humano. O termo “inteligência artificial” surgiu em uma conferência na Faculdade de Darmouth, quando John McCarthy⁴ utilizou, em 1955, a expressão para diferenciar a lógica simbólica dos estudos de robótica. A classificação da IA hoje não apresenta tanta relevância e não tem mais o objetivo de imitação do ser humano, mas sim a resolução de problemas. Existe, portanto, a tendência de desvincular o conceito da definição da inteligência humana. (DIERLE; MARQUES, 2021, p.711-712)

A IA apresenta quatro características básicas: aplicação em todas as áreas, amplia e melhora áreas onde é aplicada, intensifica pesquisa e desenvolvimento e serve como base para inovações. O impacto que causa e ainda pode causar na melhora da economia e da qualidade de vida ultrapassa outras tecnologias (ARBIX, 2020, p. 2-3). Ela é multidisciplinar, e já é possível identificar três grandes ondas da IA: a simbólica, o *machine learning* e as ondas futuras (IA geral e IA superinteligente). A IA simbólica é utilizada para ações em que o sistema tem um objetivo, como por exemplo o chinês Go. A segunda onda, tem como expoente o *machine learning* em que a máquina aprende a partir de erros e acertos, identifica padrões e realiza generalizações sem uma programação como na IA simbólica. A terceira onda, a IA futura, é aquela que se referirá à super IA e à singularidade. (DIERLE; MARQUES, 2021, p. 714-718)

2.1.1. Machine Learning e Deep Learning Neural Networks

O termo *machine learning* foi utilizado em 1959 por um funcionário da IBM chamado Arthur Samuel para designar computadores que teriam a capacidade de aprender sem serem programados para isso. (KAUFMAN, 2018, p. 20)

Em 1986, Geoffrey Hinton apresentou em um artigo na revista Nature a ideia das *neural networks* (redes neurais) que fez nascer um subcampo do *machine*

³ Teste de Turing. Disponível em <https://sites.google.com/site/inteligenciaartificialist/o-que-e-a-inteligencia-artificial/teste-de-turing>.

⁴ <https://cs.stanford.edu/index.php?q=memoriam/professor-john-mccarthy>.

learning. Este subcampo surgiu de um incremento das possibilidades de funcionamento da IA em razão de avanços de processamento com a construção de processadores alinhados, criando uma espécie de rede neural artificial. (KAUFMAN, 2018, p. 20-21)

Enquanto um algoritmo processa um dado de entrada e apresenta um resultado na saída, no *machine learning* a máquina “recebe” os dados e um resultado almejado e os algoritmos aprendizes criam outro algoritmo para que seja possível atingir o resultado proposto (DOMINGOS, 2017, 25). Os algoritmos de *machine learning* aprendem de forma estatística passando do particular ao geral por meio de inferências, ou seja, a partir de aprendizagem estimativa. A programação humana seria inviável em atividades que demandam o tratamento de grande quantidade de dados. (KAUFMAN, 2018, p. 20)

Apesar de o cérebro humano ainda ser um mistério para a ciência, e a máquina ainda não ter condições de reproduzir seu funcionamento, é ele que inspira o funcionamento da máquina. A rede neural do cérebro humano é explicada pela configuração em que se formam as sinapses, quando ocorre a passagem de informações de um neurônio para outro. No computador, cada unidade é como um neurônio e cada camada processa a informação e a transmite para outra camada. O cérebro possui dez áreas, contendo cada uma 140 milhões de neurônios. O computador mais avançado até hoje, criado pela Microsoft, e que atua em reconhecimento de imagens possui 152 camadas⁵. Hoje, o homem determina o número de camadas, mas existem projeções de que a própria máquina, sem intervenção humana, possa construir outras máquinas inteligentes⁶, e é o valor da conexão e a arquitetura (número de camadas) que define a máquina como inteligente. (KAUFMAN, 2018, p. 22-23)

A aprendizagem profunda começou a apresentar resultados mais concretos a partir dos anos de 1990. Na aprendizagem supervisionada, a máquina é treinada a

⁵ “A nova rede é mais complexa do que as redes neurais atuais, que atualmente contam com 6 ou 7 “camadas”, podendo chegar até 30. A nova rede da Microsoft contém 152 camadas” - Redes neurais estão apenas no começo, diz Microsoft”. Disponível em <https://olhardigital.com.br/2016/01/14/noticias/redes-neurais-estao-apenas-no-comeco-diz-microsoft/>. Acesso em 12 de março de 2022.

⁶ Caso recente de invenção versa sobre discussões referentes à atribuição de patente à IA mostram que talvez esta possibilidade não esteja tão distante - **IA inventa produto, mas entidades não querem registrar a patente; entenda o caso**. Disponível em <https://canaltech.com.br/inteligencia-artificial/ia-inventa-produto-mas-entidades-nao-querem-registrar-a-patente-entenda-o-caso-192517/>. Acesso em 15 de março de 2022.

partir de exemplos e recebe ajustes até apresentar uma resposta adequada por meio de um método de tentativa e erro. Na aprendizagem não supervisionada, existem milhares de processadores e não há supervisão humana, ou seja, a própria máquina aprende e chega a representações abstratas a partir dos dados. (KAUFMAN, 2018, p. 21)

Em reconhecimento de imagens, por exemplo, as primeiras camadas identificam bordas ou cantos, as intermediárias analisam características básicas e as últimas camadas formam a interpretação completa. Por meio de padrões a máquina identifica um rosto o qual ela já possui dados, como uma pessoa que identifica alguém em um álbum de fotografias após olhar várias fotos. O reconhecimento de voz é a aplicação de IA mais bem sucedida até hoje. (KAUFMAN, 2018, p. 21-22)

2.2. APLICAÇÃO GERAL E AO DIREITO

2.2.1. Inteligência Artificial fraca e forte

A computação não conseguiria trabalhar a quantidade de dados que se tem hoje com regras de programação definidas, assim, o *machine learning* é mais vantajoso por gerar os algoritmos necessários a uma tarefa e se tornou a base da IA. São os dados que definem as operações realizadas pela máquina e não os programadores, e a máquina consegue realizar tarefas específicas para a resolução de problemas. (KAUFMAN, 2018, p. 18-19)

A IA se divide em fraca (ou estreita) e forte (ou geral), e é a IA fraca a que mais vemos aplicada em todas as áreas e no nosso dia a dia, a que mais se desenvolveu. Os algoritmos utilizam dados, organizados ou não, e aprendem ou são treinados para fazer algo, e por isso a técnica recebe o nome de *machine learning* (aprendizado de máquina). Já a IA forte seria aquela que se desenvolveria para se igualar à inteligência humana, tendo consciência, emoções e motivação, podendo até superar a inteligência humana e ameaçando a sua existência. Porém, a IA forte ainda está longe do que é almejado (a singularidade). (ARBIX, 2020, p. 7)

2.2.2. Funcionamento geral e aplicação do *Deep Learning*

Os algoritmos estão por toda a parte. “Se todos os algoritmos parassem de funcionar inesperadamente, o mundo que conhecemos chegaria ao fim” (DOMINGOS, 2017, p. 20). O *deep learning* (aprendizagem profunda) consegue avaliar um grande volume de dados retornando informações úteis sobre eles e está inserido em quase todas as atividades do dia a dia: acessar um computador, pesquisar no Google, comprar na Amazon, escolher filmes na Netflix etc. A máquina aprende a partir de inferências retiradas de dados, e melhores serão as inferências quanto mais dados forem fornecidos à máquina (KAUFMAN, 2018, p. 17-18).

O salto do *deep learning* a partir de 2010 ocorreu em razão do (1) big data, do (2) incremento da capacidade computacional e da (3) evolução dos algoritmos. Em 2008 o Google identificou dois meses antes do lançamento de informações oficiais sobre aumento de gripe no leste dos Estados Unidos o aumento de pesquisas sobre sintomas gripais. Em artigo publicado na *Nature*, pesquisadores da empresa alertaram sobre nova previsão de propagação de gripe, antes da detecção do H1N1, com base em pesquisas na plataforma e foram ignorados por autoridades. O *big data* seria a geração de novas informações, padrões, relações, a partir de dados, sem apego a amostras, mas com “a disposição em adotar a desordem do mundo real dos dados, deixando de privilegiar a exatidão (...). Trata-se de sacrificar a exatidão para ter acesso à tendência geral” (KAUFMAN, 2018, p. 24).

Os dados podem servir como entrada, treinamento ou feedback, servindo este último para melhorar o aprendizado da máquina a partir de reações de usuários. Quando alguém interage em meio digital, deixa, de forma involuntária ou voluntária, seus vestígios que podem ser utilizados pela plataforma original onde o dado foi gerado ou por terceiros. Esses dados podem ser utilizados para inúmeras finalidades, até mesmo, finalidades antiéticas e/ou ilícitas. Com a internet das coisas, na qual tudo apresenta informações digitais, obtém-se inúmeras informações sobre qualquer pessoa, suscetibilidade a doenças e até mesmo se a pessoa honra as próprias dívidas. Assim, o *big data* mostra os padrões do comportamento humano e oferece uma possibilidade de entender melhor a sociedade. (KAUFMAN, 2018, p. 24-25)

“O Homo sapiens é a espécie que adapta o mundo às suas necessidades e não uma espécie que se adapta ao mundo”. O mundo se modifica com a presença do ser humano e os algoritmos agora são utilizados para isso, avaliando o que as pessoas querem (DOMINGOS, 2017, p. 12). Os algoritmos no *deep learning* fazem previsões a partir de padrões e correlações, revelando algo sobre os dados que a análise

humana é limitada para perceber. Além disso, o algoritmo não vai funcionar da mesma forma, depende do resultado almejado, por isso os algoritmos da Netflix não conseguiriam entregar livros do interesse do consumidor como a Amazon, os algoritmos nesse exemplo, respondem a modelos de negócio diferentes. O algoritmo da Amazon te leva a livros semelhantes aos que você já comprou, os da Netflix te levam a entretenimento do seu interesse, porém, preferencialmente ao entretenimento que não seja filmes de sucesso para que o valor da assinatura retorne lucro à Amazon. A Amazon comparou a habilidade de vendas de humanos e da máquina, e a máquina superou muito os profissionais humanos, já que o aprendizado delas, a partir dos dados, possibilita aumento significativo de vendas. Nem a Amazon, nem a Netflix sabem explicar como os seus algoritmos funcionam. (KAUFMAN, 2018, p. 26-29), “em um círculo virtuoso (ou vicioso, se você for o rival)” (DOMINGOS, 2017, p. 33)

O Google e o Yahoo ganham receita de propagandas, porém, o valor de mercado do Google é maior pois seus algoritmos são melhores. O contato entre produtores e consumidores pode ocorrer em um clique. A quantidade de dados é incomensurável e as empresas deixam que os algoritmos aprendam e adivinhem as necessidades dos clientes, os algoritmos “unem produtores e consumidores”. Quem melhora seus algoritmos e aprende mais rápido sobre os clientes se torna vitorioso. (DOMINGOS, 2017, p. 30-31)

As mudanças provocadas pelo *deep learning* já exigem até mesmo a regulação da interação biológica homem-máquina, como ocorre no Chile com a tentativa de positivar os neurodireitos⁷ no texto constitucional, já que com os avanços da nanotecnologia possibilitam até mesmo implantes cerebrais para mapeamento, controle e correção de funções cerebrais. “Hoje, a tecnologia de leitura do cérebro, invasiva e não invasiva, é uma realidade”⁸.

2.2.3. Inteligência Artificial aplicada ao Direito

Os serviços jurídicos evoluíram de acesso por meio de dados físicos, digitais e

⁷ LAGE, F. C.; BRANCO, P. **Neurodireito: o direito fundamental do futuro**. Disponível em <https://www.jota.info/opiniao-e-analise/artigos/neurodireito-o-direito-fundamental-do-futuro-08042022>. Acesso em 20/04/2022.

⁸ DALESE, P. **Proteção jurídica de informações neurais: a última fronteira da privacidade**. Disponível em <https://www.jota.info/opiniao-e-analise/artigos/protecao-juridica-de-informacoes-neurais-a-ultima-fronteira-da-privacidade-13032021>. Acesso em 20/04/2022.

agora por Inteligência Artificial. Antes os materiais físicos eram limitados, logo, poucas pessoas possuíam acesso e a disseminação do conhecimento era restrita. A partir de 1976 surgiu a primeira forma de pesquisa jurídica por computador. Aos poucos o uso de computadores foi se disseminando e o conhecimento era propagado por meio de *CD-Room*. Com o advento da internet, a pesquisa jurídica tornou-se ainda mais rápida e eficiente do que ocorria com os materiais físicos, e a IA ofereceu um incremento ainda maior em celeridade e eficiência de atividades. (SOARES; KAUFFMAN; CHAO, 2020, p. 107-110)

O acesso a precedentes já aumenta a chance de sucesso em um processo judicial, e as facilidades oferecidas hoje pela IA ao advogado vão além disso. Na década de 1990, a IA superou o homem no jogo de xadrez e tem hoje derrotado os melhores jogadores de pôquer⁹. Essa capacidade preditiva da máquina já começa a ser aplicada a processos judiciais, pareceres são elaborados exclusivamente pela IA, entre outras atividades executadas pela máquina. (SOARES; KAUFFMAN; CHAO, 2020, p. 105-106)

A pesquisa jurídica computacional, realizada pela IA, apresenta velocidade superior a qualquer pesquisa feita pelo homem. No Brasil, já é possível que a máquina, sem intervenção humana, após distribuição do processo, elabore o despacho de citação do réu. (SOARES; KAUFFMAN; CHAO, 2020, p. 111)

A IA é utilizada na organização e pesquisa de documentos e plataformas digitais, eliminando despesas com pessoal, neste caso, eliminando a atuação que costuma ser atribuída a advogados iniciantes, estando terceirizada¹⁰ grande parte dessa atividade para a tecnologia. A computação forense é aquela que lida com análise de provas por meio de máquina. Os contratos agora podem ser elaborados pelos algoritmos com maior correção e com inclusão de cláusulas-chave, e estabelecendo alternativas de redação do contrato que evitem discussões judiciais posteriores. Entre outras possibilidades, a IA oferece maior previsibilidade e

⁹ Science. **Superhuman AI for multiplayer poker**. Disponível em <https://www.science.org/doi/10.1126/science.aay2400>. Acesso em 1 de abril de 2022.

¹⁰ “De acordo com Vitalino Neto, a taxa de preenchimento correto dos dados de um processo no sistema da companhia é de 75% quando o trabalho é feito por humanos e de 95% com o uso da tecnologia do IBM Watson”. Exame - Inteligência artificial da IBM já ajuda advogados brasileiros. Disponível em <https://exame.com/tecnologia/inteligencia-artificial-da-ibm-ja-ajuda-advogados-brasileiros/>. Acesso em 19 de março de 2022.

segurança a serviços jurídicos. (SOARES; KAUFFMAN; CHAO, 2020, p. 118-119)

A IA é aplicada ao Direito em movimento crescente. Em escritórios de advocacia, a IA já é aplicada em ações de gestão e organização de forma superior a um advogado. Em alguns países, a IA já é utilizada em cerca de 90% dos escritórios de advocacia. Em alguns países como China e Estônia, decisões já são submetidas a juízes-robôs, como em causas de pequeno valor e litigância on-line. (DIERLE; MARQUES, 2021, p. 705-706). No Brasil, o destaque no setor público é a Advocacia Geral da União (AGU) que utiliza a IA no Sistema SAPIENS desde 2014. (DIERLE; MARQUES, 2018, p. 2)

A previsão é que a IA impacte mais fortemente áreas em que as tarefas possuem menor valor agregado e poderiam ser automatizadas como (i) pesquisa, localização e seleção de documentos, (ii) pesquisa jurídica, (iii) geração de documentos (iv) produção de documentos e (v) previsão de resultados de julgamentos. As tarefas dos juristas passarão a ser aquelas de maior valor agregado, sendo necessária adaptação e aprendizagem constante dos novos profissionais, exigindo conhecimentos interdisciplinares. Os advogados tradicionais serão mais escassos, sendo necessárias habilidades cognitivas de alto valor. Porém, a questão central do uso da IA é a eficiência que não deveria ser a base de toda essa mudança já que o valor perseguido pelo Direito é a Justiça, que pode acabar sufocada por interesses diversos. (DIERLE; MARQUES, 2021, p. 708-710)

2.3. APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL NO JUDICIÁRIO BRASILEIRO

A IA pode impactar o Judiciário, reduzindo o tempo de processos. O Judiciário tem lidado com um volume cada vez maior de processos conforme o relatório Justiça em Números do Conselho Nacional de Justiça (CNJ), gerando gastos que impactam o PIB do país. (MAIA FILHO; JUNQUILHO, 2018, p. 220-221)

A IA apresenta-se em larga expansão modificando diversas áreas, e surge no Direito como possibilidade de racionalização das atividades, principalmente no Judiciário brasileiro. A análise aprofundada de processos e o reconhecimento de

padrões e conexões pela IA pode promover uma compreensão mais profunda da prática jurídica (MAIA FILHO; JUNQUILHO, 2018, p. 222-223).

Como exemplos, o Tribunal de Justiça de Minas Gerais (TJMG) utiliza a IA Radar que analisa os processos e agrupa-os em causas similares e exemplificou o julgamento de 280 processos simultâneos com temas idênticos pela máquina (DIERLE; MARQUES, 2021, p. 706-707). No Tribunal de Justiça de Pernambuco (TJPE), o trabalho de triagem de processos e citações de execução fiscal que seriam feitas por servidores em um prazo de 18 meses, foi realizado pela IA em 15 dias. (SOARES; KAUFFMAN; CHAO, 2020, p. 111)

Segundo o “Relatório de Pesquisa Tecnologia aplicada à gestão de conflitos no âmbito do Poder Judiciário Brasileiro”, elaborado pela Fundação Getúlio Vargas sob coordenação do Ministro Luiz Felipe Salomão, do Superior Tribunal de Justiça (STJ), o Poder Judiciário contava em 2020 com cerca de cinquenta projetos de inteligência artificial (a maior parte iniciado no referido ano).

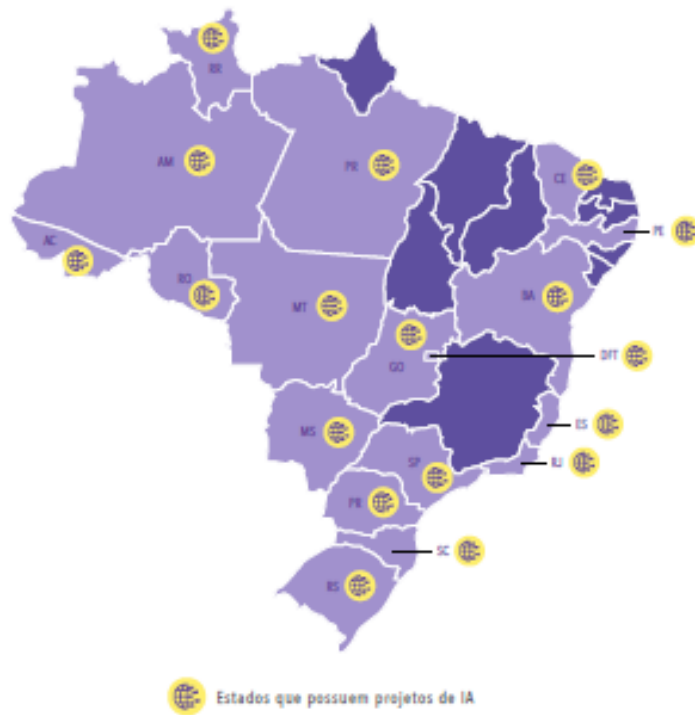
Tabela 1: Projetos de Inteligência Artificial do Poder Judiciário Brasileiro

PROJETOS DE INTELIGÊNCIA ARTIFICIAL DO PODER JUDICIÁRIO BRASILEIRO					
	IA	Órgão		IA	Órgão
1	VICTOR	STF	28	HÓRUS	TJDFT
2	ATHOS	STJ	29	ÁMON	TJDFT
3	SOCRATES	STJ	30	PROJETO 1	TJES
4	E-JURIS	STJ	31	PROJETO 2	TJES
5	TUA	STJ	32	PROJETO 3	TJES
6	BEM-TE-VI	TST	33	IA332	TJGO
7	SINAPSES	CNJ	34	LEIA	TJMS
8	ROBÔ SECOR	TRF1	35	PROJETO 1	TJMT
9	BANCO DE SENTENÇA	TRF1	36	PROJETO 1	TJPR
10	SIB	TRF1	37	ELIS	TJPE
11	ALEI	TRF1	38	PROJETO 1	TJRJ
12	PROJETO EXECUÇÃO CELERE	TRF1	39	SINAPSE	TJRO
13	ATENDENTE VIRTUAL	TRF2	40	SCRIBA	TJRR
14	SINARA	TRF3	41	MANDAMUS	TJRR
15	SIGMA	TRF3	42	PROJETO 1	TJRS
16	PREVENÇÃO	TRF3	43	PROJETO 1	TJSC
17	CLASSIFICAÇÃO DE TEMAS NA VICE-PRESIDÊNCIA E TURMAS RECURSAIS	TRF4	44	JUDI	TJSP
18	ANÁLISE DE ASSUNTO DOS PROCESSOS	TRF4	45	LEIA	TJSP
19	TRIAGEM AUTOMÁTICA DE PROCESSOS A PARTIR DA PETIÇÃO INICIAL	TRF4	46	MINERJUS	TJTO

20	SUGESTÃO DE MODELOS DE MINUTAS	TRF4	47	INTELIGÊNCIA ARTIFICIAL E EFICIÊNCIA DO JUDICIÁRIO	TRT1
21	JULIA	TRF5	48	CLUSTERIZAÇÃO DE PROCESSOS	TRT4
22	LEIA	TJAC	49	GEMINI	TRT5
23	LEIA	TJAL	50	GEMINI	TRT7
24	HÉRCULES	TJAL	51	B.I. TRT 11	TRT11
25	LEIA	TJAM	52	CONCILIA JT	TRT12
26	QUEIXA CIDADÃ	TJBA	53	GEMINI	TRT15
27	LEIA	TJCE	54	GEMINI	TRT20

Fonte: Elaborado pela autora a partir do “Relatório de Pesquisa Tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário Brasileiro”¹¹.

Figura 1: Projetos de IA do Judiciário Brasileiro



Fonte: “Relatório de Pesquisa Tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário Brasileiro”

Um dos projetos de IA de destaque é o Victor, que utiliza aprendizado de

¹¹ Relatório de Pesquisa Tecnologia aplicada à gestão de conflitos no âmbito do Poder Judiciário Brasileiro (FGV). Disponível em <https://ciapj.fgv.br/publicacoes>. Acesso em 18 de março de 2022.

máquina para que seja realizada triagem de processos em sede de controle difuso que chegam ao Supremo Tribunal Federal (STF), principalmente, para identificação de temas de repercussão geral. (MAIA FILHO; JUNQUILHO, 2018, p. 222)

A repercussão geral foi criada para que, identificados temas idênticos, pudesse ocorrer o estabelecimento de uma decisão com efeito multiplicador, em causas de relevância social, política, econômica ou jurídica que ultrapassem os interesses das partes, reduzindo a sobrecarga de processos e uniformizando a interpretação constitucional. É nesse contexto de mais de uma centena de processos anuais que o modelo de aprendizagem de máquina auxiliará o trabalho do Tribunal, em um primeiro momento desenvolvidos os algoritmos e treinados, em outro momento passa-se à fase de atividade em tempo real na triagem e classificação de processos. (MAIA FILHO; JUNQUILHO, 2018, p. 226-227)

O projeto Victor desenvolvido no STF em conjunto com a Universidade de Brasília, apresenta o envolvimento de três áreas do conhecimento: engenharia de software, ciência da computação e direito. O projeto não tem a participação da equipe de TI da Corte. A IA deverá analisar os recursos que chegam ao STF identificando temas de repercussão geral, sem tomada de decisões pela IA, mas auxiliando o tribunal na gestão dos processos. (MAIA FILHO; JUNQUILHO, 2018, p. 225-226)

A partir dessa atividade a IA auxilia a formar um banco de dados que pode ser utilizado para melhor compreensão sobre quem mais litiga, quais são os temas de maior frequência e de maior volume que chegam ao Tribunal e identificar casos excepcionais. O projeto Victor, por solicitação da equipe do STF, iniciou as atividades a partir da identificação de peças importantes. Repassados os dados, a IA teve elevada assertividade na identificação e classificação das peças. Outro aspecto é a análise pela IA de como é construída a repercussão geral, o que, reitera-se trará maiores informações sobre a atuação da Corte. O projeto Victor é inovador e inédito, visto que nenhum outro país possui projeto de IA em Corte Constitucional (MAIA FILHO; JUNQUILHO, 2018, p. 228-229)

Assim, o projeto Victor demonstrou as possibilidades de avanço da IA aplicada ao Judiciário como redução de tempo da tramitação processual, desenvolvimento de tecnologia brasileira, eficácia em pesquisa processual e jurisprudencial e isonomia das decisões em razão da identificação de casos semelhantes. A vanguarda do STF gerou

interesse de outros tribunais que também iniciaram projetos de IA. (MAIA FILHO; JUNQUILHO, 2018, p. 230)

No Judiciário, como exposto, é ainda incipiente a aplicação da IA. O fator humano não seria excluído no processamento das informações pela máquina já que uma pessoa precisa avaliar e determinar parâmetros para os algoritmos, sendo o ser humano o gestor da informação. Além disso, a IA precisa de controle e regulação além de um desenvolvimento baseado em diretrizes éticas, com proteção de direitos constitucionais. Não só a IA como a própria internet precisam de aspectos como regulação, hierarquia, organização, controle entre outros, já que tudo isso é base para o funcionamento da máquina. Seria necessário refletir sobre a qualidade do controle e a normatização da IA. (MAIA FILHO; JUNQUILHO, 2018, p. 224-225)

Existem, porém, dificuldades na aplicação da IA como vieses na programação, ausência de regulamentação na contratação de empresas e cuidados no tratamento de dados (MAIA FILHO; JUNQUILHO, 2018, p. 231), além de ausência de regulação do desenvolvimento da própria IA.

3. COMPREENDENDO OS VIESES

3.1. O VIÉS NAS DECISÕES HUMANAS

Desde 1970, a psicologia cognitiva estuda os limites da racionalidade humana, inerentes a todas as pessoas, que ocorre mesmo com as pessoas mais bem formadas, e que dificultam uma análise neutra nos julgamentos. Quando é necessário decidir, lança-se mão de heurísticas que são atalhos cognitivos que o cérebro utiliza para lidar com lacunas, seja por limitação de tempo ou de informações. Não raciocinamos de forma refletida na maior parte das vezes, e como “nossa capacidade de processamento é limitada” a heurística torna possível uma decisão dentro das limitações existentes e sem utilizar muito tempo para isso. Ou seja, toma-se uma decisão aceitável e não ótima, já que a capacidade de percepção dos fatos é limitada (HORTA; COSTA, 2017, p 17-18).

A heurística não pode ser considerada um erro, mas uma estruturação do

pensamento que auxilia de forma efetiva a tomada de decisões em alguns contextos, como em momentos em que a decisão precisa ser tomada sob estresse emocional, em curto prazo de tempo, com poucas informações ou avaliações incompletas. Esse modo de funcionamento do cérebro provoca os chamados vieses cognitivos que se expressam de forma inconsciente. Os vieses ocorrem independente de características pessoais como honestidade e boa intenção, e mesmo quem tem conhecimento e é treinado para reduzir a influência das heurísticas inconscientes acaba raciocinando com base em vieses (HORTA e COSTA, 2017, p. 19).

O que estudos sugerem é que uma das primeiras coisas a serem feitas é desconfiar que o que sabemos pode realmente descrever algum objeto de forma verdadeira. As pessoas não compreendem como funcionam seus processos cognitivos e muitas vezes não explicamos algo de forma precisa, criando na verdade racionalizações que façam com que o objeto se encaixe ao que entendemos “ser verdadeiro” e não ao que o objeto é verdadeiramente. Em experimentos, foi demonstrado que as pessoas se baseiam mais na opinião de outras pessoas ao decidir, com base em percepções morais e emocionais, do que em critérios racionais e lógicos (HORTA, 2019, p. 34).

Comprovou-se, inclusive, que o que motiva uma decisão costuma ser um critério intuitivo e que ao se questionar a razão da decisão, o decisor encontra dificuldades de justificação por desconhecimento do próprio processo decisório. Em alguns casos, as pessoas chegaram ao ponto de racionalizar uma escolha, quando na verdade os pesquisadores trocaram o objeto escolhido, e as pessoas nem ao menos perceberam (HORTA, 2019, p. 35).

Os processos mentais são elaborados em grande parte do tempo sob influência de elementos inconscientes e as pessoas normalmente, quando explicam o seu comportamento, utilizam racionalizações, não sabem os elementos envolvidos na decisão. (HORTA, 2019, p. 36) Por isso é tão complexo chegar a acordos. Se o ser humano fosse realmente racional, o melhor argumento prevaleceria (HORTA, 2019, p. 37).

Mesmo com tanta capacidade, a razão está sujeita a raciocínios incorretos, notícias falsas e outros erros cognitivos. Todos os animais são capazes de traçar algum tipo de inferência, e algumas delas são incorporadas de forma inconsciente e executa-se ações de forma automática (p. 38), como um hábito. Essas decisões

automáticas geram muitas decisões corretas, mas também leva a vieses de raciocínio, o que leva a observação e identificação de erros para que possam ser corrigidos (HORTA, 2019, p. 38).

3.1.1. Vieses em decisões judiciais

As pesquisas das ciências comportamentais provocam dúvidas se é possível encontrar uma resposta correta ou se é possível estabelecer racionalidade em decisões judiciais (HORTA, 2019, p. 37).

Reputa-se pela imparcialidade da decisão judicial, porém áreas como a psicologia e a economia levantam um debate sobre a questão demonstrando a presença de vieses em decisões que são influenciadas por questões de gênero, raça, posições político-ideológicas etc. Até mesmo a ideia de papel contra majoritário de Cortes Constitucionais pode sofrer abalos já que pesquisas demonstram, por exemplo, que a Suprema Corte dos EUA decide de acordo com a opinião da maioria. Ou seja, os dados demonstram que a imparcialidade é um ideal distante da realidade e que as decisões sofrem influências de variados tipos, visto que, esta influência de fatores, até mesmo irrelevantes nas decisões, está presente em qualquer decisão humana como demonstrado por estudos (HORTA, 2019, p. 24).

Dessa forma, o conhecimento dos vieses auxiliaria a rever as soluções clássicas da prática decisória e resolver os pontos cegos. Existe uma tendência de decisões serem tomadas por opiniões e não por dados e essa falta de postura reflexiva tem causado inclusive uma maior manipulação da opinião pública. A expectativa de que educação, informação e sistema político representativo blindariam a sociedade de fenômenos como a pós verdade acabou mostrando-se infrutífera. (HORTA; COSTA, 2017, p. 16)

Uma compreensão clássica, presente também no Direito, entende que a justiça seria alcançada com a formação de juízes tecnicamente bem formados, informados por imprensa livre e comprometidos com a concretização de direitos. Porém, o modelo das instituições e a formação dos juízes podem não atingir o fim a que se destinam e é necessário compreender a forma cognitiva adequada para julgar (HORTA; COSTA, 2017, p. 17).

A realidade é influenciada e moldada por percepções. Para correção, é necessário esforço e controle deliberado do raciocínio, ou seja, é necessário pensar além dos hábitos e mecanismos mentais inconscientes. Para tanto, duas estratégias seriam necessárias: o estímulo individual a uma atitude reflexiva e uma estrutura institucional que consiga limitar os vieses. Ou seja, é necessário o conhecimento do viés e como ele funciona em determinadas áreas para que seja minimizado (HORTA; COSTA, 2017, p. 20).

Por exemplo, o magistrado normalmente fica submetido ao viés de confirmação, que é aquele que o faz priorizar evidências que confirmem as ideias pré-concebidas. Outras influências são as crenças e preconceitos, estados físicos como cansaço excessivo, a tendência cognitiva de levar em consideração provas ilícitas¹² desentranhadas do processo, entre outros fatores que mostram que o juiz não consegue ser totalmente imparcial (HORTA; COSTA, 2017, p. 14).

Outra dificuldade é o viés de confirmação forense que se propaga (efeito bola de neve) entre todos os sujeitos envolvidos no processo (juiz, MP, delegado) que tendem a confirmar a presunção de culpa dos investigadores. Além disso, a falta de capacitação e de procedimentos adequados impedem uma boa condução do inquérito no Brasil, como o uso de perguntas fechadas em interrogatórios e o reconhecimento de suspeito sozinho (HORTA; COSTA, 2017, p.14).

3.1.2. Vieses no sistema prisional brasileiro

O Brasil possui a quarta população carcerária do mundo e mesmo com o foco em penas privativas de liberdade, não ocorreu redução efetiva da criminalidade (HORTA; COSTA, 2017, p. 23). Em maio de 2021¹³, o sistema prisional brasileiro

¹² “Agora, não haverá mesmo dúvida sobre a imparcialidade do julgador nesse contexto? Será possível eliminar os efeitos dessa prova ilícita na psique do juiz ou sequer dimensioná-los, em termos de influência sobre as demais cognições e decisões? E é razoável tolerar essa crença ingênua de que o julgador não se utilizará — consciente ou inconscientemente — dos conhecimentos ilicitamente obtidos, após excluídos do processo, durante a marcha processual subsequente e na sentença?”. LOPES, A. J.; RITTER, R. **Não basta desentranhar a prova; deve-se "desentranhar" o juiz**. Disponível em <https://www.conjur.com.br/2020-nov-06/limite-penal-nao-basta-desentranhar-prova-desentranhar-juiz>. Acesso em 17 de abril de 2022.

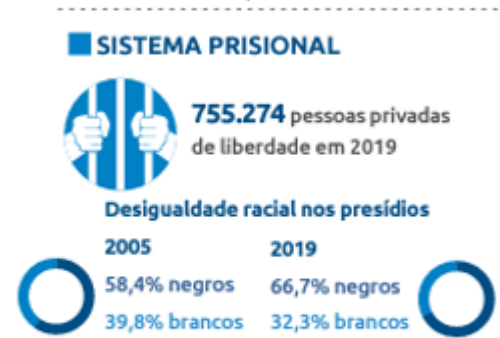
¹³ O Infopen (Levantamento Nacional de Informações Penitenciárias) de 2019 não estava disponível no momento que esta pesquisa foi realizada. Os dados sobre a população carcerária foram coletados no Monitor da violência do site G1. Disponível em <https://g1.globo.com/monitor-da>

estava com lotação cinquenta e seis por cento acima da capacidade. O estado do Amazonas é o que possui a mais alta superlotação do sistema de quase duzentos por cento, são três mil seiscentas e dez vagas que comportam dez mil seiscentas e noventa e duas pessoas. O número de pessoas presas sem julgamento passava de cerca de trinta por cento da população carcerária.

Já o mais recente Anuário Brasileiro de Segurança Pública de 2021¹⁴ informa que o número de mortes decorrentes de crimes violentos no Brasil aumentou quatro por cento, totalizando mais de cinquenta mil vítimas, sendo mais de sessenta e seis por cento de pessoas negras, mais de cinquenta e quatro por cento de jovens e mais de noventa e um por cento de homens.

Cresceu também o número de mortes de crianças e adolescentes, de feminicídios, em decorrência de ações policiais, de policiais em serviço, entre outros indicadores. A única redução percebida foi a de todos os crimes contra o patrimônio. O 14º Anuário, informou que de 2005 a 2019, houve um crescimento da proporção de negros no sistema prisional: de cada três presos, dois eram negros.

Figura 2: Distribuição de presos por raça no sistema prisional



Fonte: 14º Anuário Brasileiro de Segurança Pública

Rebeliões como as que ocorreram recentemente em alguns estados¹⁵, demonstram que o sistema prisional brasileiro entrou em colapso, pois há um superencarceramento. Além de problemas de políticas públicas, talvez a seleção de magistrados no país possa ser uma das causas desse colapso, não há muitas vezes uma boa instrução probatória e tempo suficiente para julgamento, o que gera um número elevado de prisões que muitas vezes não passam por um julgamento definitivo. Os juízes julgam de forma intuitiva e distorcida cognitivamente, já que muitas vezes as decisões são enviesadas (COSTA; HORTA, 2017, p. 13).

violencia/noticia/2021/05/17/populacao-carceraria-diminui-mas-brasil-ainda-registra-superlotacao-nos-presidios-em-meio-a-pandemia.ghtml. Acesso em 28 de março de 2021.

¹⁴ Disponível em <https://forumseguranca.org.br/anuario-brasileiro-seguranca-publica/>. Acesso em 28 de março de 2022.

¹⁵ BBC. **Rebelião em Manaus: a disputa interna de facção criminosa que levou ao massacre em presídios**. Disponível em <https://www.bbc.com/portuguese/brasil-48428432>. Acesso em 28 de março de 2022.

Um dos vieses comuns na sociedade é o viés retributivista da punição, em que a punição a ser atribuída é proporcional à gravidade do dano causado, contrário ao discurso moderno de que a pena imposta a um agente teria um caráter pedagógico ou que busca inibir repetição da mesma conduta delitiva (HORTA; COSTA, 2017, p. 20).

Em estudos, buscou-se compreender os processos mentais com relação a determinação de penas. Os entrevistados assinalaram, em abstrato, que as penas funcionavam de uma forma preventiva, porém, ao analisarem casos concretos, as penas indicadas eram proporcionais ao dano causado e não estavam ligadas a ideia de prevenção (HORTA e COSTA, 2017, p. 21).

“Ora, segundo a teoria consequencialista da prevenção, se o objetivo da punição é aumentar o custo do cometimento de crimes, para preveni-los, a punição precisa servir de exemplo dissuasivo, sendo tão mais dura quanto maior a potencialidade de reincidência, quanto mais difícil de se descobrir um crime, ou mais amplo o conhecimento do seu cometimento. Se nenhuma dessas variáveis é determinante, há então um descompasso entre a ‘preferência’ preventiva do discurso e a prática retributivista da punição”.

Ou seja, nenhuma das variáveis da teoria consequencialista estavam presentes no momento de atribuição de penas, o que mostra um desalinhamento entre discurso e prática, demonstrando que “*a punição está ligada à vingança, não à prevenção*” (HORTA; COSTA, 2017, p. 21).

Dessa forma, “*a decisão de punir é mais complexa, mais emotiva e menos racional*” (p. 21) e as pessoas tendem a atribuir penas de acordo com as emoções e depois estabelecem argumentos que racionalizem esse processo. Esta descrição pode ser visualizada na prática por legislações penais aprovadas em momentos de clamor social para agravar condutas delitivas, e a justificativa comum é evitar o cometimento de crimes semelhantes, o que tem conduzido ao alargamento das penas privativas de liberdade. Mesmo com os dados demonstrando os efeitos criminógenos¹⁶ do encarceramento, os magistrados muitas vezes sentem como injusta a manutenção do réu em liberdade, o foco da punição é a vingança e entende-se a pena privativa de

¹⁶ “A literatura especializada é rica em exemplos dos efeitos criminógenos da prisão. Enfim, a maioria dos fatores que domina a vida carcerária imprime a esta um caráter criminógeno, de sorte que, em qualquer prisão clássica, as condições materiais e humanas podem exercer efeitos nefastos na personalidade dos reclusos”. BITENCOURT, Cezar Roberto. **Análise político criminal da reincidência e o falacioso objetivo ressocializador**. Disponível em <https://www.conjur.com.br/2019-jun-13/bitencourt-reincidencia-falacioso-objetivo-ressocializador>. Acesso em 29 de março de 2022.

liberdade como justa por ser mais gravosa, não estando as práticas pautadas na reparação do dano ou pacificação social. Ampliar a criminalização, enrijecer penas e estimular o encarceramento são bases de políticas públicas que já se mostraram ineficazes, mas mesmo assim a pedagogia e a prevenção continuam a ter papéis marginais (HORTA; COSTA, 2017, p. 22).

Mostra-se necessário reduzir a incidência dos vieses (desenviesamento) por meio de estratégias, já que as pessoas não percebem a influência de fatores irrelevantes em suas decisões. Uma dessas estratégias é sempre adotar a postura de que o que se sabe pode estar incorreto, e pensar na possibilidade oposta. No processo judicial, o contraditório concretiza essa solução. Outra solução para o desenviesamento é a *accountability* já que a prestação de contas e observância da correção da resposta ajuda a reduzir vieses. A análise e decisões em grupo também ajudam a reduzir vieses, o que na lógica processual refere-se ao duplo grau de jurisdição, quando uma sentença pode ser revista por um órgão colegiado. A diversidade de pessoas que analisam uma decisão também reduziria os vieses, porém, esse pode ser um problema da magistratura brasileira, visto que o que se percebe é homogeneidade dos atores. As regras de suspeição e impedimento são eficiente forma de impedir que vieses contaminem decisões. É essencial reduzir paixões como as provocadas por exposições públicas que, apesar de importantes para a publicidade, podem provocar a influência da vaidade nas decisões, como no caso de julgamentos transmitidos pela TV Justiça (HORTA; COSTA, 2017, p. 25-26).

Entre outras soluções institucionais para redução de vieses sugere-se separação do juiz da instrução, o juiz da tutela sumária e o da sentença, impedimento de juiz que tomou contato com prova ilícita, entre outras. Outra proposta é maior rigor com relação ao ambiente do júri e que este seja até mesmo virtual para mitigar os vieses que provocados pela aparência de testemunhas em julgamentos, por exemplo. São necessárias mais pesquisas empíricas para descobrir a extensão dos vieses no Brasil (HORTA; COSTA, 2017, p. 27-28).

Existe, portanto, a necessidade de reflexão sobre os sistemas prisionais já que são permeados pelos vieses e que sejam criadas estratégias para reduzir a interferência das distorções de pensamento nos processos judiciais, deslocando os debates da verdade para os métodos. A aplicação do conhecimento científico é necessária não por ser verdadeiro, nem neutro, mas por apresentar estratégias que

reduzam os vieses. A importância de superar o senso comum e fazer uma análise reflexiva é essencial para a redução dos vieses. É necessário criar consciência de que os aparatos hermenêuticos não têm sido suficientes e que é necessário estudar o comportamento humano e nossas limitações cognitivas e que isso seja levado em consideração no desenho das instituições com vistas a mitigar vieses (HORTA; COSTA, 2017, p. 29-30).

3.2. A PRESENÇA DOS VIESES EM APLICAÇÕES DE INTELIGÊNCIA ARTIFICIAL

A IA recebe uma programação para aprender, assim como um ser humano, sendo apresentadas instruções de como entender o mundo a partir de exemplos, e retira desse aprendizado padrões que podem solucionar um problema. Ou seja, o algoritmo identifica entradas e saídas e aprende qual a melhor solução buscando uma generalização. O mesmo algoritmo pode gerar modelos diferentes por tratar diferentes bases de dados, e quanto melhor a base de dados, melhor será o modelo. A IA reproduz o conhecimento presente nos dados, logo, se neles estiver presente algum tipo de preconceito, a máquina aprenderá e disseminará os mesmos preconceitos. Dessa forma, a IA não poderia ser aplicada indiscriminadamente sem discussões sobre o seu uso (GARCIA, 2020, p. 15-16).

A IA toma decisões a partir de dados que ao não serem corretamente tratados e desagregados intensifica preconceitos e, portanto, cresce o número de resultados enviesados. Diagnósticos na Inglaterra são normalmente aplicados a mulheres de forma incorreta pois a IA não aprendeu a diferenciar que homens têm mais chances de infartos. O viés de gênero é um exemplo dentre todos os vieses. Ao analisar banco de dados desagregado por sexo, constatou-se em pesquisas recentes que a maior parte das mortes provocadas por Covid-19 ocorreu entre homens, o que poderia auxiliar na compreensão e tratamento da doença. Porém, a maior parte dos dados não são publicizados nem desagregados o que prejudica pesquisas. Há estudos em andamento que correlacionam proteção do estrogênio, hormônio feminino, ao vírus (KAUFMAN, 2021, p. 74).

Um dos problemas apresentados pela IA é a não neutralidade dos dados. Em

2016, no primeiro concurso de beleza do mundo com jurados robôs, o Beauty.AI buscou encontrar os rostos mais belos, a partir de critérios objetivos, com IA treinada para avaliação de simetria facial, rugas, uniformidade de pele etc., sem preconceitos raciais ou socioculturais. Porém, o resultado do concurso mostrou que, mesmo com os esforços de treinar os algoritmos de forma “neutra”, de 6 mil pessoas de várias partes do mundo foram escolhidos 44 vencedores, e dentre estes, apenas uma pessoa negra¹⁷. Apesar de não ter sido construída para compreender como belas apenas pessoas de pele branca, o treinamento da máquina foi realizado com imagens de atrizes de Hollywood (GARCIA, 2020, p. 14-15) e a equipe de desenvolvedores também era branca, o que também pode ter contribuído para o resultado enviesado. Tay é outro exemplo de IA enviesada. Lançada pela Microsoft em 2016, Tay, treinada para interagir como adolescente e com comentários divertidos, publicou 95 mil tweets em 16 horas, muitos deles ofensivos e preconceituosos (GARCIA, 2020, p. 16-18). Em vinte e quatro horas, a conta de Tay foi retirada do ar pela Microsoft¹⁸.

Outro problema é a validade dos dados. Em 2018, a Amazon tentou fazer uma seleção de candidatos por meio de IA. A máquina fazia a análise de currículos sendo treinada com base de dados de funcionários antigos. O algoritmo segundo a Amazon foi treinado para não ser preconceituoso, porém, a base de dados é antiga e não possuía mulheres, o que fez com que a máquina selecionasse apenas homens entre os candidatos, excluindo mulheres e homens que tivessem se formado em locais que remetesse a nomes femininos (GARCIA, 2020, p. 19).

Além da não neutralidade e da validade dos dados, eles podem camuflar vieses. Em sistema de IA treinado para uma empresa de saúde, observou-se que os pacientes mais custosos eram os doentes crônicos e que estes utilizavam mais vezes o sistema de saúde. Assim, a IA deveria identificar os segurados com maior propensão a doenças crônicas para passarem por tratamentos preventivos e reduzir dessa forma o custo com emergência e CTI. O sistema foi desenvolvido com o objetivo de não retornar resultados com viés racial, assim informações que pudessem descrever a raça dos pacientes foram retiradas da base. Apesar dos esforços, durante o período

¹⁷ “Out of 44 winners, nearly all were white, a handful were Asian, and **only one had dark skin**” (grifo meu). Disponível em <https://www.theguardian.com/technology/2016/sep/08/artificial-intelligence-beauty-contest-doesnt-like-black-people>. Acesso em 8 de março de 2022.

¹⁸ MOREIRA, Isabela. **A Microsoft criou um robô que interage nas redes sociais – e ela virou nazista**. Disponível em <https://revistagalileu.globo.com/blogs/buzz/noticia/2016/03/microsoft-criou-uma-robo-que-interage-nas-redes-sociais-e-ela-virou-nazista.html>. Acessado em 10 de junho de 2020.

em que o sistema foi utilizado, a maior parte dos segurados tratados era de pessoas brancas. Os negros utilizavam menos os sistemas em razão do custo e por receio de perder seus empregos se faltassem ao trabalho para consultas. Assim, a IA aprendeu apenas um padrão que identificasse brancos para tratamento, que eram maioria no banco de dados (GARCIA, 2020, p. 20-21).

Matemáticos estão estudando todos os passos das pessoas. O *big data* parecia imparcial e objetivo. Porém, existem erros e problemas e alguns modelos programam vieses e preconceitos (O'NEIL, 2020, p. 4-6). No beisebol esses modelos são aplicados há muito tempo. O passado é analisado e calcula-se a probabilidade de ganhar o jogo a partir dos dados, traçando-se estratégias. Vários cientistas de dados atuam no esporte, substituindo o instinto de treinadores por modelos estatísticos. Esses modelos são justos, todos sabem que eles estão sendo utilizados e é possível ter acesso aos modelos e saber como são interpretados, os dados utilizados estão sempre disponíveis (O'NEIL, 2020, p. 17-20).

O problema é que nem todos os dados estão à disposição na maior parte dos modelos e os programadores precisam utilizar proxies, dados substitutos, indicadores aproximados, para programar. Além disso, algumas correlações estatísticas utilizam dados de forma discriminatória, mantendo as pessoas em “baldes” de forma ilegal. Esses modelos destrutivos não possuem um ciclo de feedback constante para que a máquina aprenda com erros, assim como é feito nas análises de beisebol (O'NEIL, 2020, p. 20-21).

O modelo é uma abstração da realidade no qual utilizam-se informações passadas para prever respostas futuras. Todas as pessoas aplicam modelos semelhantes. Em família, é possível, a partir da convivência, prever o que cada integrante prefere para refeições ou o que não se tem interesse e a partir daí montar uma refeição em família que atenda a todos, por exemplo. Porém, ao tentar transformar esse modelo em algo que possa ser registrado e formalizado para ser utilizado por outras pessoas, algum item sempre ficará de fora. Mas quanto maior o tempo em que esses dados são registrados o modelo ficará melhor, mas nunca perfeito, pois trata-se de simplificação da realidade, que é muito complexa (O'NEIL, 2020, p. 21-23).

Esses dados que ficam de fora do modelo são chamados de pontos cegos e

“refletem o julgamento e prioridades de seus criadores”. Podem não fazer diferença se te levarem para uma rua errada no Google Maps, mas terão consequências completamente diferentes se errarem no cálculo de rota de um avião. Da mesma forma, os modelos apresentam demasiada simplificação para avaliar pessoas, buscam eficiência ao sacrifício da precisão. Reputam imparcialidade, mas na verdade refletem objetivos, valores e ideologias. “Modelos são opiniões embutidas em matemática”. O modelo funcionar ou não também acaba sendo uma questão de opinião (O’NEIL, 2020, p. 23-24).

Se crianças decidirem o melhor modelo para refeições em família, haverá sorvete todos os dias. Além disso, modelos ficam obsoletos, já que a vida é dinâmica e tudo se altera, assim, o modelo das crianças não terá sucesso para adolescentes. Alguns modelos dependem de uma única variável, como um detector de fumaça. Mas seres humanos não podem ser modelados de forma tão simples quanto um alarme de incêndio. O racismo está presente em vários modelos mentais, construídos a partir de dados incompletos, defeituosos ou generalizados. Racistas não buscam dados confiáveis ou questionam seus modelos deturpados, na verdade tendem a confirmar e fortalecer esses modelos. “Ele é alimentado por coleta irregular de dados e correlações espúrias, reforçado por injustiças institucionais e contaminado por viés de confirmação”. Muitos dos modelos matemáticos utilizados nas máquinas funcionam de forma semelhante ao racismo (O’NEIL, 2020, p. 25-26).

3.3. VIESES NA INTELIGÊNCIA ARTIFICIAL APLICADA AO DIREITO: O SISTEMA COMPAS

Nos Estados Unidos, a IA passou a ser aplicada sob o argumento de que seu uso era necessário para reduzir os gastos do sistema prisional e manter encarceradas apenas as pessoas que apresentavam maior risco à sociedade, a partir de análise de riscos futuros utilizados para dosimetria, regime de cumprimento e execução de penas etc. Os dados são apresentados aos agentes do sistema indicando um *score* (pontuação) com a probabilidade de reincidência (SULOCKI, 2019, p. 652-653).

A IA passou a ser amplamente utilizada para tentar contornar a situação de largo encarceramento do país, principalmente após algumas políticas públicas como

a guerra às drogas, com o objetivo da Defesa Social e corte de gastos. Assim, permaneceriam presos apenas as pessoas mais perigosas “desafogando o sistema” (SULOCKI, 2019, p. 653-654).

A IA, ao invés de na prática auxiliar na eficiência do sistema e redução de encarcerados, na verdade aprofundou a discriminação do sistema, afastando o devido processo legal e o julgamento justo, trazendo maior parcialidade aos processos. A subjetividade e experiências passadas não podem servir para o estabelecimento do risco de reincidência, mas a IA utiliza tais dados para estabelecer uma probabilidade de alguém voltar a cometer um crime, não a partir dos fatos trazidos ao processo. A ferramenta pode até trabalhar de forma “neutra”, mas os dados utilizados por ela refletem um sistema seletivo e discriminatório. Assim, a máquina também apresenta uma decisão enviesada, selecionando entre as pessoas de alto risco grupos já marginalizados (SULOCKI, 2019, p. 654-660).

A ONG ProPublica ao fazer uma pesquisa sobre o uso das ferramentas de *risk assessment* (avaliações de risco) descobriu que na verdade os negros eram classificados como alto risco e os brancos como baixo risco de forma incorreta. A Electronic Privacy Information Center (EPIC) também publicou análise¹⁹ do uso de IA no sistema criminal e apresentou dúvidas sobre seu uso, mais especificamente com relação à opacidade dos algoritmos, ou seja, da sua não publicidade, irrealidade e inconstitucionalidade e até mesmo, erros de cálculo. Assim, a aplicação da IA no sistema criminal tem representado uma violação a direitos individuais, o que indica a necessidade de maior controle sobre as ferramentas (SULOCKI, 2019, p. 660-664).

Um das ferramentas de IA utilizada em tribunais americanos e que tem provocado reflexões é o COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), que atribui a réus um *score* de um (baixa periculosidade) a dez (alta periculosidade) indicando o risco de cometimento de novos crimes. O COMPAS determina o *score* após analisar as respostas dos réus a um questionário com cento e trinta e sete perguntas sobre histórico familiar, ocupação, endereço, entre outros (FEET; MEDEIROS, 2021, p. 758).

Segundo a análise da ProPublica²⁰, apenas vinte por cento das sete mil pessoas

¹⁹ Disponível em <https://epic.org/issues/ai/ai-in-the-criminal-justice-system/>.

²⁰ ProPublica. **Machine Bias**. Disponível em <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Acesso em 4 de junho de 2020.

presas descritas como de alto risco pelo COMPAS em Broward, Flórida, cometeram novos crimes. Além disso, o COMPAS atribuiu o alto risco duas vezes mais a negros do que a brancos, sendo que estes normalmente eram considerados baixo risco. Retirando os dados de raça e histórico criminal dos dados dos presos em um teste, a ProPublica identificou que ainda assim a máquina atribuía aos negros um risco maior de cometerem novos crimes. A ONG atribuiu o resultado ao viés de raça das avaliações de risco ao próprio sistema criminal americano que prende majoritariamente negros. A ProPublica questionou também que a razão pela qual as ferramentas de IA foram utilizadas, redução do encarceramento, não foi alcançada. A Virgínia, um dos primeiros estados a utilizar IA em 2005, apresentou uma redução de 5% da população carcerária após dez anos de uso da ferramenta, sendo que na década anterior, a população carcerária foi reduzida em 31%.

4. IMPACTOS, POSSIBILIDADES E REGULAÇÃO NO USO DA INTELIGÊNCIA ARTIFICIAL EM MATÉRIA PENAL

4.1. IMPACTOS DA INTELIGÊNCIA ARTIFICIAL APLICADA À MATÉRIA PENAL

As pessoas associam a probabilidade de cometimento de crimes a pessoas negras, seja de forma explícita ou não, este costuma ser fator decisivo em sentenças. Nos Estados Unidos, os homens negros recebem penas vinte por cento mais duras que homens brancos e representam quarenta por cento da população carcerária, mesmo representando treze por cento da população. Os modelos de reincidência foram criados para reduzir o preconceito, evitando os vieses das decisões dos juízes. “A questão é que de fato eliminamos o viés humano ou o camuflamos com tecnologia” (O’NEIL, 2020, p. 26-29).

Os questionários a que uma pessoa presa é submetida deixarão em situação difícil um jovem negro e não uma pessoa privilegiada. Em avaliações como envolvimento anterior com a polícia, um negro, que normalmente vive em locais com alto número de abordagens, provavelmente já foi parado sem motivos. Em pesquisas,

negros²¹ e latinos representam mais de quarenta por cento das abordagens, e destas, noventa por cento das pessoas não tinha feito ou portava nada ilegal. “Se envolvimento precoce com a polícia indica reincidência, pessoas pobres e minorias raciais parecem muito mais perigosas”. (O’NEIL, 2020, p. 29).

Outra das perguntas, sobre parentes e amigos, não terá impacto para quem é de classe média, mas para um homem pobre fará diferença. Esses questionários foram utilizados por milhares de detentos desde sua criação, em 1995. As perguntas são irrelevantes para o processo e injustas, pois utilizam critérios que em outras circunstâncias causariam fortes objeções. As pessoas deveriam ser julgadas não por quem são, mas pelo que fazem. A avaliação faz com que a pena seja maior, aumenta a probabilidade de que a pessoa volte a ser presa e ao voltar à sociedade a pessoa tem dificuldades em encontrar emprego, o modelo cria um ciclo destrutivo que o retroalimenta (O’NEIL, 2020, p. 30-31).

O primeiro problema é que o modelo de reincidência é opaco, se as pessoas soubessem como ele funciona, não se autoincriminariam, o modelo e seus objetivos não são claros como no jogo de beisebol. Modelos transparentes, como o do beisebol, são a exceção. Os modelos são caixas-pretas impenetráveis, propriedade intelectual, que podem valer milhões como os algoritmos das gigantes de tecnologia, como, por exemplo, o Google (O’NEIL, 2020, p. 31-33). “Quando uma nova tecnologia é tão difusa e revolucionária como o *machine learning*, não é sábio deixá-la como uma caixa preta. A opacidade abre a porta para o erro e a utilização incorreta” (DOMINGOS, 2017, p. 13).

As caixas-pretas precisam ser abertas e seu conteúdo avaliado, já que “ciência e poder podem criar experiências desiguais”, e, por exemplo, o reconhecimento facial demonstra como os vieses provocam consequências morais, éticas e sociais graves. A IBM até mesmo encerrou as pesquisas em reconhecimento facial após os impactos

²¹ “Uma pesquisa realizada pelo Centro de Estudos de Segurança Pública e Cidadania (Cesec) mostra que 63% das abordagens policiais na cidade do Rio de Janeiro têm como alvo pessoas negras. Os dados inéditos revelam, segundo a coordenadora do estudo, a socióloga Silvia Ramos, o caráter racista como centro da atividade policial do estado do Rio de Janeiro. De acordo com o levantamento, **um quinto (17%) dessas pessoas já foi abordada pela polícia mais de 10 vezes**. Ao olhar para o local das abordagens, os pesquisadores perceberam que as atividades comuns para pessoas brancas são vistas como suspeitas para pessoas negras” (grifo meu). Disponível em <https://www.cnnbrasil.com.br/nacional/negros-correspondem-a-63-das-pessoas-abordadas-por-policiais-no-rio-de-janeiro/>. Acesso em 17 de abril de 2022.

que a sua tecnologia gerou e tem se posicionado contra o monitoramento e vigilância em massa. Mais de noventa por cento dos presos pelo reconhecimento facial no Brasil são negros (RUBACK; AVILA.; CANTERO, p. 2-8). Durante o uso dessa tecnologia em quatro dias de carnaval na Bahia, noventa e seis por cento dos alertas, após captura de mais de um milhão de rostos, não resultaram em prisões. No Rio de Janeiro, uma mulher foi indicada como procurada pela justiça, mas a verdadeira criminosa estava presa há quatro anos. (NUNES, 2019, p. 70)

Além disso, outra questão é que esses modelos podem até surgir de boas intenções, porém, costumam ir contra interesses e direitos das pessoas, são injustos. “Um fator crucial desse sofrimento é o ciclo nocivo de feedback, ou retroalimentação”. O terceiro problema é que esses modelos são aplicados em escala, estão em praticamente todas as áreas da vida, acabam “rapidamente estabelecendo normas gerais que exercem sobre nós algo muito próximo do poder da lei”. Isso ocorre no sistema de empréstimos, se um modelo de risco for aplicado a você, todo o sistema pode vê-lo como “caloteiro”. (O’NEIL, 2020, p. 33-34).

Buscou-se no passado solucionar os problemas do sistema prisional através da “imparcialidade” dos dados. Mas a aplicação dos sistemas de reincidência fez nascer uma arma de destruição matemática a partir de um modelo opaco, em escala e que provoca danos. Esses danos não são sentidos por todos, os modelos beneficiam “alguns sortudos” às custas do sofrimento de muitos (O’NEIL, 2020, p. 34-36).

O problema do uso da IA no Direito é que essa aplicação não chega precedida de estudos e pesquisas dos impactos do seu uso, e essa compreensão é essencial, já que a IA tem uma natureza dual (MEDEIROS, 2021, p. 749-750). Um ponto que chama atenção é a tendência de o Brasil copiar de forma pouco refletida essa tecnologia, sendo necessário, portanto, que a IA tenha foco no ser humano e uma aplicação ética, pautada no princípio da Dignidade Humana (SULOCKI, 2019, p. 660-664).

A IA tem encantado a muitos e apenas características como eficiência estão conduzindo o processo. Porém, no avanço da IA e no uso em decisões, podem ser gerados graves problemas. A IA apresenta riscos e sua aplicação deve ser feita de forma crítica. Um dos problemas que surge é a falta de transparência em razão da opacidade dos algoritmos. As empresas privadas que desenvolvem os projetos não

explicam os algoritmos em razão do segredo comercial e da proteção da propriedade intelectual (MEDEIROS, 2021, p. 762).

Caso referência sobre o COMPAS foi o julgamento de *Eric Loomis*, condenado a seis anos de prisão por cometer cinco crimes em um tiroteio em *La Crosse*. *Loomis* respondeu ao questionário do COMPAS que o classificou com alta probabilidade para cometimento de um novo crime. O resultado do COMPAS foi usado pelo juiz para a dosimetria da pena e negativa da liberdade condicional. Em moção, a defesa questionou o uso do COMPAS, alegando que foi violado o direito à sentença individualizada e de ser condenado por informações precisas. A moção foi negada, e em recurso à Suprema Corte de *Winsconsin* a decisão foi mantida sob os argumentos de que o uso do COMPAS não desrespeitou os direitos do réu e que ele receberia a mesma pena se não existisse a pontuação. A Suprema Corte dos Estados Unidos negou o “*writ of certiorari*” de *Loomis* entendendo que o tema é recente para ser apreciado (MEDEIROS, 2021, p. 759).

Dentre os fatores mais preocupante para o Direito está a falta de transparência que pode impactar as decisões judiciais. No caso do COMPAS, por exemplo, não há explicação do porquê a máquina atribuiu a pontuação negativa a *Loomis* e a ausência de transparência da decisão viola garantias processuais, como o contraditório e a ampla defesa, e prejudica a atuação das partes no processo. Além disso, podem ocorrer falhas na entrada dos dados que irão influenciar a decisão judicial. Até o Victor, do STF, pode decidir de forma equivocada ao negar recursos sem que seja possível auditar a máquina (MEDEIROS, 2021, p.765).

Um exemplo de viés é o de ancoragem que induz uma pessoa a decidir algo por ter tido contato com uma informação prévia. Como fora citado, provas ilícitas, mesmo quando desentranhadas tendem a influenciar a decisão do júri. Nesse ponto, o COMPAS causa grande influência, enviesando decisões dos que tomam contato com a pontuação da máquina. A Suprema Corte de *Winsconsin*, após o julgamento do caso *Loomis*, determinou que juízes que usam a pontuação devem demonstrar a existência de outras razões para a sentença. Ou seja, o contato com a avaliação do COMPAS impede que o juiz tome uma decisão sem ser influenciado pela pontuação, podendo ser considerado uma agressão ao princípio da presunção de inocência (MEDEIROS, 2021, p. 770-771). Em casos avaliados pela ProPublica, os juízes

anulam acordos entre promotoria e defesa por tomarem contato com a altas pontuação dada pelo COMPAS. A ONG identificou até mesmo declarações de juiz em audiência de que atribuiria pena menor a um réu se não tivesse tido contato com a avaliação da máquina.

Outro risco apresentado refere-se à consolidação de desigualdades, quando o aprendizado de máquina propaga vieses e preconceitos a partir dos dados. Assim, os algoritmos não são imparciais, e em alguns programas, como o COMPAS, a máquina atribui a negros e outros grupos vulneráveis alto risco para cometimento de novos crimes. Além dos dados, o viés pode decorrer dos próprios desenvolvedores. Como fora destacado, na análise da ProPublica foi verificado que o COMPAS atribuiu risco mais alto de negros cometerem novos crimes do que para brancos. Mas análises posteriores da ONG demonstraram que os brancos reincidiram mais do que os negros. Em outra pesquisa, o desempenho do COMPAS foi comparado ao desempenho de pessoas com pouco perícia, obtendo resultados semelhantes a quem não tem proximidade com o sistema criminal. (MEDEIROS, 2021, p. 775-776)

A perseguição da eficiência tem impedido o exercício de garantias e princípios do processo. Sistemas como o COMPAS impedem o exercício de garantias fundamentais como a ampla defesa e o contraditório. Tal constatação é crítica já que o contraditório é o principal meio para reduzir os vieses. Não se sabe que respostas determinam o resultado da máquina. O algoritmo pode então reproduzir a sociedade, perpetuando estereótipos, ou até ampliando desigualdades sociais. Além disso, corre-se o risco de o uso da IA “fossilizar as decisões”. Outra barreira é uma aura de inatacabilidade das decisões da IA por estarem revestidas de “imparcialidade”. Por isso a importância da transparência algorítmica que possibilite maior participação na decisão. (MEDEIROS, 2021, p. 777-783)

Todo o processo de aplicação da IA tem sofrido grande influência do neoliberalismo, que teve como marco histórico principal as reformas implementadas por Regan, nos Estados Unidos, e Thatcher, na Inglaterra. Os princípios do neoliberalismo têm norteado o Estado, que acaba deixando direitos em segundo plano e se adequando ao mercado. Dessa forma, a ideologia neoliberal permeia toda a sociedade moldando a política, a sociedade e a cultura, e claro, influenciando também o Direito (DIERLE E MARQUES, 2021, p. 718-721).

O neoliberalismo modificou o modelo processual brasileiro a partir da década de 1990, adequando o processo às necessidades do mercado, e reformas que acabaram por prejudicar direitos fundamentais. A “moda” seria a eficiência e a razoável duração do processo sem considerar direitos em jogo. Influenciado pela sociedade de consumo o processo precisaria agora produzir mais decisões, e não necessariamente produzir Justiça (DIERLE; MARQUES, 2021, p. 721-723).

Ao desenvolver um modelo existe uma escolha entre justiça e eficiência. A Constituição elege a presunção de inocência como princípio, mas para um modelador, que precisa desenvolver um sistema que avalie culpa ou inocência, a presunção de inocência é um entrave. Os sistemas tendem a favorecer a eficiência e não a justiça, justiça é difícil de quantificar, é algo que os computadores não entendem, mesmo com os avanços das máquinas nas áreas da lógica e da linguagem. Como justiça não pode ser calculada, injustiças são externalidades negativas comuns aos sistemas. É necessário escolher o que se deve sacrificar, talvez seja melhor menos eficiência e mais justiça. Para existir justiça é necessário igualdade e essa igualdade precisa ser refletida na justiça criminal, “justiça não pode ser apenas algo que uma parte da sociedade impõe sobre outra”. (O’NEIL, 2020, p. 106-108)

A pontuação de risco de cometer novos crimes supostamente é científica, neutra, e juízes acabam dando penas maiores a quem “aparenta” oferecer maior risco. Para saber o que realmente impacta a possibilidade de cometer novos crimes é necessário pesquisas, como por exemplo, analisar os efeitos do tempo que se passa preso e a relação com a reincidência. Mas dados sobre o sistema prisional costumam ser mais utilizados para manter e justificar o sistema do que o modificar e melhorar. A Amazon, por exemplo, tem uma grande equipe de cientistas de dados buscando aperfeiçoar seus algoritmos e as experiências dos clientes. Porém, no sistema prisional são feitas muitas vezes correlações que não representam verdadeiras soluções aos problemas. Muitos podem ser os impactos que o próprio sistema causa no comportamento dos presos que pode provocar reincidência, mas isto não é objeto de avaliações e de pesquisa. (O’NEIL, 2020, p. 109-111)

Um programa sério no sistema de justiça criminal teria como objetivo otimizar o sistema, assim como ocorre com os algoritmos da Amazon, para entender o sistema e seus efeitos, identificar pessoas que possam ser auxiliadas e beneficiar detentos e

sociedade. Presídios não querem passar por tais pesquisas ou mudanças. Assim como em companhias aéreas, presídios só dão lucro²² quando apresentam capacidade elevada. “Então, em vez de analisar e otimizar os presídios, lidamos com eles como caixas-pretas. Os detentos entram e somem de nossas vidas”. Ao entrar no sistema, e quanto mais tempo se passa nele, maiores são as chances de cometer novos crimes após soltura. (O’NEIL, 2020, p. 111-112)

A virada tecnológica no Direito que poderia auxiliar a resolver problemas está na verdade aprofundando-os. As *legal techs* “vendem” a ideia de IA como um milagre e pregam uma total desregulação e prevalência de liberdade. Muitos profissionais do Direito têm “comprado” tal discurso. Para os futuristas, a IA no Direito pode implementar um ambiente de maior justiça e promover o Estado de Direito. Porém, a eficiência não é o único valor que deve ser levado em conta, sendo necessário observar a ordem constitucional democrática. A própria IBM admitiu em estudo que com o passar do tempo os vieses algorítmicos serão ampliados (DIERLE; MARQUES, 2021, p. 724-727), o que demonstra não ser possível uma aplicação desregulada e com total liberdade da IA com a atual tecnologia.

Em matéria²³ da Associação dos Magistrados Brasileiros sobre o Expojud, evento denominado como “Congresso de Inovação, Tecnologia e Direito para o Ecosistema de Justiça do Brasil”²⁴, e que tem como patrocinadores diversas empresas de tecnologia, é possível perceber o posicionamento que alguns magistrados estão adotando sobre o uso da IA no Judiciário Brasileiro. A matéria informa que, em um painel composto por três juízes, os magistrados acreditam que a IA seja a ferramenta capaz de auxiliar os juízes a eliminar vieses em decisões judiciais.

“No mundo todo, esses sistemas de IA têm auxiliado os juízes a decidirem, nunca estão decidindo sozinhos”, contou a juíza federal Caroline Tauk.

²² “Ao privatizar unidades prisionais, o Estado transfere o seu poder punitivo a uma entidade com interesses meramente lucrativos. A Umanizzare recebeu em 2016 R\$ 326,3 milhões do governo estadual. Inclusive, o Ministério Público do Amazonas protocolou nesta semana um pedido ao Tribunal de Contas do Estado para o rompimento desse contrato, pois há indícios de superfaturamento, mau uso do dinheiro público, conflito de interesses empresariais e ineficácia da gestão”. **O massacre no Amazonas e as prisões privatizadas: o lucro como alma do negócio**. Disponível em https://brasil.elpais.com/brasil/2017/01/05/opinion/1483625278_386473.html. Acesso em 17 de abril de 2022.

²³ Associação dos Magistrados Brasileiros. **Juízes apontam como algoritmos podem ajudar Justiça a melhorar julgamentos**. Disponível em <https://www.amb.com.br/juizes-apontam-como-algoritmos-podem-ajudar-justica-melhorar-julgamentos/>. Acesso em 24 de março de 2022.

²⁴ <https://www.expojud.com/>

Ela afirmou que os vieses algorítmicos podem ser aprimorados desde a etapa de treinamento do algoritmo. “Primeiro: remover atributos sensíveis – raça, gênero – dos dados que treinam aquele algoritmo. Quando se tem uma pergunta no formulário de treinamento, mudar a forma como é feita aquela pergunta. Se, em princípio, os dados de treinamento estão livres de discriminação, **espera-se** que aquele modelo computacional aprendido também seja livre disso.” (grifo meu)

Como será demonstrado na próxima sessão deste trabalho²⁵, a IA realmente pode auxiliar na identificação de vieses decisórios. Mas é necessário maior cautela com os sistemas de inteligência artificial. Mesmo com a tentativa de evitar vieses no desenvolvimento da IA, como ocorreu no caso da seleção de colaboradores da Amazon e na seleção de melhores candidatos a tratamento preventivo por seguradora de saúde, exemplos citados anteriormente, o contexto de onde os dados é retirado pode ter sido construído com base em desigualdades, e os resultados da máquina serão enviesados, além da presença de vieses da escolha dos dados e da programação. Se a maior porcentagem de pessoas no sistema prisional é negra, retirar dados de gênero e raça pode não fazer diferença para treinar o algoritmo. Assim como no COMPAS não havia perguntas sobre raça, mas, mesmo assim, os dados mostraram a prática de encarceramento da população negra e o resultado foi enviesado, o que tende a ser perpetuado ou ampliado caso a IA seja utilizada na área criminal.

Outro ponto relevante, é a análise da teoria do Direito e verificação da possibilidade de matematização do processo. O Direito apresenta conexão com a linguagem, e não poderia sofrer uma tradução algorítmica. Direito é construção argumentativa, e é necessária abertura argumentativa para a continuidade do Direito. A legitimidade do Direito está baseada no discurso, por meio da estabilização de expectativas, da concordância entre os que estão submetidos à norma. Desta forma, a linguagem constitui o Direito, decisões legítimas são aquelas estabelecidas argumentativamente no processo e passíveis de responsabilização. (DIERLE; MARQUES, 2021,p. 727-730)

A cultura é formada por combinação de linguagem e contexto e serve de substrato para emoções, que variam de lugar em razão disso. Isso deixa claro que resposta justa não é resposta média, é preciso analisar o sujeito. As decisões

²⁵ No item “4.2. – Possibilidades de aplicação da Inteligência Artificial em matéria penal”, será demonstrado como por meio de pesquisas utilizando jurimetria foi possível identificar a presença de vieses de seleção em decisões do TJSP.

humanas partem de processos cognitivos nem sempre racionais, como por exemplo, elementos culturais, decisões inerentes ao ser humano e que não podem ser imitadas pela IA. Ou seja, é até possível estabelecer padrões do comportamento humano, mas existe diferença entre cognição e racionalidade. Dessa forma, seria impossível, diante da atual tecnologia, atribuir à máquina a capacidade de compreender o ser humano e decidir sobre restrições à liberdade, sendo possível a sua aplicação apenas a outros atos do processo penal. (PEDRINA, 1593-1594)

Portanto, a abertura semântica do Direito não se encaixa à exatidão matemática e se tal empreendimento for colocado em prática, a IA provocará uma redução do Direito, e tornará a sua aplicação limitada, pelo menos, com a tecnologia existente hoje. Dessa forma, no Direito o homem tem vantagem sobre a máquina. Uma demonstração de tal limitação é o projeto Debater da IBM que apresentou argumentos a favor e contra o uso de jogos violentos por menores de idade. Apesar de apresentar argumentos adequados, a IA deixou de apresentar vários argumentos abordados na decisão judicial do caso *Vídeo Software Dealers Assoc. v. Schwazenegger* na qual foi analisada a constitucionalidade de artigo do Código Civil que restringia a venda de jogos a menores. (DIERLE; MARQUES, 2021, p. 730-733)

Além disso, a decisão judicial é validada pela participação do cidadão, a partir do contraditório e da ampla defesa (influência e não surpresa) que ficam prejudicados caso a IA seja utilizada como um juiz-robô. A decisão judicial é o resultado de razões e não podem ser compreendidas ou debatidas com o uso da IA, o que prejudicaria a lógica processual. Dessa forma, a IA não pode ser utilizada, não com a tecnologia atual, sendo possível sua aplicação na gestão do processo. Não deve ser permitida a contaminação do Direito pelas influências do mercado permitindo que a IA passe a ser utilizada no processo decisório sob o risco de “atingir o pilar em que se funda o direito – o discurso racional” (p. 737) (DIERLE; MARQUES, 2021, p. 733-737).

Neste ponto, importante verificar como alguns magistrados percebem os sistemas de IA. Segundo José Faustino Macêdo, juiz de direito do TJPE, ele

“utiliza no dia a dia o sistema, batizado de Elis, e diz que, após conhecer ferramentas tecnológicas de outros tribunais do país, **classifica o sistema local como o único que ‘usa inteligência artificial no processo decisório’.** **‘Elis de certa forma decide.** Ela diz se o processo está ok ou não, e bota na minha caixa para eu assinar. **É como se me substituísse até.**

Agora, não digo que ela me substitui porque eu tenho que parar, logar e **posso** olhar, verificar se está certo ou não', afirma Macêdo²⁶". (grifo meu)

Destaca-se que o magistrado informa que “pode” analisar a decisão da máquina que já destina uma minuta para assinatura, o que demonstra que talvez, na prática, já tenhamos decisões realizadas pela IA no Judiciário Brasileiro, visto que analisar a decisão da máquina ou simplesmente só assinar a decisão parece ser uma faculdade do magistrado.

A pandemia de Covid-19 acelerou o uso da tecnologia no judiciário. Tais alterações ocorreram sem grandes debates, e as Cortes online passaram a ser utilizadas com base nas Resoluções Online de Disputas, que estão centradas na resolução de problemas e não tem um enfoque legal. Têm-se a ideia de judiciário como um serviço que deve atender a um resultado qualitativo, atendendo a pressupostos neoliberais. O processo não deve ser visto somente como uma forma, mas como garantidor de um espaço democrático. Portanto, as mudanças implementadas no período da pandemia devem levar em conta tais considerações e ampliar debates sobre a questão. (DIERLE; MARQUES, 2021, p. 738-741)

Os vieses já eram estudados no Direito, principalmente com relação às decisões pois, mesmo que inconscientemente, eles podem influenciar uma decisão interferindo no julgamento. Um dos vieses mais abordados é o de confirmação que faz com que o julgador busque e interprete informações que confirmem suas concepções. Outros vieses muito estudados são o social, que confirma as percepções sobre alguém com base em dado grupo social, o de ancoragem, que dificulta que afastemos uma primeira impressão, e o viés de conformidade que nos faz sempre estar ligados a quem tem a mesma opinião que nós. Apesar dos diversos estudos, juízes resistem ao tema e não desenvolvem técnicas capazes de evitar o enviesamento. Porém, em um processo decidido por humanos é possível impugnar uma decisão enviesada, em uma decisão de IA isso se torna quase impossível. (DIERLE; MARQUES, 2018, p. 8)

Os valores que permeiam a máquina apresentam maior relevância em questões mais sensíveis, como as que envolvem direitos, o que exige “cuidados e

²⁶ Folha de São Paulo. **Inteligência artificial atua como juiz, muda estratégia de advogado e 'promove' estagiário.** Disponível em <https://www1.folha.uol.com.br/poder/2020/03/inteligencia-artificial-atua-como-juiz-muda-estrategia-de-advogado-e-promove-estagiario.shtml>. Acesso em 28 de fevereiro de 2022.

conhecimentos especiais em seu desenvolvimento”. O viés pode ser relevante, mas difícil de observar, é preciso reduzir ao máximo o enviesamento da máquina. As decisões precisam ser controladas e é necessário transparência (*accountability*) em respeito ao devido processo constitucional. O uso irrefletido da IA “poderá lesar princípios constituidores do Estado Democrático de Direito”, devido processo constitucional e acesso à Justiça, por exemplo (DIERLE; MARQUES, 2018, p.10).

O uso da IA no mundo já demonstra problemas no processo decisório em razão de sua opacidade e impossibilidade de conhecer motivos e da possibilidade de impugnação da decisão. É necessário assumir a questão e reduzir vieses, assim como ocorre com os humanos (ou deveria ocorrer) e aplicar a IA a partir de reflexões e garantir a transparência algorítmica (DIERLE; MARQUES, 2018, p. 11).

Em resumo, a IA não funciona de forma semelhante ao cérebro humano e não pode substituir a capacidade cognitiva do julgador humano por falta de elementos que estão além da razão. Além disso, o uso da IA em decisões pode também intensificar desigualdades sociais por contaminação de dados. É necessário também transparência e desenvolvimento de algoritmos em acordo com a Constituição e com os direitos humanos. É necessário não permitir que a IA se torne elemento pré-constituído de culpa, buscando uma suposta “neutralidade” no processo (PEDRINA, 1603). Dessa forma, a IA seria utilizada apenas como apoio, na gestão processual, ou em atos que não envolvem decisões.

4.2. POSSIBILIDADES DE APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL EM MATÉRIA PENAL

O uso da IA pode ser estabelecido no estado da arte atual em teses em tribunais, no estabelecimento de standards probatórios, busca por evidências e na formulação de perguntas. Com relação à aplicação de teses, a IA consegue por meio da leitura dos processos identificar o conteúdo dos autos e separa-os por temas e até devolve processos aplicando tese já aprovada ou faz o sobrestamento até a decisão, como ocorre com o Victor do STF. Os *standards* probatórios podem ser elaborados a partir da análise de decisões que podem indicar qual seria a quantidade de entorpecente que caracterizaria o crime de tráfico, por exemplo. Com relação às evidências, a IA pode auxiliar a encontrar informações nos autos, apoiando

investigações e testemunhos e reduzindo confusões narrativas. Por meio de perguntas, a IA pode auxiliar o juiz a tomar decisões mais neutras em inquirições. Necessário em todos os casos transparência da IA para possibilitar a análise dos resultados e verificação de vieses. (PEDRINA, 1597-1598)

A IA é também uma importante ferramenta a ser utilizada na pesquisa jurídica. Em pesquisas utilizando *machine learning* nos Estados Unidos, foi possível identificar que os argumentos técnicos na análise de trezentos e oitenta e sete mil casos em que houve dissenso em decisões desde 1880 somavam quase cinquenta por cento do total contra cerca de trinta e sete por cento de questões subjetivas. Assim, grande parte dos argumentos é técnico na maior parte das decisões, ficando a cargo de decisões difíceis uma maior possibilidade de influência de argumentos subjetivos (políticos, ideológicos etc.). (HORTA, 2019, p. 45-47).

A “opinião pública” divide-se quando o assunto é o início da execução penal. Alguns argumentam que a demora na execução só serve a protelação do início do cumprimento da pena e que isso congestiona os tribunais com vários recursos. Para outros, a presunção de inocência deve ser preservada sob pena de cometimento de injustiças. Tanto a presunção de inocência quanto a efetividade do processo são importantes. Qual a proporção de decisões que são reformadas em apelação? (NUNES; TRECENTI, 2015, p. 1-2).

Para verificar tal questão, foi elaborado estudo jurimétrico com base em 157.379 decisões em segunda instância das Câmaras de Direito Criminal do Tribunal de Justiça do Estado de São Paulo (TJSP) e das Câmaras Extraordinárias utilizando as técnicas de Inteligência Artificial. Os dados, que são públicos, foram organizados e informações ausentes foram tratadas estatisticamente. De todas as decisões, 68.044 referiam-se a uma única pessoa (em alguns casos havia dois processos, uma apelação do réu e outra do Ministério Público). Deste total, foram isoladas apenas as apelações em que o réu pedia anulação ou reforma da sentença para redução da pena, restando 57.625 processos, dos quais 618 processos não puderam ser analisados por ausência de dados (NUNES; TRECENTI, 2015, p. 2-5).

Como resultado, verificou-se que a taxa de reforma das sentenças foi de cerca de cinquenta por cento, confirmando estudos que perceberam que tal taxa refere-se a

um viés de seleção²⁷. A lide se estabelece apenas quando as duas partes acreditam que podem ganhar em um processo, caso contrário, haveria um acordo (autor tem razão) ou não seria exercido o direito de ação (réu tem razão). O viés de seleção já identificado refere-se ao processo civil, e, presentes as mesmas taxas no processo penal, pode indicar a ocorrência de viés de seleção desconhecido (NUNES; TRECENTI, 2015, p. 5-6).

Dessa forma, aprofundou-se a análise para verificar como essas taxas eram distribuídas. Dividindo-se por assunto, os crimes mais graves possuem menor taxa de reforma e os menos graves taxas mais altas.

“Os assuntos com maiores taxas de recursos negados são homicídio qualificado, latrocínio, roubo e estupro de vulnerável, justamente os tipos de crimes considerados mais graves. Os três assuntos com menores taxas de recursos negados são os crimes de trânsito, violação de direito autoral e outros assuntos, que agrupa todos os assuntos envolvendo trinta ou menos recursos” (NUNES; TRECENTI, 2015, p. 7)

Ao separar os processos em razão do órgão julgador, considerando que as câmaras ordinárias e extraordinárias não teriam diferenças materiais segundo o regramento do TJSP e, portanto, têm as mesmas competências, “observamos discrepâncias enormes, encontrando câmaras com mais de setenta e cinco por cento de recursos negados (quarta e sexta) e câmaras com menos de trinta por cento de recursos negados (primeira, segunda e décima segunda)” (NUNES; TRECENTI, 2015, p. 7-8).

Figura 3: Frequências dos assuntos dos processos segundo órgão julgador.

²⁷ “Quando a amostra de unidades usada para uma análise estatística é limitada de alguma maneira por processos não aleatórios, existe a possibilidade de viés de seleção. A seletividade da amostra pode render resultados distorcidos porque a natureza das relações entre variáveis nos casos incluídos na amostra não é necessariamente igual às relações que observaríamos em uma amostra aleatória”. **Vieses e a importância de sempre analisar o cenário completo.** Disponível em <https://medium.com/blogcarlavieira/vieses-estatisticos-d12b38a27c4c>. Acesso em 17 de abril de 2022.

Câmara	Furto	Furto Qual.	Outros	Roubo Maj.	Tráf. Drogas	Total
4ª Câmara	242 (6.4%)	419 (11.0%)	1641 (43.2%)	568 (14.9%)	933 (24.5%)	3803 (100%)
3ª Câmara	233 (6.6%)	431 (12.2%)	1503 (42.6%)	522 (14.8%)	842 (23.8%)	3531 (100%)
14ª Câmara	246 (7.1%)	399 (11.6%)	1411 (41.0%)	535 (15.5%)	853 (24.8%)	3444 (100%)
3ª Câmara Extr.	230 (6.8%)	397 (11.7%)	1337 (39.5%)	640 (18.9%)	785 (23.2%)	3389 (100%)
4ª Câmara Extr.	201 (5.9%)	403 (11.9%)	1295 (38.3%)	605 (17.9%)	881 (26.0%)	3385 (100%)
9ª Câmara	174 (5.4%)	298 (9.2%)	1345 (41.6%)	501 (15.5%)	916 (28.3%)	3234 (100%)
1ª Câmara Extr.	193 (6.0%)	402 (12.5%)	1262 (39.3%)	560 (17.5%)	792 (24.7%)	3209 (100%)
10ª Câmara	184 (5.8%)	326 (10.3%)	1458 (45.9%)	435 (13.7%)	772 (24.3%)	3175 (100%)
2ª Câmara Extr.	196 (6.2%)	410 (12.9%)	1272 (40.1%)	581 (18.3%)	716 (22.6%)	3175 (100%)
7ª Câmara	188 (6.0%)	330 (10.5%)	1328 (42.3%)	476 (15.2%)	817 (26.0%)	3139 (100%)
8ª Câmara	131 (4.4%)	258 (8.6%)	1220 (40.6%)	541 (18.0%)	852 (28.4%)	3002 (100%)
16ª Câmara	151 (5.4%)	298 (10.7%)	1193 (43.0%)	436 (15.7%)	696 (25.1%)	2774 (100%)
11ª Câmara	181 (6.6%)	293 (10.6%)	1141 (41.3%)	448 (16.2%)	699 (25.3%)	2762 (100%)
12ª Câmara	171 (6.6%)	253 (9.8%)	1136 (43.9%)	407 (15.7%)	623 (24.1%)	2590 (100%)
6ª Câmara	167 (6.6%)	264 (10.4%)	1103 (43.3%)	372 (14.6%)	641 (25.2%)	2547 (100%)
13ª Câmara	102 (4.1%)	235 (9.6%)	874 (35.5%)	462 (18.8%)	786 (32.0%)	2459 (100%)
2ª Câmara	136 (5.9%)	239 (10.4%)	992 (43.0%)	359 (15.6%)	580 (25.2%)	2306 (100%)
1ª Câmara	110 (5.0%)	230 (10.4%)	935 (42.2%)	332 (15.0%)	611 (27.5%)	2218 (100%)
5ª Câmara	95 (5.4%)	158 (8.9%)	683 (38.6%)	307 (17.4%)	525 (29.7%)	1768 (100%)
Total	3406 (5.91%)	6171 (10.7%)	23845 (41.4%)	9376 (16.3%)	14827 (25.7%)	57625 (100%)

Fonte: Tabela retirada do artigo “Reformas de decisão nas câmaras de direito criminal em São Paulo”.

Como uma primeira hipótese, buscou-se verificar se havia discrepância no momento da distribuição, se assuntos relativos a crimes mais graves ficavam concentrados em algum órgão julgador, porém, não havia tal diferença. Uma segunda hipótese estaria ligada ao comportamento do relator, e ao distribuir os processos por relator, em dois casos a negativa para apelações dos réus ultrapassava 85% (NUNES; TRECENTI, 2015, p. 9-11).

Figura 4: Frequências dos resultados dos processos segundo relator

Relator	Negaram	Outros	Parcialmente	Provido	Total
Eduardo Abdalla	599 (50.3%)	6 (0.5%)	369 (31.0%)	218 (18.3%)	1192 (100%)
Aguinaldo de Freitas Filho	804 (68.2%)	2 (0.2%)	189 (16.0%)	184 (15.6%)	1179 (100%)
Cesar Augusto A. de Castro	610 (53.1%)	2 (0.2%)	322 (28.0%)	214 (18.6%)	1148 (100%)
Alexandre Almeida	454 (40.1%)	1 (0.1%)	385 (34.0%)	293 (25.9%)	1133 (100%)
Julio Caio Farto Salles	747 (66.1%)	1 (0.1%)	193 (17.1%)	189 (16.7%)	1130 (100%)
Luis Augusto de S. Arruda	565 (50.0%)	6 (0.5%)	388 (34.4%)	170 (15.1%)	1129 (100%)
Mauricio Valala	502 (47.2%)	6 (0.6%)	341 (32.1%)	214 (20.1%)	1063 (100%)
Edison Brandão	931 (87.7%)	9 (0.8%)	100 (9.4%)	22 (2.1%)	1062 (100%)
Ivana David	883 (86.5%)	6 (0.6%)	71 (7.0%)	61 (6.0%)	1021 (100%)
Zorzi Rocha	852 (84.4%)	1 (0.1%)	98 (9.7%)	59 (5.8%)	1010 (100%)
Luiz Antonio Cardoso	509 (51.1%)	9 (0.9%)	375 (37.6%)	104 (10.4%)	997 (100%)
Roberto Mortari	601 (63.5%)	5 (0.5%)	232 (24.5%)	109 (11.5%)	947 (100%)
Ruy Alberto Leme Cavalheiro	672 (72.0%)	5 (0.5%)	118 (12.6%)	138 (14.8%)	933 (100%)
Lauro Mens de Mello	536 (57.9%)	4 (0.4%)	250 (27.0%)	135 (14.6%)	925 (100%)
Total	31059 (53.9%)	347 (0.638%)	17877 (31%)	8342 (14.5%)	57625 (100%)

Fonte: Tabela retirada do artigo “Reformas de decisão nas câmaras de direito criminal em São Paulo”.

Em síntese, os dados sugerem que (i) existe viés de seleção em casos criminais, que (ii) a taxa de reforma da sentença em crimes mais graves é mais baixa (menos de dez por cento) e que (iii) a negativa aos réus varia entre dezesseis a oitenta e um por cento a depender da câmaras. Sobre a discrepância na negativa dos recursos entre as câmaras as hipóteses seriam ou que a distribuição dos processos não seria aleatória, ou que o comportamento dos magistrados influencia esse resultado e, independente de qual seja a hipótese, ambas atestariam grande insegurança jurídica. Por fim, os dados revelaram que seria inapropriado iniciar o cumprimento da pena após a sentença visto que “os dados mostram que é relevante a quantidade de apelações contra o Ministério Público que são parcialmente e totalmente providos em segunda instância”. (NUNES; TRECENZI, 2015, p. 12-13).

Dessa forma, além da aplicação da máquina na gestão processual, a pesquisa jurídica poderá se valer da pesquisa empírica a partir da análise de dados tratados por IA, apresentando argumentos mais sólidos e embasados em dados para auxiliar a compreender e melhorar a prática jurídica.

4.3. REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL

Em 2021²⁸, surgiu o primeiro acordo global sobre ética na aplicação da IA

²⁸ United Nations. **193 countries adopt first-ever global agreement on the Ethics of Artificial Intelligence** Disponível em <https://news.un.org/en/story/2021/11/1106612>. Acesso em 15 de fevereiro de 2022.

assinado pelos membros da Unesco. O documento buscou estabelecer diretrizes sobre regulação e fiscalização da IA e as vantagens e desvantagens do seu uso. Entre as diretrizes principais estão o controle de dados pessoais e a vedação de sistema de IA para pontuação social e vigilância em massa. Recentemente, com o incremento da tecnologia e a acurácia elevada, modelos baseados em IA estão sendo amplamente utilizados, em todas as áreas da vida e seu uso acaba por gerar violações a direitos humanos. (KAUFMAN, 2022, p. 50)

Por isso, é urgente a regulação dessa tecnologia para redução de impactos éticos e sociais, porém as tentativas de regulação privada mostram-se infrutíferas. Muito da fonte regulatória na área era proveniente do *Asilomar Principals* que buscou estabelecer uma IA benéfica. Porém, os princípios são abstratos e difíceis de serem traduzidos para linguagem matemática. (KAUFMAN, 2022, p. 50-51)

A autorregulamentação das empresas privadas de tecnologia é uma ilusão. Temas éticos que antes eram discutidos apenas em âmbito acadêmico passaram a ser considerados pela opinião pública, como enviesamento, notícias falsas, proteção à privacidade, entre outros. Inicialmente havia tendência de estabelecer apenas *soft laws*²⁹ para o setor, porém todas as iniciativas estabelecidas pelo setor não tiveram resultados concretos. Como exemplos, cita-se o *Advanced Technology External Advisory Council* criado pelo Google e ironizado pelo MIT³⁰ por ter durado apenas uma semana e o órgão independente para revisão de decisões sobre conteúdo criado pelo Facebook e que não possui legitimidade para operar alterações na plataforma. É necessário que as empresas de tecnologia participem do processo, mas é o poder público que deve avaliar como elas impactam a sociedade. As mais fortes empresas hoje nasceram em um momento de grande confiança no mercado, porém eventos

²⁹ “Refere-se como soft law, geralmente, aqueles instrumentos ou documentos particulares não vinculantes ou ainda provisões não vinculantes em tratados. A terminologia indica que o instrumento ou provisão em questão não é “lei”, mas sua importância dentro da estrutura geral do desenvolvimento legal internacional e tal que merece uma atenção especial. Soft law não é lei, mas tem influência na política internacional. Sua utilização através de documentos que estipulem recomendações, diretrizes, códigos de condutas ou padrões é significativa em sinalizar a evolução de diretrizes, que podem, no fim das contas, serem convertidas em regras legalmente vinculantes. A soft law é importante e influente, mas não constituem normas legais por si só.” NASSER, Salem H. Direito internacional público. [Digite o Local da Editora]: Grupo GEN, 2013. 9788522475223. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788522475223/>. Acesso em: 11 abr. 2022.

³⁰ MIT Technology Review. **Hey Google, sorry you lost your ethics council, so we made one for you.** Disponível em <https://www.technologyreview.com/2019/04/06/65905/google-cancels-ateac-ai-ethics-council-what-next/>. Acesso em 28 de fevereiro de 2022.

recentes as deixam sob suspeita, como a Grande Recessão dos anos 2000 (KAUFMAN, 2022, p. 50-52), o escândalo da *Cambridge Analítica*³¹, entre outros.

Existe, portanto, a necessidade de regulação pelo poder público. A comissão Europeia e a China divulgaram recentemente proposta de regulação, os Estados Unidos pretendem reapresentar projeto parado no Senado americano. A regulação tende a ser elaborada de forma participativa, em uma sinergia entre poder público, empresas de tecnologia e sociedade civil. O Instituto de Engenheiros Eletricistas e Eletrônicos (IEEE), por exemplo, elaborou uma espécie de certificado³² para organizações que utilizam IA em conformidade a critérios éticos básicos. (KAUFMAN, 2022, p. 52-53)

Uma proposta de regulação de IA deveria atender a algumas diretrizes como (1) mapeamento de riscos e análise de impacto, (2) gestão de dados utilizados, (3) registro de desenvolvimento e funcionamento do sistema, (4) registro automático de eventos, (5) interface que facilite auditorias, (6) informação transparente para as pessoas que um modelo em uso é baseado em IA, (7) transparência de resultados e (8) implementação de testes de segurança. Porém, existem limitações à regulação como a já exposta lacuna de conhecimento geral sobre a IA, dificuldades de auditoria em razão do sigilo comercial. (KAUFMAN, 2022, p. 55)

No Brasil, o Projeto de Lei 21/2020³³ foi aprovado na Câmara dos Deputados e encaminhado ao Senado após um período de debates de apenas três meses. Os debates na Câmara demonstraram que o legislador não tem conhecimento básico do assunto, o que deixa claro o distanciamento entre desenvolvedores e reguladores. O projeto inclusive prevê um ponto que tem causado controvérsia que é a responsabilização subjetiva de desenvolvedores o que provocou reação de juristas brasileiros em documento enviado ao Senado intitulado “Carta Aberta de Juristas ao

³¹ ALVES, Paulo. **Facebook e Cambridge Analytica: sete fatos que você precisa**. Disponível em <https://www.techtudo.com.br/noticias/2018/03/facebook-e-cambridge-analytica-sete-fatos-que-voce-precisa-saber.ghtml>. Acesso em 4 de março de 2022.

³² “The IEEE CertifAIEd™ Program offers a risk-based framework supported by a suite of AI ethical criteria that can be contextualized to fit organizations’ needs— helping them to deliver a more trustworthy experience for their users. IEEE CertifAIEd Ontological Specifications for Ethical Privacy, Algorithmic Bias, Transparency, and Accountability are an introduction to our AI Ethics criteria. Fill out the form to start receiving these specifications”. Disponível em <https://engagestandards.ieee.org/ieeecertifaiied.html>. Acesso em 15 de fevereiro de 2022.

³³ Disponível em <https://www.camara.leg.br/propostas-legislativas/2236340>.

Senado Federal”³⁴ que destaca que é preciso maior conhecimento e envolvimento de todos sobre o tema (KAUFMAN, 2022, p. 53-54).

Segundo informa o art. 1º do Projeto de Lei nº 21 de 2020, de autoria do Deputado Federal Eduardo Bismarck, e que, se aprovado, será o marco legal da Inteligência Artificial no Brasil, o PL 21/2020 busca estabelecer

“princípios, direitos, deveres e instrumentos de governança para o uso da inteligência artificial no Brasil e determina as diretrizes para a atuação da União, dos Estados, do Distrito Federal e dos Municípios, pessoas físicas e jurídicas, de direito público ou privado, e entes sem personalidade jurídica em relação à matéria”.

O PL alcança diversas áreas e por isso não pode ter uma tramitação acelerada, sem realização de debates, consultas e audiências públicas. Na Câmara dos Deputados, o Projeto tramitou em regime de urgência o que impediu que fosse dada maior contribuição à proposta. Em comparação, a recente proposta de regulamento do Parlamento Europeu contou com mais de mil contribuições de diferentes instituições (FGV, 2021, p. 13-14). O Marco Civil da Internet tramitou por três anos na Câmara dos Deputados e mais dois anos no Senado Federal e contou com mais de 800 contribuições (KAUFMAN, 2022, p. 53).

O PL se coloca como vinculativo, determinando até mesmo a demonstração do cumprimento das normas nele constantes, impõe diversos deveres, tanto ao Poder Público como a particulares, com atributos de “abstratividade, generalidade, imperatividade e coercibilidade”, porém (FGV, 2021, p. 16)

“em relação à sua eficácia normativa, o texto não prevê claramente as responsabilidades, e tampouco fixa as penalidades em caso de descumprimento, como faz, por exemplo, o artigo 71 da Proposta de Regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de inteligência artificial”. (FGV, 2021, p. 14-16)

Outros países estão em processos regulatórios incipientes ou em construção, como o Brasil. Em relação especificamente à Justiça, a Comissão Europeia para a Eficiência da Justiça publicou a Carta Ética Europeia para o uso da IA no Judiciário³⁵ em 2018. Em 2019, a Organização para a Cooperação e Desenvolvimento Econômico

³⁴ Disponível em <https://www.change.org/p/senado-federal-carta-aberta-de-juristas-ao-senado-federal-contr-o-artigo-6o-inciso-vi-do-pl-21-a-2020>.

³⁵ Comissão Europeia para a Eficiência da Justiça (CEPEJ). **Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu ambiente**. Disponível em <https://rm.coe.int/carta-etica-traduzida-para-portugues-revista/168093b7e0>. Acesso em 30 de março de 2022.

(OCDE)³⁶ estabeleceu documento com princípios para que a “IA seja confiável e respeite os direitos humanos e os valores democráticos” (p. 17), e o Brasil é um dos signatários do documento. Comparando o PL 21/2020 com diplomas como o Marco Civil da Internet e a Lei Geral de Proteção de Dados, estes têm um número maior de artigos, mas apresentam uma estrutura semelhante ao PL 21/2020, “o que reforça que não se trata apenas de orientação” (FGV, 2021, p. 16-18).

Outro exemplo é a Estratégia Brasileira de Inteligência Artificial (Ebia)³⁷ construída pelo Ministério da Ciência, Tecnologia e Inovações que aborda “eixos transversais (legislação, regulação, uso ético, governança e aspectos internacionais) e verticais (educação, força de trabalho e capacitação, empreendedorismo, aplicação no Poder Público e segurança pública)”. Assim, o Projeto precisa de alterações e aprofundamento em alguns pontos (FGV, 2021, p. 18).

O conceito de IA trazido pelo Projeto de Lei precisa ser aprimorado, os termos relativos à IA precisam ser melhor descritos para evitar insegurança jurídica e o texto apresenta algumas expressões vagas³⁸. O PL parece se inspirar no documento construído pela OCDE, mas a “moldura ética” não está bem regulamentada, há imprecisão e vagueza (FGV, 2021, p. 18-23).

Nessa linha, o Conselho Nacional de Justiça (CNJ) tratou do assunto por meio da resolução 332/2020 e da portaria 271/202032. A primeira dispôs sobre a ética, a transparência e a governança na produção e no uso de inteligência artificial no Poder Judiciário”; enquanto a segunda regulamentou o uso de inteligência artificial no âmbito do Poder Judiciário” (FGV, 2021, p. 23).

Não há clareza no Projeto sobre responsabilidade, gerenciamento de risco e prestação de contas com relação ao uso da IA. Neste ponto, sugere-se observar as discussões internacionais e estabelecer diferentes níveis de gerenciamento de risco a depender da complexidade do sistema em uso, estabelecendo-se um maior gerenciamento em razão de sistema de IA que apresente maior risco. O Projeto precisaria, portanto, passar pelo crivo de grupo de trabalho criado para melhor analisar a forma como foi construído seu texto e por novas consultas públicas, aprofundando-se assim os debates sobre o tema (FGV, 2021, p. 25-27).

³⁶ OECD Legal Instruments. **Recommendation of the Council on Artificial Intelligence**. Disponível em <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>. Acesso em 2 de abril de 2022.

³⁷ Ministério da Ciência, Tecnologia e Inovações. **Estratégia Brasileira de Inteligência Artificial**. Disponível em <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial>. Acesso em 15 de fevereiro de 2022

³⁸ “O termo ‘objetivos definidos pelo homem’ é por demais vago e destoa do que pode ser encontrado na literatura da área”. (FGV, 2021, p. 21)

No Judiciário Brasileiro, os primeiros projetos de IA foram implementados em 2018, totalizando 54 projetos até dezembro 2020, segundo o relatório da FGV. A resolução 332 de agosto de 2020 do Conselho Nacional de Justiça, sendo, portanto, posterior a maior parte dos projetos, versa sobre “sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário” traz em seu bojo algumas considerações sobre os vieses algorítmicos.

A resolução do CNJ inspirou-se no documento publicado pela Comissão Europeia para a Eficiência da Justiça de 2020. O documento destaca a necessidade de proteção de direitos fundamentais, a segurança jurídica e a igualdade das partes em casos idênticos, respeito a igualdade, não discriminação, julgamento justo, e a redução e preconceitos nos julgamentos. (FGV, 2020, p. 20-21)

“Essa preocupação com a escalada do preconceito nas decisões com suporte da IA fez com o CNJ optasse por estabelecer a necessidade de homologação dos projetos de inteligência artificial desenvolvidos nos tribunais, a fim de detectar vieses dos algoritmos ou tendência discriminatória no funcionamento da tecnologia. Caso essa inclinação seja verificada, os desenvolvedores deverão realizar as correções devidas. Os projetos de IA que utilizam técnicas de reconhecimento facial exigem uma autorização prévia do CNJ para implementação.” (FGV, 2020, p. 21)

A resolução determina que as decisões da IA devem fornecer explicações tornando possível auditar as decisões. Os projetos podem ser desenvolvidos pelos tribunais ou com apoio, seja de universidades, instituições públicas ou privadas e os modelos devem ser construídos de forma comunitária. Não existiria vinculação à uma decisão de IA. Em matéria penal, haveria maior proteção já que o documento não estimula o uso da IA para decisões preditivas. Haveria exceções para uso no gerenciamento do processo, verificação de reincidência, prescrição e cálculo de penas. (FGV, 2020, p. 21-22)

Não ser estimulado o uso da IA em decisões preditivas em matéria penal seria diferente de “não deve ser aplicado o uso de IA em decisões preditivas em matéria penal”? A redação do art. 23 da referida resolução impede que algum tribunal brasileiro utilize decisões preditivas? Sobre decisões judiciais fornecidas pela IA, já é possível até mesmo encontrar artigos³⁹ de juristas brasileiros capitaneando alterações

³⁹ Ferro, S. (2021). **Permissibilidade do juiz robô no sistema jurídico brasileiro**. REVISTA ELETRÔNICA DA PGE-RJ, 4(1). Disponível em <https://revistaeletronica.pge.rj.gov.br/index.php/pge/article/view/208#:~:text=Nesse%20contexto%2C%20o%20objetivo%20do,no%20C3%A2mbito%20jurisdicional%2C%20utilizando%2Dse>. Acesso em 18 de março de 2022.

constitucionais para tornar possível que a máquina e não um ser humano possa decidir em processos judiciais. Assim,

“o mundo real, com toda sua bagunça, fica de lado. A inclinação é de substituir pessoas por rastros de dados, transformando-as em compradores, eleitores ou trabalhadores mais eficientes, para otimizar algum objetivo. Isso é fácil de fazer – e de justificar – quando o sucesso viola na forma de uma pontuação anônima, e quando as pessoas afetadas permanecem tão abstratas quanto os números dançando pela tela”. (O’NEIL, 2020, p. 55).

Existe uma separação entre os modelos e as pessoas, e não há preocupação com as consequências morais disso. A falsa sensação de segurança faz com que se aplique esses modelos imperfeitos, e quem se opõe é considerado um ludista⁴⁰ nostálgico. Os algoritmos aumentam as desigualdades fazendo com que pessoas vulneráveis fiquem em estado ainda mais crítico (O’NEIL, 2020, p. 55). Tal termo foi utilizado em análise⁴¹ sobre decisão judicial em São Paulo que impediu que tivesse continuidade um projeto de reconhecimento facial no metrô do estado. O reconhecimento facial, como destacado na matéria citada, tem tido os projetos interrompidos pois detecta-se violação de direitos humanos em seu uso.

A IA não poderia ainda fornecer um cálculo da pena maior que o magistrado calcularia sem este recurso em prejuízo ao réu (FGV, 2020, p. 22). Qual seria, portanto, o momento em que o magistrado teria acesso ao cálculo da máquina? Ficaria registrada a decisão do magistrado antes que ele tivesse acesso a uma possível decisão de IA? No uso do COMPAS, magistrados informaram que dariam penas menores, mas pelo efeito do viés de ancoragem, eles acabaram calculando pena superior após contato com a decisão da máquina, e em alguns casos, não homologando acordos já realizados entre defesa e a promotoria, que possuíam períodos de reclusão menores aos das penas efetivamente aplicadas.

Os vieses podem ser reduzidos quando é necessário justificar uma decisão para uma audiência ou quando institucionalmente é estabelecido um ambiente para

⁴⁰ “Ludismo ou luddismo foi um movimento de trabalhadores ingleses do ramo de fiação e tecelagem, ativo no início do século XIX, nos primórdios da Revolução Industrial, e que se notabilizou pela destruição de máquinas como forma de protesto”. Disponível em <https://artsandculture.google.com/entity/m04jn3?hl=pt>. Acesso em 30 de março de 2022.

⁴¹ “Os ludistas queimavam os teares mecânicos para defender o trabalhador. A decisão barra a tecnologia com medo de seu mal uso. Perde a oportunidade para discutir o que é realmente relevante”. **Ludismo judicial e o reconhecimento facial em serviços públicos**. Disponível em <https://www.jota.info/opiniao-e-analise/colunas/publicistas/ludismo-judicial-reconhecimento-facial-em-servicos-publicos-29032022>. Acesso em 30 de março de 2022.

tal. A necessidade de explicar argumentos melhora a decisão e reduz vieses, a vigilância epistêmica⁴² do grupo acaba ativando um processo cognitivo mais elaborado. Mesmo em discussões valorativas mais sensíveis há evidências de que esse processo de discussão em grupo gera uma decisão melhor. Para evitar ambientes em que haja disseminação de informações incorretas é necessário que o ambiente de discussão seja diverso, possibilitando a identificação de erros de raciocínio (HORTA, 2019, p. 41-42).

A resolução, em seu art. 20⁴³, acertaria ao buscar a diversidade na formação de equipes de pesquisa, desenvolvimento e capacitação em IA. Porém, abre a possibilidade em seu parágrafo segundo para que a diversidade não seja atendida caso as equipes não possam ser assim estabelecidas, por ausência de profissionais nos tribunais. Diante da necessidade de redução de vieses da máquina em seu desenvolvimento, seria possível estabelecer parcerias em pesquisa e desenvolvimento para promover a diversidade? Ao invés de apresentar possibilidade de motivar a ausência de diversidade, seria possível apresentar alternativas ao parágrafo segundo?

Provavelmente será difícil estabelecer diversidade de gênero e raça, por exemplo, se for levado em conta apenas a equipe dos tribunais em possível projeto. Em um país em que é necessário estabelecer cotas em universidades e concursos públicos, é provável que os tribunais não possuam diversidade racial⁴⁴. A área de tecnologia ainda possui número reduzido de mulheres⁴⁵, cerca de vinte por cento do

⁴² “Vigilância epistêmica’ – o fato de as pessoas serem muito mais exigentes com argumentos proferidos por outros indivíduos do que com as próprias crenças, mesmo que elas sejam mal justificadas.” (HORTA, 2019, p. 39)

⁴³ Art. 20. A composição de equipes para pesquisa, desenvolvimento e implantação das soluções computacionais que se utilizem de Inteligência Artificial será orientada pela busca da diversidade em seu mais amplo espectro, incluindo gênero, raça, etnia, cor, orientação sexual, pessoas com deficiência, geração e demais características individuais.

⁴⁴ “De acordo com a pesquisa, com base em 2008, 59% do contingente de 845.479 profissionais do setor são de homens e mulheres brancos – 37% e 22%, respectivamente. Já os negros, pardos e indígenas são responsáveis por 30% dessa força de trabalho, sendo 11% mulheres e 19% homens. Considerando que, segundo o IBGE, o Brasil possui 56,10% de pessoas que se declaram negras ou pardas, a desproporção de acesso ao mercado de trabalho por este grupo indica que há muito a ser feito”. Forbes. **Setor de tecnologia brasileiro precisa protagonizar a luta antirracista**. Disponível em <https://forbes.com.br/forbes-tech/especial-inovadores-negros/2020/07/presenca-de-negros-em-tecnologia-ainda-precisa-avancar/>. Acesso em 11 de abril de 2022.

⁴⁵ Associação das Empresas de Tecnologia da Informação e Comunicação (TIC) e de Tecnologias Digitais. **Mulheres na tecnologia: ainda precisamos bater nessa tecla!** Disponível em <https://brasscom.org.br/mulheres-na-tecnologia-ainda-precisamos-bater-nessa-tecla/>. Acesso em 11 de abril de 2022.

setor. Se for analisada a necessidade de interdisciplinaridade na formação das equipes, provavelmente ocorrerá a mesma dificuldade de diversidade dos membros que tenham formação jurídica ou em outra área do conhecimento⁴⁶.

“Meu argumento é que oceanos de dados comportamentais, nos próximos anos, irão sustentar diretamente os sistemas de inteligência artificial. E estes permanecerão como caixas pretas aos olhos humanos. Ao longo desse processo, raramente saberemos a qual tribo “pertencemos” ou o porquê. Na era da inteligência de máquina, a maioria das variáveis permanecerão um mistério. (...) Se não tomarmos de volta um certo controle, essas futuras ADMs parecerão misteriosas e poderosas. Farão conosco o que quiserem, e mal saberemos o que está acontecendo” (O’NEIL, 2020, p. p. 194)

Enquanto não houver limites e responsabilização, as fragilidades e vulnerabilidades humanas serão grande fonte de lucro. O ser humano, mesmo imperfeito, pode evoluir e decidir melhor, já a máquina não decide melhor sozinha, permanece parada no tempo, alterando-se apenas se um ser humano intervier. A sociedade perde-se nesta nova revolução industrial, mas pode tomar como referência a revolução industrial anterior. O progresso não pode ser dar a custas de vidas humanas, não é o livre mercado que por conta própria adotará limites. (O’NEIL, 2020, p. 223-229)

Primeiro, cientistas de dados precisam compreender as consequências de sua atividade se não for bem conduzida, adotando valores, uma base filosófica. Aos que não conseguirem por conta própria adotar um comportamento autorregulado, a lei precisará adequar-se para responsabilizá-los. É necessário que os sistemas sejam permeados por valores humanos, mesmo que se reduza a eficiência. A matemática deve ser uma ferramenta, não uma opressora. Necessário também análises constantes de impacto dos sistemas e de auditorias. Para sistemas injustos, a única solução é o descarte. Devem ser adotados ciclos positivos de feedback. É necessário verificar os sistemas também para encontrar vieses e eliminá-los. Existem iniciativas que criam robôs que estudam a forma que são tratados para atingir essa finalidade. Gigantes da tecnologia não poderiam conduzir suas próprias auditorias pois estariam contaminadas com os próprios vieses. Nesse ponto o governo precisará regular o setor para evitar que mais excessos e injustiças ocorram assim como foi necessário na última revolução industrial. A legislação europeia, principalmente a relativa a

⁴⁶ Agência Brasil. Cresce total de negros em universidades, mas acesso é desigual. Disponível em <https://agenciabrasil.ebc.com.br/geral/noticia/2020-11/cresce-total-de-negros-em-universidades-mas-acesso-e-desigual>. Acesso em 11 de abril de 2022.

proteção de dados, pode auxiliar os governos oferecendo uma estratégia inicial para a criação de legislações do setor. Modelos precisam ser construídos com o objetivo de auxiliar pessoas, mas muitos não serão bem-vistos pois podem confrontar o *status quo*. Portanto, os modelos são resultados de escolhas, e essas escolhas “são fundamentalmente morais” (O’NEIL, 2020, p. 229-244)

5. CONSIDERAÇÕES FINAIS

O presente trabalho buscou contribuir com o debate sobre a Inteligência Artificial aplicada ao Direito, e sobre os impactos dos vieses na Inteligência Artificial aplicada em matéria penal.

Para isso, buscou-se analisar os impactos da inteligência artificial em matéria penal, compreendendo a Inteligência Artificial, os vieses e os impactos, possibilidades e regulação da IA.

Não seria possível, portanto, a utilização da máquina em decisões judiciais em matéria penal em razão da não neutralidade da Inteligência Artificial e das dificuldades de identificação e correção de vieses.

Até mesmo em casos em que a máquina sirva como auxiliar da decisão, é necessário refletir sobre como ela será utilizada, para proteção da dignidade da pessoa humana e princípios constitucionais, como o devido processo legal, e até mesmo para reduzir a possibilidade de vieses, como, por exemplo, o de confirmação.

É necessário estabelecer regulação sobre o tema após amplo debate com especialistas, principalmente por ser necessário ao legislador e à sociedade civil melhor compreensão sobre o que é e como a inteligência artificial funciona.

Os vieses decisórios, da máquina ou do homem, despertam ainda mais a necessidade de mudança de comportamento dos juristas. É necessário buscar maior conhecimento sobre as dificuldades que o cérebro possui para decidir e de que forma essas dificuldades podem ser mitigadas na atuação profissional e nas decisões judiciais.

A IA aplicada ao Direito, como exposto na literatura, possui o condão de alterar significativamente a realidade, mas é necessário cautela, regulação e diretrizes éticas que possibilitem que essas alterações sejam mais justas e respeitem direitos humanos.

Existe grande desinformação sobre a IA por parte dos juristas e legisladores.

Lucra-se com a desinformação até que se regule, controle e direcione atividades, assim como ocorreu no início da revolução industrial. Percebe-se que a IA é classificada como uma nova revolução, e, se observarmos a história, o mercado se valeu da desinformação para maximizar lucros, até que se tomasse consciência dos abusos e houvesse maior regulação do Estado. Já está mais do que demonstrado que o mercado não consegue se autorregular sobre o tema. Não se pode permitir que eficiência e lucro desmedido, principalmente, às custas da dignidade da pessoa humana, dirijam todo o processo.

Urge a necessidade de melhor regulação e responsabilização no uso da IA. E diante de todas as experiências recentes, que colocam em xeque seu uso principalmente em razão da ética, somando-se ao fato de que a IA mantém o *status quo*, principalmente em um contexto de segregação racial no qual funciona o sistema prisional brasileiro, com a quarta população carcerária do mundo, torna-se urgente a regulação, controle e responsabilização no uso da IA para que não se repita no Brasil experiências como as descritas na literatura em contextos semelhantes, como nos Estados Unidos. No atual estado da arte, é necessário impedir a aplicação da IA em matéria penal limitando-se o uso à gestão processual sob o risco de intensificarmos preconceitos sociais em razão dos vieses humanos e algorítmicos.

A aplicação da IA no Judiciário Brasileiro precisa pautar-se nos direitos humanos, centrar-se na dignidade da pessoa humana e buscar o maior dos princípios ao qual o Direito serve, a Justiça. A resolução 332/2020 do CNJ representa importante marco para atingir esta finalidade, trazendo orientações sobre mitigação de vieses e até mesmo o abandono de projetos em razão de erros que não possam ser solucionados. Diante dos problemas e dificuldades descritos pela literatura sobre os vieses, principalmente em decisões judiciais, essencial a não aplicação da IA em matéria penal, conforme preceitua a Resolução.

REFERÊNCIAS BIBLIOGRÁFICAS

ARBIX, Glauco. **A transparência no centro da construção de uma IA ética**. Novos estudos CEBRAP [online]. 2020, v. 39, n. 2, pp. 395-413. Disponível em

<https://www.scielo.br/j/nec/a/pD9k5gtHpXwsgFcsMC5gbJg/abstract/?format=html&lang=pt#>. Acesso em 25 de julho de 2021.

CNJ. Resolução 332/2020 do Conselho Nacional de Justiça. Disponível em <https://atos.cnj.jus.br/atos/detalhar/3429>. Acesso em 5 de setembro de 2020.

COSTA, A. A. e HORTA, R.L. **Vieses na decisão judicial e desenho institucional: uma discussão necessária na era da pós-verdade.** Cadernos Adenauer, XVIII, nº1, s/l, s/e, 2017, p.11-34. Disponível em <https://www.kas.de/de/statische-inhalte-detail/-/content/cadernos-adenauer>. Acesso em 12 de agosto de 2021.

DOMINGOS, Pedro. O algoritmo mestre: como a busca pelo algoritmo de machine learning definitivo recriará nosso mundo. Novatec Editora; 1ª edição. São Paulo, 2017.

FEET, N.R.; MEDEIROS, F.V. **Uso da Inteligência Artificial no processo de tomada de decisões jurisdicionais: potenciais riscos e possíveis consequências.** In: Inteligência Artificial e Direito Processual: os impactos da virada tecnológica no Direito Processual. Coordenadores: Dierle Nunes, Paulo Henrique dos Santos Lucon, Erik Navarro Wolkart - Salvador: JusPodivm, 2021, p. 749-786.

FGV. **Nota Técnica Sobre o Projeto de Lei 21/2020.** Centro de Inovação, Administração e Pesquisa do Judiciário. Fundação Getúlio Vargas, 2021. Disponível em <https://ciapj.fgv.br/publicacoes>. Acesso em 13 de março de 2022.

FGV. **Relatório de Pesquisa: Tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário Brasileiro.** Fundação Getúlio Vargas, 2020. Disponível em <https://ciapj.fgv.br/publicacoes>. Acesso em 13 de março de 2022.

FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA. Anuário Brasileiro de Segurança Pública: 2021. São Paulo: FBSP, 2021. Disponível em <https://forumseguranca.org.br/anuario-brasileiro-seguranca-publica/>. Acesso em 25 de março de 2022.

FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA. Anuário Brasileiro de Segurança Pública: 2020. São Paulo: FBSP, 2020. Disponível em <https://forumseguranca.org.br/anuario-14/>. Acesso em 25 de março de 2022.

GARCIA, Ana Cristina Bicharra. **Ética e Inteligência Artificial.** In Computação – Revista da Sociedade Brasileira de Computação – Nov. 2020, n 43, pag 14-22. Disponível em <https://sol.sbc.org.br/journals/index.php/comp-br/article/view/1791>. Acesso em 8 de março de 2022.

HORTA, Ricardo de Lins e. **Direito e tomada de decisão: elementos para uma teoria da decisão jurídica**. Brasília, 2019. 177 p. Tese (Doutorado em Direito – UnB). PósGraduação em Direito da UnB.

KAUFMAN, Dora. **Dossiê: Deep learning: a Inteligência Artificial que domina a vida do século XXI**. *Teccogs: Revista Digital de Tecnologias Cognitivas*, TIDD | PUC-SP, São Paulo, n. 17, p. 17-30, jan-jun. 2018. Disponível em <https://revistas.pucsp.br/index.php/teccogs/article/view/48585>. Acesso em 14 de fevereiro de 2022.

KAUFMAN, Dora. **Inteligência Artificial: caminhos para proteger a sociedade**. *Revista Interesse Nacional*, Edição 56, dez 2021. Disponível em <https://interessenacional.com.br/2021/12/21/inteligencia-artificial-caminhos-para-protoger-a-sociedade/>. Acesso em 15 de fevereiro de 2022.

MAIA FILHO, M. S.; JUNQUILHO, T. A. **Projeto Victor: perspectivas de aplicação da inteligência artificial ao direito**. *Revista de Direitos e Garantias Fundamentais*, v. 19, n. 3, p. 218-237, 29 dez. 2018. Disponível em <https://sisbib.emnuvens.com.br/direitosegarantias/article/view/1587>. Acesso em 3 de fevereiro de 2021.

NUNES, P. **Novas ferramentas, velhas práticas: reconhecimento facial e policiamento no Brasil. Retratos da violência: cinco meses de monitoramento, análise e descobertas (Rede de Observatório de Segurança) - 2019**. Disponível em <http://observatorioseguranca.com.br/wp-content/uploads/2019/11/1relatoriorede.pdf>. Acesso em 18 de abril de 2022.

NUNES, D; LUCON, P.H.S.; WOLKART, E.N. **Decisão judicial e inteligência Artificial: é possível a automação da decisão?** In: *Inteligência Artificial e Direito Processual: os impactos da virada tecnológica no Direito Processual*. Coordenadores: Dierle Nunes, Paulo Henrique dos Santos Lucon, Erik Navarro Wolkart - Salvador: JusPodivm, 2021, p. 703-748.

NUNES, D; MARQUES, A.L.P.C. **Inteligência Artificial e Direito Processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas**. *Revista de Processo*, vol. 285/2018, Nov / 2018. Disponível em <https://bit.ly/3p7ntRm>. Acesso em 4 de abril de 2021.

O'NEIL, Cathy. **Algoritmos de destruição em massa: como o big data aumenta a desigualdade e ameaça a democracia**. 1ª edição. Ed.: Rua do Sabão, São Paulo, 2020.

PEDRINA, Gustavo M. L. **Consequências e perspectivas da aplicação de inteligência artificial a casos penais**. *Revista Brasileira de Direito Processual Penal*, Porto Alegre, vol. 5, n. 3, p. 1589-1606, set./dez. 2019.

<https://doi.org/10.22197/rbdpp.v5i3.265>

Relatório de Pesquisa Tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário Brasileiro – FGV. Disponível em <https://ciapj.fgv.br/publicacoes>. Acesso em 18 de março de 2022.

RUBACK, L; AVILA, S.; CANTERO, L. **Vieses no Aprendizado de Máquina e suas Implicações Sociais: Um Estudo de Caso no Reconhecimento Facial.** In: WORKSHOP SOBRE AS IMPLICAÇÕES DA COMPUTAÇÃO NA SOCIEDADE (WICS), 2. , 2021, Evento Online. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2021. p. 90-101.

SOARES, M. N., KAUFFMAN, M. E., & CHAO, K.-M. (2020). **Inteligência Artificial: impactos no Direito e na advocacia.** *Direito Público*, 17(93). Disponível em <https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/3555>. Acesso em 13 de janeiro de 2021.

SULOCKI, Victoria de. **Novas Tecnologias, velhas discriminações: ou da falta de reflexão sobre o sistema de algoritmos na Justiça Criminal.** In: *Inteligência artificial e direito: ética, regulação e responsabilidade*; coordenação Ana Frazão e Catilin Mulholland. São Paulo: Thomson Reuters Brasil, 2019, p. 651.

Schwab, Klaus. **A quarta revolução industrial.** Tradução Daniel Moreira Miranda. - São Paulo: Edipro, 2016.