

**REDE DE ENSINO DOCTUM
CURSO DE DIREITO
UNIDADE DE MANHUAÇU**

**Dalmo Werner Lopes
Gabriel Pereira Lopes
Wallace de Souza Silveira Lage
Winissan Felipe de Souza**

**AS PROBLEMÁTICAS JURÍDICAS NA UTILIZAÇÃO DA INTELIGÊNCIA
ARTIFICIAL (IA) NA TOMADA DE DECISÕES DO STF**

Manhuaçu/MG

2024

**REDE DE ENSINO DOCTUM
CURSO DE DIREITO
UNIDADE DE MANHUAÇU**

**Dalmo Werner Lopes
Gabriel Pereira Lopes
Wallace de Souza Silveira Lage
Winissan Felipe de Souza**

**AS PROBLEMÁTICAS JURÍDICAS NA UTILIZAÇÃO DA INTELIGÊNCIA
ARTIFICIAL (IA) NA TOMADA DE DECISÕES DO STF**

Trabalho de conclusão de curso apresentado à Rede Doctum de Ensino na Unidade de Manhuaçu/MG, como requisito parcial à obtenção do grau de bacharel em Direito.

Prof. Supervisor: Prof^a. Soraya Cesar Sanglard

Manhuaçu/MG

2024

RESUMO

Este artigo trata das implicações jurídicas e éticas do uso da Inteligência Artificial (IA) pelo Supremo Tribunal Federal (STF), com ênfase no Programa Victor. O estudo explora os desafios da automação judicial, especialmente quanto ao Art. 93, Inciso IX da Constituição Federal, que exige decisões fundamentadas. Com base em revisão de literatura, identifica-se três áreas críticas: responsabilidade nas decisões assistidas por IA, transparência dos algoritmos e mitigação de vieses e discriminações. A pesquisa aponta a necessidade de supervisão humana e discute dificuldades na intervenção eficaz, propondo soluções como painéis multidisciplinares e sistemas híbridos de decisão. Conclui-se que, embora a IA possa aprimorar a eficiência judicial, o uso responsável exige auditorias robustas e mecanismos de explicabilidade para assegurar decisões justas e transparentes, resguardando os direitos fundamentais.

Palavras-chave: Inteligência Artificial (IA); fundamentação das decisões; vieses algorítmicos; Direito Digital.

SUMÁRIO

1. INTRODUÇÃO	4
2. PROBLEMAS ÉTICOS E JURÍDICOS NA UTILIZAÇÃO DA INTELIGÊNCIA ARTIFICIAL PELO JUDICIÁRIO	5
2.1. Breve histórico da IA.....	5
2.2. Os problemas éticos e sociais	5
2.3. Problemáticas jurídicas	7
3. TRANSPARÊNCIA, RESPONSABILIDADE, <i>ACCOUNTABILITY</i>, <i>AUDITABILITY</i> E VIÉSES DISCRIMINATÓRIOS DOS ALGORITMOS DE IA NO STF	10
3.1. A transparência dos algoritmos de IA	10
3.2. Da responsabilidade, <i>accountability</i> e <i>auditability</i> nas decisões algorítmicas	12
3.3. Vieses e discriminações nos algoritmos de IA	13
4. A MITIGAÇÃO DOS DESVIOS E TRANSPARÊNCIA NAS FERRAMENTAS DE IA UTILIZADAS PELO JUDICIÁRIO	15
4.1. Relatórios de Impacto Algorítmico	15
4.2. Painéis Multidisciplinares.....	16
4.3. Sistemas Híbridos De Decisão.....	17
5. CONSIDERAÇÕES FINAIS.....	19
REFERÊNCIAS BIBLIOGRÁFICAS	20

1. INTRODUÇÃO

A introdução progressiva da Inteligência Artificial (IA) no Poder Judiciário brasileiro, especialmente no Supremo Tribunal Federal (STF), vem fomentando um debate profundo, mas necessário sobre os impactos sociais, éticos e jurídicos que envolvem essa tecnologia. Entre as iniciativas recentes desenvolvidas em colaboração com a Universidade de Brasília, com o propósito de otimizar a triagem de processos e a análise de repercussão geral, destaca-se o Programa Victor. Esse fenômeno que já alcança o Judiciário, faz parte do que tem se chamado de a Quarta Revolução Industrial que é marcado pela integração da IA em diversas esferas do conhecimento humano e aplicações cotidianas, simples ou complexas, o que inclui o Poder Judiciário. Entretanto, a aplicação de IA no Judiciário levanta questionamentos críticos e por meio de revisão bibliográfica, foram identificadas três áreas críticas: a responsabilidade por decisões assistidas por algoritmos, a transparência dos sistemas empregados e os vieses potenciais que podem comprometer direitos fundamentais dos cidadãos.

Importante destacar que a Constituição Federal de 1988 em seu Art. 93, inciso IX, estabelece a fundamentação das decisões judiciais como imperativo, assegurando, assim, a individualização, publicidade e a transparência em todas as decisões. Com a introdução da IA, impõem-se desafios específicos acerca de como garantir que os algoritmos possam justificar, de maneira clara e acessível, os fundamentos das decisões. Esta premissa constitucional revela a necessidade de se examinar as complexidades envolvidas na transparência e na supervisão humana das decisões automatizadas, considerando a possibilidade de perpetuação de vieses e discriminações na análise de dados.

Assim, o presente artigo se propõe a examinar os impactos centrais do uso de IA no Judiciário, com especial enfoque nas questões éticas e jurídicas, na transparência dos algoritmos e nas medidas essenciais para a mitigação de vieses e desigualdades. A partir de uma análise do Programa Victor, buscam-se apontamentos para que a IA contribua para um Judiciário mais célere e eficiente, sem comprometer a legitimidade e a imparcialidade das decisões judiciais.

2. PROBLEMAS ÉTICOS E JURÍDICOS NA UTILIZAÇÃO DA INTELIGÊNCIA ARTIFICIAL PELO JUDICIÁRIO

2.1. BREVE HISTÓRICO DA IA

A Inteligência Artificial (IA) tem suas raízes na década de 1950, com os primeiros trabalhos de Alan Turing e John McCarthy, que cunhou o termo "Inteligência Artificial". Desde então, a IA evoluiu significativamente, passando de programas simples capazes de resolver problemas específicos para sistemas complexos que utilizam aprendizado de máquina e redes neurais profundas. Inicialmente, a IA era restrita a áreas técnicas e científicas, mas com o avanço tecnológico, seu uso se expandiu para diversos setores, incluindo o jurídico, podendo ser considerada uma tecnologia multidisciplinar e, recentemente, essa multidisciplinaridade tem ganhado destaque – o Aprendizado de Máquina. Este campo possibilita a “educação” da máquina por meio de exemplos relacionados ao problema que se busca solucionar, minimizando a necessidade de uma programação mais extensa (LAGE, 2022).

No contexto jurídico, desde 2016 a IA começou a ser adotada de forma mais robusta, quando a redução nos custos de armazenamento e o aumento da capacidade de processamento de dados permitiram a criação de sistemas mais eficientes. Esses sistemas são capazes de analisar grandes volumes de dados, prever resultados de casos e auxiliar na redação de decisões judiciais. Nessa esteira, o Programa Victor foi pensado pelo STF – que em parceria com a Universidade de Brasília (UnB) – com o objetivo de agilizar a classificação de ações judiciais e identificar precedentes, representando um marco significativo na utilização da IA pelo poder judiciário brasileiro.

O uso de Inteligência Artificial (IA) pelo Judiciário apresenta sérias questões sociais, éticas e principalmente, jurídicas. Embora possa haver vantagens no uso de IA que, sem dúvidas, leva a um aumento da eficiência e da celeridade dos processos judiciais, em um escopo mais amplo, há preocupações fundamentais quanto sobre a justiça, a moralidade e a confiança pública.

2.2. OS PROBLEMAS ÉTICOS E SOCIAIS

O uso de IA no direito — especialmente à medida que a tecnologia avança e começa a assumir funções críticas típicas da jurisdição — tem gerado uma série de debates éticos e sociais. É verdade que o escopo principal do Projeto Victor do STF está na automação do processamento de recursos de repercussão geral por meio do uso de algoritmos de IA para

auxiliar juízes na tomada de decisões. Mas, mais importante, ele questiona como tal tecnologia, que é essencialmente baseada em dados e algoritmos, lidará com questões relacionadas à moralidade, justiça e cultura humana que seu funcionamento nunca pode replicar completamente.

Não restam dúvidas de que os impactos éticos e sociais são de suma importância, pois afetam não apenas a legitimidade, mas até mesmo a confiança pública nas decisões judiciais.

Contudo, por mais que a IA seja um produto dos dados que recebe e dos algoritmos que processam essas informações, sua neutralidade é tão boa, isenta e imparcial quanto a qualidade isenção e imparcialidade dos dados que ela foi treinada. Seja como for, se os dados históricos forem tendenciosos em quaisquer termos — raça, gênero ou status socioeconômicos — a IA inevitavelmente perpetuará o mesmo viés. Citada por Maia Filho e Junquilha (2018), Cathy O'Neil denuncia em seu argumento que tais sistemas podem se tornar verdadeiras "armas de destruição matemática", reforçando desigualdades em vez de resolvê-las.

Cathy O'Neil (2016) denuncia a formação de vieses que podem ocorrer de forma intencional ou não intencional na programação. Isso porque, quando os dados utilizados para a elaboração da tecnologia são processados “fritamente” por matemáticos e cientistas da computação – sem o devido tratamento que vise a eliminar preconceitos e atribuir transparência aos algoritmos - pode-se chegar a medidas de inovação que perpetuam a opressão social. Isso é, que têm potencial para se tornar verdadeiras “armas de destruição matemática” (expressão que dá título ao livro da autora)” (MAIA FILHO E JUNQUILHO, 2018, p. 234).

Há também a possibilidade de que a automação possa tornar o processo de decisão judicial desumanizado, pois não há como uma IA capturar a intuição, a humanidade e o conhecimento das sensibilidades culturais que um juiz individual reflete em uma decisão. Os algoritmos desenvolvidos no Projeto Victor, conforme enfatizado por Maia Filho e Junquilha (2018), são direcionados a aumentar a eficácia do STF, por meio do qual os humanos devem garantir que as decisões propostas pela IA sejam validadas e interpretadas. Tal percepção pode questionar a credibilidade de todo o judiciário, uma vez que tais decisões frias e desumanas vão contra a realidade social – a própria ideia sobre a qual repousa a legitimidade de um sistema judicial.

Esses desafios éticos e sociais levantam questões críticas que deverão ser exploradas com maior profundidade. Como garantir que os algoritmos usados pela IA no Judiciário sejam treinados em dados imparciais e livres de preconceitos e vieses? Quais mecanismos podem ser implementados para garantir que a supervisão e responsabilidade humana continue a desempenhar um papel central, mesmo em um sistema cada vez mais automatizado? E,

finalmente, como a sociedade pode confiar em um sistema onde as decisões, que afetam direitos fundamentais tão caros à sociedade e à democracia, são tomadas ou influenciadas por máquinas? Esses são alguns dos problemas que precisam ser analisados, discutidos e solucionados à medida que o uso da IA se expande no Judiciário brasileiro.

2.3. PROBLEMÁTICAS JURÍDICAS

Em relação às questões que refletem no sistema de justiça, as falhas jurídicas cometidas por meio da prolação de decisões por uma máquina, podem, de fato, ter consequências graves. A Constituição Brasileira, em seu art. 5º, LXXV, prevê que "o Estado indenizará o condenado por erro judiciário, assim como o que ficar preso além do tempo fixado na sentença". Julgamentos incorretos proferidos por AI podem perpetuar erros, pelos quais a indenização e outros danos acumulam-se à reputação negativa da instituição dentro do judiciário.

Além disso, a Constituição Brasileira também estabelece no Art. 93, IX, que "todos os julgamentos dos órgãos do Judiciário serão públicos, e todas as decisões serão fundamentadas, sob pena de nulidade". A transparência e a fundamentação das decisões são elementos fundamentais na construção de qualquer sistema que vise a trazer justiça e legalidade a uma sentença. Um sistema de IA não pode, de forma alguma, fundamentar sua decisão de forma compreensível e detalhada, infringindo, assim, o requisito constitucional de justificação e tornando suas decisões nulas.

De acordo com a Seção II - Dos Elementos e dos Efeitos da Sentença do Código de Processo Civil de 2015, a sentença deve conter elementos essenciais, como o relatório, os fundamentos e o dispositivo. A decisão deve ser fundamentada, enfrentando todos os argumentos deduzidos no processo e explicando claramente a aplicação dos conceitos jurídicos. A IA pode não cumprir isso, pois limita sua atividade à reprodução de atos normativos ou à invocação de precedentes sem entrar em detalhes na análise exigida por essas alegações; portanto, suas decisões ficam aquém das regras legais e, conseqüentemente, são impugnáveis.

As decisões judiciais são baseadas, influenciadas e abalizadas nos valores morais, culturais e costumeiros de uma sociedade. Esses fatos podem ser lidos por juízes humanos que então os interpretam para trazer à tona uma sensibilidade contextual em seus processos de tomada de decisão que a IA, constituída apenas por dados e algoritmos matemáticos, nunca alcançará. Portanto, por exemplo, um algoritmo, que é dependente de precedentes legais anteriores, não será capaz de reconhecer mudanças delicadas ocorrendo na aplicação da lei ou da justiça social de maneiras que um juiz humano poderia captar mais instintivamente. Morais

(2021) destaca que as ferramentas de IA, ao serem aplicadas no campo da justiça, podem acabar restringindo direitos fundamentais por não compreenderem as condições performativas e morais de cada caso específico.

Nesse aspecto, o julgador assume o dever moral de estar consciente sobre as consequências da sua decisão em outro sujeito moral. Ou seja, a identificação e classificação de um recurso como de repercussão geral pelo programa computacional não são capazes de entender as condições performativas decorrentes da sua ação específica, o que pode ser a restrição de um direito fundamental (MORAIS, 2021, p. 322).

Além disso, como já dito, os algoritmos são tão imparciais quanto os dados nos quais foram treinados. Se esses bancos de dados de treinamento da IA forem viciados com vieses históricos, como discriminações raciais, de gênero ou socioeconômicas, certamente essa tendência será perpetuada nas decisões proferidas pela IA. Um dos exemplos mais famosos foi o sistema COMPAS, nos Estados Unidos, que mostrou vieses significativos contra réus negros, classificando-os com maior probabilidade de reincidência em comparação com réus brancos – perpetrando assim a discriminação histórica e latente contra a comunidade negra.

Como exemplo paradigmático, cita-se o episódio envolvendo o emprego do sistema *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS), usado para o cálculo do risco de reincidência de uma pessoa presa. Estruturado na análise de 137 fatores de uma base de dados composta da realidade de informações de pessoas previamente presas, o referido sistema acabou por estabelecer índices de periculosidade mais altos em desfavor de presos negros do que em relação a presos de pele branca (PASSOS, 2022, p. 111, apud REICHELDT, 2021).

Outra questão de extrema importância é a confiança pública. A legitimidade de uma decisão judicial pressupõe um nível de confiança por parte das pessoas em relação ao sistema judicial. Com as máquinas tomando a decisão, a sociedade pode perder a confiança em tais julgamentos. A IA não tem a capacidade de ter empatia, fazer julgamentos morais ou mesmo entender as complexidades humanas contra os desafios em casos judicialmente desafiadores e carregados de emoção e complexidade, sendo essa capacidade valorativa algo inerente apenas aos seres humanos, como argumenta Hugo Passos:

[...] o julgamento, a valoração de algo como bom (juízo de valor) ou como devido/correto (juízo de dever) são atividades inerentes ao ser humano, pois, ao avaliarem a realidade e estimarem o que é bom e/ ou devido, conduzem as escolhas individuais e sociais, guiam os indivíduos e as ações estatais. Em outras palavras, delimitam a vida humana tanto individual quanto socialmente. (PASSOS, 2022, p. 86).

Essa falta de empatia e juízo humano pode corroer a aceitação e a legitimação de decisões proferidas por sistemas automatizados. Influenciados, portanto, pela moral, cultura e costumes da sociedade, somente os magistrados humanos podem ter uma perspectiva contextual, sensíveis à esses valores sociais e culturais, refletidos em suas decisões, o que está muito além do alcance de qualquer inteligência artificial.

Como já argumentado, a IA funciona com dados e algoritmos — matemática pura — e, portanto, decisões que, embora tecnicamente corretas, podem não representar a justiça social ou os valores culturais de uma sociedade. Nesse sentido, o juiz assume dever moral com as consequências que suas decisões têm sobre outro sujeito moral, não apenas em uma decisão como uma sentença ou um acórdão, mas até mesmo para identificar e classificar recursos de repercussão geral que hoje já são feitos pelo programa Victor do STF. Tampouco, nessas decisões de aplicação de sentenças de repercussão geral, o programa de computador consegue compreender as condições performativas decorrentes de sua ação específica, o que pode levar à restrição de um direito fundamental - como argumenta Moraes (2021, p. 322):

Assim, essa crescente implantação de inteligência artificial no Judiciário levanta questões complexas de ética e de justiça. Enquanto, por um lado, a IA promete procedimentos judiciais mais eficientes e rápidos ao facilitar a triagem e análise de casos, por outro lado, suas operações de algoritmos matemáticos e baseados em dados levantam preocupações de imparcialidade, transparência e justiça social. A adoção de IA nas decisões judiciais acentua o problema do viés inerente aos dados; é também um processo de tomada de decisão desumana, que muito provavelmente pode não compreender totalmente as complexidades culturais e morais da sociedade. Esses problemas revelam a necessidade de mecanismos contínuos de vigilância e controle humano para garantir justiça, transparência e conformidade com os princípios constitucionais em decisões judiciais.

3. TRANSPARÊNCIA, RESPONSABILIDADE, *ACCOUNTABILITY*¹, *AUDITABILITY*² E VIÉSES DISCRIMINATÓRIOS DOS ALGORITMOS DE IA NO STF

3.1. A TRANSPARÊNCIA DOS ALGORÍTMOS DE IA

O uso da inteligência artificial no STF tem se mostrado promissor na otimização dos processos judiciais, com o uso do Programa Victor. No entanto, um dos maiores desafios é garantir a transparência dos algoritmos envolvidos. A opacidade desses sistemas, especialmente em algoritmos complexos de *deep learning*³, pode minar a confiança pública e comprometer a legitimidade das decisões assistidas por IA. Como observado por Débora Bonat e Fabiano Peixoto (2020):

Os algoritmos internalizam dados de maneiras que criam dificuldade para uma auditabilidade e compreensão precisas. Tal dificuldade afeta princípios e propósitos para uma proposta de IA robusta e deve ser conscientemente observada. Essa forma de internalização a partir de referenciais é influenciada por seres humanos falíveis, que podem facilmente gerar também vieses na ação produzida pela aplicação do algoritmo, desde ferramentas de resolução de disputas, algoritmos de coleta de dados até discriminações em razão de raças ou religiões. (BONAT E PEIXOTO, 2020, p. 60, apud TRITES, 2019).

Para assegurar uma aplicação justa e responsável dessas ferramentas, é crucial que os algoritmos sejam acessíveis e explicáveis.

A transparência nos algoritmos utilizados pelo STF deve ser vista como um requisito fundamental para assegurar que as decisões tomadas ou assistidas por IA sejam passíveis de escrutínio e controle. Conforme mencionado no Projeto Victor, a IA “não decide ou julga; ela está sendo treinada tão somente para atuar na organização dos processos” (BRAGANÇA, 2019, p. 71). Essa distinção é essencial, uma vez que o papel do algoritmo é apenas auxiliar a celeridade processual, enquanto a decisão judicial permanece sob controle humano. No entanto,

¹ Não existe uma palavra em português que traduz *accountability*, mas sim uma dezena de termos que tentam conceituá-la, tais como controle, fiscalização, responsabilização, prestação de contas, compromisso, proatividade e transparência. Disponível em: <https://clp.org.br/o-que-e-accountability>. Acesso em: 15 set 2024.

² Em outras palavras, auditabilidade diz respeito a qualquer informação que possa ser utilizada para descrever os dados e processos, como por exemplo: a natureza dos dados, o que inclui sua origem, destino e as dependências associadas a eles; as etapas de processamento; e outras informações contextuais, como aspectos de proteção de dados, configuração do sistema, ações da equipe, etc (Singh et al., 2018). Disponível em: <https://booksol.sbc.org.br/index.php/sbc/catalog/download/110/497/770-1>. Acesso em: 15 set 2024.

³ Subconjunto do aprendizado de máquina que usa redes neurais de várias camadas, chamadas de redes neurais profundas, para simular o complexo poder de tomada de decisão do cérebro humano. Alguma forma de *deep learning* alimenta a maioria das aplicações de inteligência artificial (IA) em nossas vidas atualmente. Disponível em: <https://www.ibm.com/br-pt/topics/deep-learning>. Acesso em: 15 set. 2024.

há uma crescente preocupação com a falta de explicabilidade, o que pode comprometer a confiança pública e a responsabilização dos operadores da IA.

A falta de transparência pode ser abordada por meio de auditorias algorítmicas, que permitem a revisão e o escrutínio dos sistemas. Bonat e Peixoto (2020) destacam que um dos principais desafios relacionados à IA é a complexidade dos algoritmos, especialmente em redes de *deep learning*, onde “milhares de neurônios artificiais trabalham de forma difusa, formando a chamada caixa-preta” (p. 61). A dificuldade na visualização e compreensão dessas estruturas matemáticas aumenta a percepção de falta de transparência, dificultando a auditabilidade do processo de decisão.

A transparência dos algoritmos inclui a divulgação dos dados utilizados para treinar os modelos de IA e os critérios empregados para a tomada de decisões. Isso é crucial para evitar a perpetuação de vieses e discriminações presentes nos dados históricos que alimentam esses sistemas. Além disso, relembramos que a Constituição Brasileira, em seu art. 93, inciso IX, exige que todas as decisões judiciais sejam fundamentadas, implicando que as decisões algorítmicas também devem ser explicáveis de maneira compreensível, atendendo às exigências legais de transparência e justificativa.

Para além disso, a revisão humana do controle operacional sobre ferramentas é crucial dentro do Poder Judiciário. É o que afirma Moraes (2021), ao enfatizar que a Resolução nº 332/2020 do Conselho Nacional de Justiça (CNJ) é uma medida fundamental para assegurar o direito de supervisão humana para todos que possam ser implicados pelo uso da tecnologia (CONSELHO NACIONAL DE JUSTIÇA, 2020). Sem prejuízo de que nem todas as decisões automatizadas são revistas por intervenção humana, deve haver um espaço institucional para questionar e corrigir um erro, criando uma instância revisora em caso de impugnação por parte dos interessados.

A explicabilidade e a transparência são diretrizes essenciais para o uso ético da IA no Judiciário, como destacado por Moraes (2021). Elas garantem que, quanto maior o impacto sobre os direitos fundamentais, mais clara deve ser a explicação sobre o funcionamento do algoritmo. A supervisão humana também é indispensável para permitir questionamentos em casos de erro, assegurando um espaço de revisão institucional. Essas medidas são inspiradas em princípios de sopesamento ético, visando maior proteção dos direitos fundamentais.

Para garantir a compreensão dos algoritmos utilizados no STF, é necessário adotar medidas como auditorias públicas e relatórios de impacto. BRAGANÇA (2019) menciona que os dados devem ser curados com cuidado para evitar vieses que comprometam a justiça das decisões. O estudo do Projeto Victor também aponta a importância de interfaces acessíveis, que

forneçam relatórios detalhados sobre o funcionamento do algoritmo e tornem as decisões mais transparentes, fortalecendo o princípio da publicidade dos atos processuais.

3.2. DA RESPONSABILIDADE, ACCOUNTABILITY E AUDITABILITY NAS DECISÕES ALGORÍTMICAS

Os desenvolvedores de sistemas de IA têm grande responsabilidade em garantir que os algoritmos sejam justos, transparentes e isentos de viés. Tais compromissos incluem escolher os dados de treinamento cuidadosamente, instituir técnicas robustas para detecção e mitigação de viés e executar testes extensivos na operação do sistema garantindo uma operação conforme o esperado. Dada a complexidade de aviso em algoritmos de IA encontrar erros e vieses se torna mais difícil, aumentando a responsabilidade desses desenvolvedores nos resultados finais.

Nesse contexto das aplicações da IA no Judiciário, fica óbvio que o conhecimento das ciências jurídicas não deve se restringir apenas aos juízes, promotores, advogados e serventuários, mas também deve se estender a profissionais de Ciência de Dados, programadores e cientistas da computação visto que não apenas aqueles, mas estes novos profissionais que doam todo o seu conhecimento técnico para as ciências humanas aplicadas ao Direito, devem dela beber como fonte de conhecimento, garantindo que os algoritmos realmente reflitam os princípios de justiça e equidade, não apenas resultados matemáticos frios.

Outra questão a ser considerada é sobre o processo de automatização interna das decisões. As decisões automatizadas sofrem um processo intenso de critérios que muitas vezes não são conhecidos ou bem explicados por seus criadores, o que pode influenciar significativamente o dia a dia das pessoas sem que elas percebam. Como destaca Pinto (2020), algumas questões precisam ser objetivamente respondidas para garantir a integridade dessas decisões.

Se, durante algum tempo, era suficiente a abertura para a auditoria dos códigos-fonte dos algoritmos para demonstrar a sua conformidade com leis ou padrões exigidos por governos, empresas, sociedade civil organizada e organizações internacionais, diante da atual diversidade de novos sistemas de IA, percebe-se que o grau de eficiência do seu monitoramento pode ser de eficácia reduzida. Um sistema de IA baseado num sistema analítico, por exemplo, pode ser de fácil controle; todavia, em sistemas inteligentes cujo modo operacional se dá por machine learning, seus mecanismos de monitoramento podem apresentar resultados totalmente insatisfatórios (PINTO, 2020, p. 56),

A tecnologia intelectual de hoje opera como uma modalidade de cálculo racional da vida, aumentando exponencialmente a produtividade e eficiência. Contudo, isso não deve

reduzir a importância da ética no comportamento humano, especialmente porque as relações humanas estão cada vez mais intermediadas por tecnologias da informação. A *accountability* e *auditability* em sistemas de IA refere-se à necessidade de transparência e responsabilidade sobre como as decisões são tomadas. É essencial que os algoritmos utilizados no Judiciário sejam compreensíveis não apenas para os técnicos, mas também para juizes, advogados e para a sociedade em geral.

Como os algoritmos de aprendizado de máquina são tão complexos e se adaptam a novos dados, a auditoria de responsabilidade não pode se basear apenas na auditoria do código-fonte. Ela também deve incluir o registro detalhado do treinamento e da calibração do sistema para que as decisões possam ser revisadas e explicadas, permitindo assim a identificação e a correção de possíveis vieses e erros.

Não há outra alternativa; a responsabilidade e a *accountability* nas decisões algorítmicas são indispensáveis para preservar a solidez e a credibilidade do sistema judicial. Garantir que os sistemas de IA sejam transparentes, auditáveis e governados pelos princípios da ética promoverá a justiça e impedirá a implantação de desigualdades, bem como discriminação, e isso deve ser feito.

3.3. VIESES E DISCRIMINAÇÕES NOS ALGORITMOS DE IA

Os algoritmos de inteligência artificial no Judiciário apresentam um potencial significativo para aumentar a eficiência e a rapidez nas decisões. (Maia Filho e Junquilha, 2018, 230). No entanto, esses sistemas podem também introduzir ou perpetuar vieses e discriminações que afetam negativamente a justiça. Tais problemas ocorrem porque a IA aprende com dados históricos que podem refletir desigualdades e injustiças.

A identificação de vieses nos algoritmos de IA começa com a análise dos dados usados para seu treinamento. Como observa BRAGANÇA (2019), a “máquina, por conseguinte, não é inteiramente isenta, uma vez que depende de uma atividade humana preliminar de seleção de informações”, e isso inclui decisões sobre quais dados são considerados relevantes (p. 69). Quando esses dados contêm preconceitos históricos, a IA pode replicar essas distorções, tornando essencial a curadoria cuidadosa e a revisão constante dos dados. Isso significa que os desenvolvedores e operadores da IA no STF devem estar atentos às fontes e qualidade dos dados.

Como já tratado, o caso do sistema COMPAS destaca as falhas de um algoritmo enviesado, que rotulava réus negros como de maior risco de reincidência em comparação com

réus brancos. A aplicação de um sistema com tal viés, sem a devida supervisão humana, resultou em sentenças desproporcionais, perpetuando desigualdades raciais no Judiciário (ROQUE; SANTOS, 2021, p. 68).

Nesse sentido, a identificação e a mitigação desses vieses são fundamentais para garantir a neutralidade e equidade das decisões judiciais assistidas por IA o que deve ser feito através da revisão dos dados de treinamento, mas também a criação de mecanismos que permitam a intervenção humana em decisões automatizadas. Morais (2021) argumenta que “deve-se garantir institucionalmente um espaço para questionamento diante de potencial equívoco” no uso da IA (p. 323). Isso garante que, no caso de um algoritmo apresentar um viés ou erro, as partes afetadas possam contestar a decisão com base na força de argumentos transparentes e bem fundamentados que eventualmente retornariam ao ponto de responsabilização pelas decisões tomadas pela IA.

Também é importante que as ferramentas de IA sejam sujeitas a auditoria constante para garantir a aplicação adequada dos princípios de igualdade e justiça. Conforme explicado na pesquisa, os algoritmos devem estar sob avaliação constante para minimizar seus resultados tendenciosos, um processo realizado por meio de auditorias externas independentes que consideram os dados e os critérios de decisão envolvidos (GONÇALVES, 2021). São as etapas a serem tomadas para que os algoritmos não impactem injustamente os grupos minoritários.

4. A MITIGAÇÃO DOS DESVIOS E TRANSPARÊNCIA NAS FERRAMENTAS DE IA UTILIZADAS PELO JUDICIÁRIO

A transparência dos algoritmos no Judiciário é um pilar fundamental para garantir a legitimidade das decisões judiciais assistidas por IA. A opacidade desses sistemas pode gerar desconfiança entre os profissionais do direito e a sociedade, especialmente em casos que envolvem direitos fundamentais. Morais (2021) postulou que a explicação dos processos decisórios de IA sem transparência compromete a compreensão de como os dados são processados e aplicados; isso gera um sentimento de falta de controle sobre o resultado final. Nesse sentido, mecanismos que aumentem a clareza e a auditabilidade precisam ser implementados para garantir a confiança pública nas decisões tomadas com o auxílio de algoritmos.

4.1. RELATÓRIOS DE IMPACTO ALGORÍTMICO

Os Relatórios de Impacto Algorítmico nascem como uma das principais soluções para reduzir ou mitigar a opacidade dos algoritmos. Esses relatórios se apresentam como instrumentos para documentar e explicar como os processos de tomada de decisão algorítmica foram implementados, trazendo à tona o processo e a transparência das operações. Morais (2021) afirma que a usabilidade da IA no Judiciário deve ser condicionada ao preceito de que suas decisões possam ser compreendidas pelas partes e posteriormente, se necessário, contestadas. Os relatórios devem ser aqueles que forneçam detalhes dos dados, dos fatores e das razões para se chegar a uma decisão.

Além disso, as ferramentas de IA precisam passar por auditorias contínuas para assegurar que princípios de igualdade e imparcialidade estejam sendo aplicados de maneira correta. No entanto, Sanches (2021) citando Green e Kak (2021) argumenta que formas mais significativas de supervisão humana são incrivelmente difíceis de serem realizadas na prática. Mesmo que os operadores humanos recebam treinamento intensivo para trabalhar com IA, torna-se extremamente complexo "intervir para equilibrar adequadamente as percepções humanas e algorítmicas". Isso ocorre porque o processo de tomada de decisão é muitas vezes obscurecido pela complexidade do algoritmo, dificultando que os operadores avaliem com precisão os resultados gerados e intervenham de forma eficaz.

Mesmo operadores humanos especializados podem não executar praticamente uma intervenção adequada e equilibrada devido ao nível de complexidade que o algoritmo acarreta. Portanto, a implementação de relatórios de impacto algorítmico não apenas aumenta a

transparência, mas fornece um canal para o qual a responsabilização e a supervisão podem ser atualizadas.

As avaliações de impacto algorítmico devem fornecer informações de auditoria detalhadas e contínuas e, em detalhes, disponibilizar o funcionamento do algoritmo internamente em termos acessíveis não apenas para os especialistas em tecnologia, mas também para os profissionais do direito e o público, bem como o histórico de decisões, comparando-as com decisões passadas para mostrar como os dados foram processados e em quais parâmetros cada decisão judicial foi tomada. Também deve ser incluída uma reflexão de dados sobre provável viés e propostas explícitas para sua mitigação e ajuste futuro. Para controle e supervisão humanos, os relatórios poderiam ser revisados periodicamente por painéis de diferentes disciplinas, o que também confirmaria a neutralidade e a justiça das decisões algorítmicas — dessa forma, ajudando a construir mais confiança e responsabilidade pelo uso de IA no Judiciário.

4.2. PAINÉIS MULTIDISCIPLINARES

A proposta de criação de “Painéis Multidisciplinares” seria uma outra estratégia desenvolvida como medida de transparência e controle no parâmetro do uso da IA pelo Judiciário. Sua composição deveria envolver profissionais de vários campos, não se limitando ao direito e da tecnologia, mas da ética, psicologia, antropologia, psicologia, ciências sociais, etc. Sua finalidade deveria ser a de empreender uma reflexão sobre o impacto das decisões judiciais exaradas por IA. A combinação dessas especialidades permitiria uma abordagem equilibrada e rigorosa, garantindo que as decisões judiciais não sejam baseadas apenas em critérios técnicos, mas também em valores éticos e sociais que preservam a justiça e sopesem o respeito aos direitos e garantias fundamentais e princípios democráticos assentadas na nossa Constituição Cidadã. Isso geraria um mecanismo extra, por meio do qual as decisões algorítmicas poderiam ser periodicamente verificadas para a identificação de possíveis vieses e erros que possam surgir no processo decisório.

Esses painéis também poderiam propor diretrizes e recomendações para o uso futuro de IA no Judiciário, bem como quaisquer outras questões que considerem necessárias para uma discussão substancial e justa entre especialistas jurídicos e cientistas de dados para estabelecer critérios objetivos sobre o uso de algoritmos respeitando princípios constitucionais, como igualdade e dignidade humana. Os eticistas teriam que fazer uma contribuição substancial para essa avaliação sobre os efeitos sociais de decisões automatizadas e especialmente aquelas

decisões que estão conectadas com direitos fundamentais. Dessa forma, padrões mais altos de transparência e responsabilização no uso de IA podem ser definidos por meio de seu envolvimento, o que aumentaria a confiança dos cidadãos nas decisões judiciais.

Além disso, para garantir que esses painéis possam cumprir sua função adequadamente, as atividades devem ser incorporadas às instituições e regulamentadas, com mecanismos de auditoria externa e relatórios periódicos de resultados. A transparência deve orientar o princípio, tornando o processo de revisão acessível a todos e isso inclui as partes interessadas em processos judiciais. Além disso, a inclusão de diferentes setores no painel realizaria uma análise mais completa das decisões que permitem a correção de falhas sistêmicas que prejudicam os cidadãos antes de sua remediação e que promove uma integração mais ética e justa da IA e do Direito no contexto judicial.

4.3. SISTEMAS HÍBRIDOS DE DECISÃO

A implementação de sistemas híbridos de tomada de decisão no Judiciário, que unem a eficiência dos algoritmos com a análise crítica de operadores humanos, parece ser um equilíbrio adequado para maior transparência e responsabilização em decisões apoiadas por IA. Tais sistemas tornarão possível que as decisões propostas pelos algoritmos sejam consideradas e aprovadas na hora por um juiz humano. Assim, a intervenção humana garante que os algoritmos não visem apenas automatizar processos, mas respeitar as sutilezas culturais e sociais implicadas em cada caso. O julgamento individual, quando aplicado em conjunto com a capacidade de análise rápida da IA, pode ser útil para detectar qualquer provável viés e, então, tomar uma decisão mais justa e equilibrada.

A eficácia do sistema híbrido depende do treinamento adequado de operadores humanos nos aspectos técnicos das operações de algoritmos para permitir intervenções adequadas. Isso inclui as habilidades para identificar qualquer possível erro, viés ou inconsistência nos resultados de saída obtidos de uma IA. Um design de sistema híbrido eficaz deve ser amigável aos usuários, tendo interfaces prontas para os operadores interagirem com algoritmos. O desenvolvimento de tais interfaces deve garantir processos de tomada de decisão transparentes, possibilitando que um operador acompanhe como o algoritmo eventualmente fez uma determinada recomendação. Os protocolos de revisão também precisam ser claros sobre quando no processo a intervenção humana é necessária e a extensão de tal intervenção.

Em conjunto com os painéis multidisciplinares, os julgadores poderiam verificar, controlar e intervir para que haja adequação dos resultados através de intervenções humanas. A

interação entre tecnologia e julgamento humano nos sistemas híbridos, portanto, não apenas promoveria a justiça, mas também fortaleceria a confiança pública na utilização da IA no Judiciário.

5. CONSIDERAÇÕES FINAIS

Mesmo quando a integração da Inteligência Artificial ao sistema jurídico brasileiro promete ganhos significativos de eficiência e velocidade, esse movimento entra em conflito com a transparência e a justificação das decisões. Por exemplo, como todas as decisões judiciais precisam ser públicas e motivadas pelo Artigo 93, inciso IX, da Constituição Federal, a inexplicabilidade do processo algorítmico pode comprometer a legitimidade de quaisquer decisões tomadas com a ajuda da IA. Na implementação do Projeto Victor, o STF é obrigado a garantir que as ferramentas tecnológicas garantam a adesão a esse princípio, tornando as decisões compreensíveis e auditáveis

A criação dos relatórios de impacto algorítmico e dos sistemas híbridos de decisão, surgem como soluções viáveis para mitigar os riscos de opacidade e falta de fundamentação, pois mesclam a capacidade de análise dos algoritmos com a supervisão humana. Os Painéis Multidisciplinares – em que há a colaboração entre operadores do Direito, cientistas de dados e especialistas nas áreas das ciências humanas e sociais – tornam-se essenciais para garantir que os sistemas de IA no Judiciário operem de forma justa, transparente e responsiva aos valores constitucionais. A auditoria contínua desses sistemas é outro item crucial para identificar e corrigir eventuais vieses ou falhas.

Dessa forma, a implementação de IA no Judiciário requer um equilíbrio entre tecnologia e supervisão humana. É fundamental que as inovações tecnológicas sejam acompanhadas de mecanismos rigorosos de controle, garantindo que os princípios de publicidade e fundamentação das decisões sejam respeitados, preservando, assim, a confiança pública no sistema judicial brasileiro.

REFERÊNCIAS BIBLIOGRÁFICAS

BONAT, Débora; PEIXOTO, Fabiano Hartmann. **Racionalidade no Direito: Inteligência Artificial e Precedentes**. Curitiba: Alteridade, 2020.

BRAGANÇA, Fernanda; BRAGANÇA, Laurinda Fátima da F. P. G. **Revolução 4.0 no Poder Judiciário: levantamento do uso de inteligência artificial nos tribunais brasileiros**. Revista da Seção Judiciária do Rio de Janeiro, v. 23, n. 46, p. 65-76, jul./out. 2019. Disponível em: <https://doi.org/10.30749/2177-8337.v23n46p65-76>. Acesso em: 14 set. 2024.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicaocompilado.htm. Acesso em: 15 set. 2024.

BRASIL. **Código Civil**, 2015. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/2002/L10406compilada.htm. Acesso em: 15 set. 2024.

CONSELHO NACIONAL DE JUSTIÇA (Brasil). **Resolução nº 332, de 21 de agosto de 2020**. *Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências*. Diário da Justiça Eletrônico: Conselho Nacional de Justiça, Brasília, DF, n. 274, p. 4-8, 25 ago. 2020. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3429>. Acesso em: 14 set. 2024.

GONÇALVES, Mariana Sbaite. **Viés algorítmico e discriminação: Como os algoritmos de IA podem perpetuar e amplificar vieses sociais**. Migalhas, 2021. Disponível em: <https://www.migalhas.com.br/depeso/415125/vies-algoritmico-e-discriminacao-ia-pode-amplificar-vieses-sociais>. Acesso em: 15 set. 2024.

GUTIERREZ, Andriei. **É possível confiar em um sistema de inteligência artificial?: práticas em torno da melhoria da sua confiança, segurança e evidências de accountability**. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo: Revista dos Tribunais, 2019. p. 83-97.

HAN, J.; PEI, J.; KAMBER, M. **Data mining: concepts and techniques**. Amsterdam: Elsevier, 2011. Disponível em: https://archive.org/details/dataminingconcep00hanj_0/page/746/mode/2up. Acesso em: 13 set. 2024.

LAGE, Fernanda de Carvalho. **Manual de inteligência artificial no direito brasileiro**. 2. ed. Salvador: JusPODIVM, 2022.

MAIA FILHO, Mamede Said; JUNQUILHO, Tainá Aguiar. **Projeto Victor: Perspectivas de aplicação da inteligência artificial ao direito**. Revista de Direitos e Garantias Fundamentais, Vitória, v. 19, n. 3, p. 219-238, set./dez. 2018. DOI: <http://dx.doi.org/10.18759/rdgf.v19i3.1587>. Acesso em: 14 set. 2024.

MORAIS, Fausto Santos de. **O uso da inteligência artificial na repercussão geral: desafios teóricos e éticos**. Revista de Direito Público, Brasília, v. 18, n. 100, p. 306-326, out./dez. 2021.

Disponível em: <https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/6001>. Acesso em: 14 set. 2024.

NUNES, Dierle; MARQUES, Ana Luiza Pinto Coelho. **Inteligência Artificial e Direito Processual: vieses, algoritmos e os riscos de atribuição de função decisória às máquinas.** Revista de Processo, v. 285, p. 421-447, nov. 2018.

PASSOS, Hugo Assis. **Inteligência artificial e a repercussão geral da questão constitucional: análise crítica e parâmetros de utilização.** 2022. Tese (Doutorado em Direito Constitucional) – Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP), Brasília, DF, 2022. Disponível em: https://repositorio.idp.edu.br/bitstream/123456789/4673/1/Tese_HUGO%20ASSIS%20PASSOS%20_Doutorado_2022.pdf. Acesso em: 17 set. 2024.

PINTO, Henrique Alves. **A utilização da inteligência artificial no processo de tomada de decisões: por uma necessária accountability.** Revista de Informação Legislativa: RIL, Brasília, DF, v. 57, n. 225, p. 43-60, jan./mar. 2020. Disponível em: http://www12.senado.leg.br/ril/edicoes/57/225/ril_v57_n225_p43. Acesso em: 10 out. 2024.

ROQUE, André Vasconcelos; SANTOS, Lucas Braz Rodrigues dos. **Inteligência artificial na tomada de decisões judiciais: três premissas básicas.** Revista Eletrônica de Direito Processual, Rio de Janeiro, v. 22, n. 1, p. 58-78, jan./abr. 2021. Disponível em: <https://www.e-publicacoes.uerj.br/redp/article/view/53537/36309>. Acesso em: 15 set. 2024.

SANCHES, Vanessa Karam de Chueiri. **A supervisão humana e a discriminação algorítmica no sistema judiciário.** Conjur, 2021. Disponível em: <https://www.conjur.com.br/2021-jun-25/nunes-supervisao-humana-decisoes-ia-reduz-riscos/>. Acesso em: 15 set. 2024.