



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE MESTRADO ACADÊMICO
DEPARTAMENTO DE DIREITO PÚBLICO
PROGRAMA DE PÓS-GRADUAÇÃO EM DIREITO

INTELIGÊNCIA ARTIFICIAL NO PODER JUDICIÁRIO: ANÁLISE DO RISCO DE
VIESES ALGORÍTMICOS EM DECISÕES JUDICIAIS

FORTALEZA-CE

2025

ALINE BESSA SAMPAIO

**INTELIGÊNCIA ARTIFICIAL NO PODER JUDICIÁRIO: ANÁLISE DO RISCO DE
VIESES ALGORÍTMICOS EM DECISÕES JUDICIAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Direito da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Direito. Área de concentração: Constituição, Sociedade e Pensamento Jurídico .

Orientador: Prof. Dr. Felipe Braga Albuquerque.

FORTALEZA-CE

2025

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

S181i sampaio, Aline bessa.
inteligência artificial no poder judiciário : análise do risco de vieses algorítmicos em decisões judiciais / Aline bessa sampaio. – 2025.
140 f.

Dissertação (mestrado) – Universidade Federal do Ceará, Faculdade de Direito,
Programa de Pós-Graduação em Direito, Fortaleza, 2025.

Orientação: Prof. Dr. Felipe Braga Albuquerque.

1. inteligência artificial. 2. poder judiciário. 3. vieses algorítmicos. I. Título.

CDD 340

ALINE BESSA SAMPAIO

**INTELIGÊNCIA ARTIFICIAL NO PODER JUDICIÁRIO: ANÁLISE DO RISCO DE
VIESES ALGORÍTMICOS EM DECISÕES JUDICIAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Direito da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Direito. Área de concentração: Constituição, Sociedade e Pensamento Jurídico .

Aprovada em 29/04/2025.

BANCA EXAMINADORA

Prof. Dr. Felipe Braga Albuquerque (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Robério Fontenele de Carvalho
Universidade de Fortaleza (UNIFOR)

Prof. Dr. Renaud Ponte Aguiar
Faculdade Luciano Feijão (FLF)

A Deus.

Aos meus pais, Teresinha de Jesus e Domingos José , meu esposo Igor, meus amigos e professores, especialmente meu orientador, prof. Felipe Braga.

AGRADECIMENTOS

A Deus, pela fortaleza e inspiração durante toda a minha jornada acadêmica.

Aos meus pais, Teresinha de Jesus e Domingos José, cujo amor incondicional e incentivo foram fundamentais para cada conquista obtida ao longo deste percurso.

Ao meu esposo, Igor, pela compreensão, paciência e companheirismo inestimáveis.

Aos meus irmãos, Alessandra, Daniele e Domingos Neto, por todo o apoio e pelas palavras de encorajamento que me motivaram a prosseguir.

Ao Prof. Dr. Felipe Braga, pela excelente orientação, orientando-me com rigor e dedicação, e contribuindo decisivamente para o amadurecimento deste trabalho.

“Não sabendo que era impossível, ele foi lá e fez.” (Jean Cocteau).

RESUMO

A presente dissertação tem como objetivo analisar a aplicação de sistemas de Inteligência Artificial (IA) no Poder Judiciário brasileiro, com ênfase na identificação e mitigação de vieses algorítmicos em decisões judiciais. Considerando o potencial dessas tecnologias para promover maior eficiência, celeridade processual e padronização de entendimentos, a pesquisa parte da premissa de que sua adoção, quando realizada sem critérios técnicos e jurídicos rigorosos, pode comprometer princípios constitucionais como equidade, imparcialidade, contraditório e segurança jurídica. A metodologia adotada baseia-se em revisão bibliográfica e análise documental de normas, artigos científicos e casos práticos. O trabalho estrutura-se em quatro capítulos: o primeiro apresenta os conceitos e a evolução da IA; o segundo analisa sua aplicação no Judiciário brasileiro; o terceiro aprofunda o estudo do viés algorítmico, discutindo suas fontes, impactos e riscos; e o quarto propõe estratégias de mitigação, como governança de dados, auditoria algorítmica e regulação normativa. Entre os principais resultados, destaca-se que os vieses frequentemente derivam de dados históricos enviesados, os quais, ao serem utilizados no treinamento dos algoritmos, podem perpetuar desigualdades e comprometer a legitimidade das decisões automatizadas. Diante disso, evidenciam-se a importância da curadoria de dados, da explicabilidade dos sistemas e da supervisão humana ao longo de todo o ciclo de vida da IA. Ressalta-se, ainda, a necessidade de marcos regulatórios sólidos e adaptativos, que estabeleçam parâmetros mínimos de transparência, responsabilização e controle sobre os sistemas utilizados. A pesquisa conclui que a articulação entre inovação tecnológica e valores constitucionais deve ser orientada por uma política pública clara, por práticas de governança transparentes e por uma formação continuada dos profissionais envolvidos. Assim, a IA deve ser compreendida como instrumento de apoio à atividade jurisdicional, e não como substituto da atuação humana, assegurando que a tecnologia seja empregada de maneira ética, democrática e alinhada aos direitos fundamentais.

Palavras-chave: inteligência artificial; poder judiciário; vieses algorítmicos.

ABSTRACT

This dissertation aims to analyze the application of Artificial Intelligence (AI) systems in the Brazilian Judiciary, with an emphasis on identifying and mitigating algorithmic biases in judicial decisions. Considering the potential of these technologies to promote greater efficiency, procedural speed, and standardization of understandings, the research is based on the premise that their adoption, when carried out without rigorous technical and legal criteria, can compromise constitutional principles such as equity, impartiality, adversarial proceedings, and legal certainty. The methodology adopted is based on a bibliographic review and documentary analysis of standards, scientific articles, and practical cases. The work is structured in four chapters: the first presents the concepts and evolution of AI; the second analyzes its application in the Brazilian Judiciary; the third delves deeper into the study of algorithmic bias, discussing its sources, impacts, and risks; and the fourth proposes mitigation strategies, such as data governance, algorithmic auditing, and regulatory regulation. Among the main results, it is worth highlighting that biases often derive from biased historical data, which, when used to train algorithms, can perpetuate inequalities and compromise the legitimacy of automated decisions. In view of this, the importance of data curation, system explainability and human supervision throughout the entire AI life cycle is highlighted. It is also worth highlighting the need for solid and adaptive regulatory frameworks that establish minimum parameters of transparency, accountability and control over the systems used. The research concludes that the articulation between technological innovation and constitutional values must be guided by a clear public policy, transparent governance practices and ongoing training of the professionals involved. Thus, AI must be understood as an instrument to support jurisdictional activity, and not as a substitute for human action, ensuring that technology is used in an ethical, democratic manner and in line with fundamental rights.

Keywords: artificial intelligence; judiciary; algorithmic biases.

SUMÁRIO

1	INTRODUÇÃO	14
2	INTELIGÊNCIA ARTIFICIAL.....	18
2.1	Conceito e Evolução da Inteligência Artificial.....	20
2.2	Funcionamento dos Algoritmos.....	27
2.2.1	<i>Machine Learning, Deep Learning e Redes Neurais Artificiais</i>	30
2.2.2	<i>Inteligência Artificial Generativa</i>	40
3	INTELIGÊNCIA ARTIFICIAL NO PODER JUDICIÁRIO.....	48
3.1	Aplicações de IA no Poder Judiciário Brasileiro.....	50
3.2	Benefícios do Uso da IA na Tomada de Decisão Judicial.....	55
3.3	Riscos do Uso da IA na Tomada de Decisão judicial.....	60
3.3.1	<i>Opacidade</i>	67
3.3.2	<i>Discriminação</i>	73
4	VIÉS ALGORÍTMICO.....	77
4.1	Definição de viés algorítmico e discriminação automatizada	81
4.2	Fontes de viés nos sistemas de IA.....	90
4.3	Impactos do viés na equidade, imparcialidade e confiança no Judiciário	96
4.4	Mecanismos de mitigação e regulação.....	103
5	CONCLUSÃO	121
	REFERÊNCIAS	126

1 INTRODUÇÃO

O avanço das tecnologias digitais nas últimas décadas tem transformado, de forma substancial, a maneira como indivíduos, instituições e governos interagem entre si e com o mundo ao seu redor. Entre as inovações tecnológicas que mais suscitam debates atualmente, destaca-se a Inteligência Artificial (IA), não apenas por seu elevado potencial de inovação, mas também pelos significativos desafios éticos, jurídicos e sociais que impõe. A incorporação de sistemas automatizados, baseados em algoritmos, em diferentes esferas da vida humana tem alterado radicalmente a lógica de funcionamento de setores como a saúde, a educação, o mercado financeiro e, mais recentemente, o Poder Judiciário.

A aplicação de algoritmos em processos judiciais representa uma oportunidade ímpar para modernizar os sistemas de justiça, promover maior eficiência, agilizar a tramitação processual e padronizar entendimentos jurídicos. Todavia, a introdução da IA no âmbito judicial também levanta relevantes questionamentos de natureza ética, técnica e jurídica, especialmente no que tange à possibilidade de ocorrência de vieses algorítmicos capazes de comprometer princípios fundamentais, como a equidade, a imparcialidade e a segurança jurídica. Nesse contexto, o presente trabalho tem como objetivo central a análise crítica desses desafios, com foco nas armadilhas e oportunidades que envolvem a adoção de tecnologias de IA nos tribunais brasileiros.

A integração da IA ao sistema de justiça é, frequentemente, celebrada como uma solução tecnológica para lidar com a sobrecarga de processos e o atraso estrutural que afeta os tribunais. Ferramentas baseadas em IA demonstram potencial para automatizar tarefas repetitivas, tais como a triagem de casos, a análise de jurisprudência e até mesmo a elaboração de minutas de decisões. Entretanto, a implementação dessas tecnologias exige atenção criteriosa, de modo a evitar que a busca pela eficiência tecnológica se sobreponha aos valores constitucionais que estruturam o sistema jurídico brasileiro.

Nesse sentido, a chamada opacidade algorítmica — entendida como a dificuldade em compreender os processos internos que conduzem à tomada de decisões por algoritmos — constitui uma preocupação central no debate atual. A

ausência de transparência quanto ao funcionamento interno dos sistemas pode dificultar a identificação de erros ou vieses, comprometendo não apenas a legitimidade das decisões judiciais automatizadas ou assistidas, mas também a confiança pública no sistema de justiça como um todo.

Entre as principais fontes de risco associadas à aplicação de IA no Judiciário, destaca-se o viés algorítmico, que pode emergir em diferentes fases do desenvolvimento e da implementação dos sistemas. Tais vieses podem estar presentes, por exemplo, nos dados utilizados para treinar os algoritmos, os quais frequentemente refletem desigualdades históricas e estruturais presentes na sociedade. Conseqüentemente, decisões judiciais automatizadas ou assistidas por IA podem, ainda que de forma não intencional, reproduzir preconceitos que desfavorecem determinados grupos sociais, violando os princípios constitucionais de igualdade e justiça.

Assim, torna-se imprescindível a adoção de práticas rigorosas de governança de dados, que envolvam desde a seleção e o tratamento adequado das informações até a realização de auditorias periódicas com vistas à avaliação da qualidade, precisão e imparcialidade dos sistemas utilizados. Nesse contexto, a adoção de abordagens interdisciplinares — que integrem conhecimentos do Direito, da Ciência da Computação, da Estatística e da Ética aplicada — mostra-se essencial para o enfrentamento dos desafios relacionados à opacidade e aos vieses algorítmicos.

A formação continuada de profissionais do Direito, em diálogo com especialistas em ciência de dados e inteligência artificial, revela-se, nesse cenário, como elemento fundamental para a construção de soluções que respeitem os direitos fundamentais e promovam a legitimidade das decisões judiciais mediadas por tecnologias. Outrossim, a criação de marcos regulatórios sólidos e específicos para o uso de IA no Judiciário é apontada como uma prioridade estratégica, a fim de assegurar que as inovações tecnológicas sejam compatíveis com os valores fundamentais do Estado Democrático de Direito.

No Brasil, a introdução da IA no sistema de justiça representa uma nova fronteira de modernização institucional, acompanhada de promessas de maior celeridade processual, redução de custos operacionais e aumento da eficiência decisória. Iniciativas como o projeto *Victor*, desenvolvido pelo Supremo Tribunal Federal (STF), ilustram de maneira concreta essa realidade emergente. Contudo, a

par das oportunidades, surgem também riscos relevantes que não podem ser negligenciados, sobretudo quando tais tecnologias são aplicadas em contextos decisórios sensíveis, como ocorre nos processos judiciais.

Nesse cenário, um dos problemas mais complexos e urgentes é, justamente, o viés algorítmico. Diferentemente de falhas técnicas pontuais, trata-se de um fenômeno estruturante e, muitas vezes, invisível, decorrente de uma cadeia de fatores que inclui a coleta e a curadoria dos dados, os parâmetros adotados no treinamento dos modelos, e os contextos nos quais os algoritmos são aplicados. Se não for devidamente compreendido e mitigado, o viés algorítmico pode comprometer gravemente valores jurídicos fundamentais, como a imparcialidade, a equidade, a dignidade da pessoa humana e a confiança no Poder Judiciário.

A presente pesquisa tem como objetivo geral analisar criticamente o uso da Inteligência Artificial no âmbito do Poder Judiciário brasileiro, com ênfase no risco de vieses algorítmicos em decisões judiciais. Para alcançar tal objetivo, o trabalho está estruturado em quatro grandes partes, inter-relacionadas, que fornecem um panorama teórico e prático sobre o tema proposto.

O Capítulo 2 apresenta os fundamentos da Inteligência Artificial, abordando sua evolução histórica, os principais conceitos e o funcionamento dos algoritmos. São exploradas as tecnologias mais relevantes no contexto atual, como *Machine Learning*, *Deep Learning*, Redes Neurais Artificiais e Inteligência Artificial Generativa. A compreensão desses elementos técnicos é essencial para o entendimento de suas implicações jurídicas.

O Capítulo 3 é dedicado à análise da utilização da IA no Poder Judiciário brasileiro. Discute-se, nesse capítulo, o panorama atual das aplicações de IA no sistema de justiça, os benefícios esperados — como maior celeridade, uniformidade e redução de carga de trabalho —, bem como os principais riscos identificados, com destaque para a opacidade dos sistemas e a possibilidade de reprodução de desigualdades sociais preexistentes.

O Capítulo 4, núcleo central da dissertação, aprofunda a discussão sobre o viés algorítmico. São abordados, nessa seção: a definição do conceito e suas manifestações na forma de discriminação automatizada (item 4.1); as principais fontes dos vieses em sistemas de IA, especialmente relacionadas à coleta e ao tratamento de dados (item 4.2); os impactos desses vieses na imparcialidade, equidade e confiança institucional do Judiciário (item 4.3); e, por fim, as estratégias

de mitigação e regulação (item 4.4), com destaque para políticas de governança algorítmica, auditoria, técnicas de IA explicável (*Explainable AI – XAI*), além da recente Resolução CNJ nº 615/2025.

A escolha do tema justifica-se pela crescente e, muitas vezes, acrítica adoção de tecnologias algorítmicas no sistema de justiça, o que demanda uma análise mais aprofundada de seus impactos sobre os direitos fundamentais, a inclusão e a justiça social. A proposta deste trabalho não se insere em uma perspectiva contrária à inovação, mas sim em um projeto de construção de um modelo de justiça que integre os avanços tecnológicos com responsabilidade, transparência e adesão aos princípios democráticos e constitucionais.

Ao final desta investigação, espera-se contribuir para o debate jurídico contemporâneo sobre a digitalização do Judiciário, evidenciando tanto os benefícios quanto os desafios associados à adoção da Inteligência Artificial em decisões judiciais. O presente trabalho busca reafirmar a centralidade do ser humano no processo judicial e defender que a IA seja utilizada como ferramenta de apoio, e não como substituta da atuação humana, sempre subordinada aos princípios da dignidade da pessoa humana, do devido processo legal e do acesso à justiça.

Portanto, a pesquisa reforça a necessidade urgente de desenvolver políticas públicas e regulamentos que harmonizem a inovação tecnológica com os princípios de justiça, equidade e transparência. A adoção de mecanismos robustos de supervisão, auditoria e avaliação, juntamente com a implementação de critérios claros de explicabilidade e controle social, constitui estratégia indispensável para assegurar que a IA seja aplicada de maneira ética, responsável e condizente com os valores do Estado de Direito. Dessarte, espera-se que a aplicação dessas tecnologias contribua não apenas para maior eficiência e racionalidade processual, mas também para o fortalecimento da segurança jurídica e da confiança da sociedade no sistema de justiça, consolidando, assim, uma integração equilibrada entre inovação tecnológica e valores constitucionais.

2 INTELIGÊNCIA ARTIFICIAL

No presente capítulo, são introduzidos conceitos e pressupostos fundamentais para a compreensão desta pesquisa, que, embora vinculada ao campo do Direito, possui uma interface relevante com a área da Tecnologia e suas consequências. Além disso, discute-se como a Inteligência Artificial se desenvolveu desde a década de 1940 até a atual ênfase na IA generativa .

Conforme Castro (2021), considera-se antropológica toda investigação que tenha como objeto entidades dotadas de inteligência. Nesse sentido, a contemporaneidade é marcada por três áreas centrais de estudo antropológico, associadas às três “inteligências”: a cultural, a natural e a artificial, tendo a Terra como ponto de convergência. Na tradição filosófica moderna, esse triângulo é reconhecido como “humanos, animais e máquinas”. A interação entre essas inteligências pode ocorrer de modo colaborativo, sobreposto e até conflituoso, decorrendo de uma disputa epistemológica produtiva acerca de qual seria a inteligência primordial ou mais geral, na qual cada uma pode ser concebida como uma variante das outras.

Seguindo essa perspectiva, Castro (2021) observa que a inteligência cultural se dedica a investigar as diferenças nos modos como os seres humanos “constroem o mundo”. Já a inteligência natural está voltada ao estudo das formas de interação de entidades não humanas tanto entre si quanto com diferentes modos humanos, abrangendo, assim, estudos multiespécies. Já a terceira inteligência é designada como “artificial”, centrando-se em dispositivos e arranjos lógico-materiais habitualmente desenvolvidos por humanos, com capacidade efetiva ou suposta de construir mundos e de estabelecer conexões, ou mesmo se dissociar, dos mundos humanos. Nesse âmbito se inserem os diversos estudos sobre IA, as reflexões filosóficas e antropológicas acerca da tecnologia, o recente interesse pelas “cosmotécnicas” e as especulações metafísicas acerca de questões como “pós-humano” e “inumano”.

A análise proposta por Castro (2021) conduz a um questionamento central no estudo da inteligência artificial: estariam as máquinas efetivamente equiparadas à inteligência humana, a ponto de se poder afirmar que elas pensam?

Nesse sentido, Han (2022) sugere que, em razão de fatores como a ausência de afeto, de comoção, de espírito ou de uma escuta do pensamento — ou, possivelmente, de todos esses aspectos ao mesmo tempo —, a inteligência artificial não chega a pensar de fato, mas apenas realiza cálculos.

Segundo Han (2022), a inteligência artificial baseia-se em aprendizagem a partir do passado, de modo que o “futuro” que ela calcula não corresponde ao conceito de futuro em sua acepção plena. Ela é incapaz de perceber eventos inesperados; enquanto isso, o pensamento se caracteriza como um acontecimento, introduzindo algo inteiramente novo no mundo. Falta, portanto, à inteligência artificial a “negatividade da ruptura”, indispensável para que surjam elementos inéditos em sentido enfático, o que faz com que, em essência, tudo permaneça igual. A noção de “inteligência” implica a ideia de escolha (interlegere), porém, na prática, a IA apenas seleciona entre opções previamente definidas — em última instância, entre 0 e 1 —, sem transcender o que já aconteceu em direção ao não acontecido.

Conforme Gabriel (2021), não seria plausível conceber as ferramentas de busca como dotadas de inteligência no mesmo sentido em que a possuímos. Isso porque qualquer resultado fornecido por um sistema de inteligência artificial advém de uma base de dados finita — embora em constante expansão devido ao Big Data. Desse modo, as máquinas são capazes de reproduzir apenas determinados aspectos da cognição humana, como a realização de cálculos mentais (atividade que há muito tempo já é automatizada por calculadoras), mas não se estendem a todas as dimensões do pensamento humano. Sob essa perspectiva, adentramos uma área ainda pouco elucidada nos estudos de tecnologia, conhecida como inteligência artificial geral.

Ainda não se verificou, no presente estágio do desenvolvimento tecnológico, a existência desse tipo de inteligência artificial, embora haja quem acredite que a humanidade esteja mais ou menos próxima de alcançar tal ruptura — conforme sugere Gabriel (2021). Dessa forma, convém ressaltar que a chamada inteligência artificial forte, geral ou universal não constitui o objeto de análise deste trabalho. A investigação aqui proposta se limita ao estudo de inteligências artificiais fracas já em operação no contexto atual. Tal conceito será abordado ao longo do presente capítulo.

2. 1 Conceito e Evolução da Inteligência Artificial

Desde os primórdios da história humana, a disseminação de informações tem sido uma constante, manifestando-se tanto através da tecnologia quanto na forma como nos comunicamos. Esta interação entre linguagem e tecnologia se reflete na fabricação de diversos artefatos evolutivos, como roupas, armadilhas, armas, entre outros, que não apenas servem como meios de sobrevivência, mas também como veículos para transmitir conhecimento e facilitar a vida cotidiana.

Tanto a tecnologia quanto a linguagem são produtos das culturas específicas em que surgem, moldadas por suas necessidades e contextos. Elas evoluem conforme as esperanças e os receios de cada sociedade, adaptando-se continuamente ao ambiente em que estão inseridas.

Carvalho (2021) aponta que é essencial notar que toda inovação, especialmente novas tecnologias, suscita tanto esperanças quanto receios. Um exemplo emblemático remonta à descoberta do domínio do fogo, que pode ser empregado tanto para o aquecimento quanto para causar danos. Da mesma maneira, todo progresso científico, tecnológico e inovador pode acarretar tanto vantagens quanto desvantagens, dependendo da sua aplicação.

No atual cenário complexo, as ciências tecnológicas emergem como um dos pilares essenciais da sociedade contemporânea. Desde as invenções mais rudimentares até os sistemas mais avançados, como a Inteligência Artificial (IA), a evolução tecnológica tem sido um elemento central na trajetória humana.

Em 1956, durante a Conferência de Dartmouth, John McCarthy cunhou o termo "inteligência artificial", reunindo especialistas interessados na área emergente. Segundo Sichman (2021), a IA surgiu na década de 1950, acompanhando o desenvolvimento dos computadores. A Conferência de Dartmouth é considerada seu marco inicial, contando com a participação de pioneiros como McCarthy, Marvin Minsky, Alan Newell e Herbert Simon, que estabeleceram as bases desse campo da Computação.

A Inteligência Artificial enfrentou períodos de entusiasmo e ceticismo, conhecidos como "invernos da IA", marcados pela redução de financiamento devido a expectativas infladas e desafios técnicos. Segundo Sichman (2021), esses ciclos

se repetiram, alternando fases de avanços e investimentos com momentos de desilusão, como ocorreu entre 1975-1980 e 1987-1993.

No entanto, a virada do século XXI foi marcada por um renascimento notável na área da Inteligência Artificial, impulsionado principalmente pelos avanços em algoritmos de aprendizado profundo e pelo aumento da capacidade computacional.

Klaus Schwab(2016) denominou como Quarta Revolução Industrial a esse período de transformação tecnológica caracterizado pela fusão de tecnologias digitais, físicas e biológicas , a qual se configura como uma revolução científico-tecnológica precipuamente caracterizada pela velocidade, amplitude, profundidade e impacto sistêmico, cujo elemento central é a IA, que será devidamente conceituado a seguir .

Desde o início dos anos 2000, com a crescente disseminação da internet, temos testemunhado um contínuo avanço na adoção da tecnologia e na sua crescente integração nas vidas das pessoas. A inovação trouxe conveniências que estão modificando a maneira como enfrentamos os desafios cotidianos. Exemplos dessas mudanças em nossas vidas incluem o uso do aplicativo Waze para encontrar o trajeto mais eficiente até um destino e o uso de redes sociais que incorporam algoritmos cada vez mais eficazes. O progresso tecnológico e a chamada Revolução 4.0 estão claramente visíveis em todos os aspectos da sociedade, com a introdução e aplicação de ferramentas de inteligência artificial e tecnologias avançadas. As interações sociais já acontecem nesse contexto há algum tempo, abrangendo desde algoritmos que classificam serviços e estabelecem preços até a automação de processos industriais, serviços financeiros, entretenimento e comunicação. Em outros países, inclusive no setor público, a situação é semelhante, com o uso de inteligência artificial para prever crimes, otimizar a alocação de recursos e automatizar tarefas burocráticas e administrativas do governo (Ferrari,2018).

Autores como Froomkin, Kerr e Pineau (2019) destacam que essa evolução tem influenciado até mesmo os diagnósticos médicos, sugerindo que as máquinas poderão superar os humanos nessa área. Essa percepção decorre do progresso significativo nas áreas de conhecimento relacionadas à tecnologia, em conjunto com outras disciplinas científicas.

A trajetória da Inteligência Artificial, desde seus primórdios mitológicos até suas realizações modernas, representa um testemunho da incessante busca

humana por compreender, replicar e, eventualmente, aprimorar a própria inteligência. Essa jornada visa também aperfeiçoar o desenvolvimento de funções de forma mais eficiente com o uso dessa tecnologia emergente.

Ao vislumbrar o futuro, torna-se evidente que a história da Inteligência Artificial ainda está em curso, prometendo revoluções ainda mais profundas nas experiências, no trabalho e no pensamento humano, ao mesmo tempo em que levanta questões controversas que precisam ser abordadas.

Cozman (2021) corrobora e discorre que a construção de inteligências artificiais sempre suscitou controvérsias, não apenas em relação aos seus limites, mas também quanto aos objetivos a serem alcançados. Parece haver dois estilos de abordagem fundamentalmente diferentes na Inteligência Artificial: por um lado, um estilo empírico, que se baseia fortemente em observações da biologia e psicologia de seres vivos e está disposto a adotar arquiteturas complexas que surgem da interação de vários módulos distintos; por outro lado, um estilo analítico, fundamentado em princípios gerais e organizacionais, interessado em concepções abstratas da inteligência e apoiado em argumentos matemáticos e lógicos.

Alcolea (2009) expõe que na Inteligência Artificial pode-se identificar duas abordagens distintas: a primeira concebe a IA como uma tentativa de desenvolver uma tecnologia capaz de fornecer habilidades de raciocínio computacional semelhantes às da inteligência humana; a segunda aborda a IA como uma pesquisa relacionada aos mecanismos da inteligência humana, sendo estes utilizados em simulações para validar teorias.

Historicamente, a Inteligência Artificial, conforme é compreendida atualmente, começou a surgir no século XX. Em 1950, Alan Turing, matemático e lógico britânico, propôs o que agora é conhecido como o "Teste de Turing" - um método para avaliar se uma máquina é capaz ou não de demonstrar comportamento inteligente.

O teste de Turing foi concebido para oferecer uma definição operacional viável de inteligência. Um computador passaria no teste se um interrogador humano, após fazer algumas perguntas por escrito, não conseguisse distinguir se as respostas provêm de uma pessoa ou de um computador. Nesse contexto, a Inteligência Artificial registrou seus primeiros êxitos, com programas capazes de jogar xadrez e resolver problemas matemáticos, oferecendo uma perspectiva promissora para o futuro. (Russell; Norvig, 2013).

No entanto, não há um consenso sobre a definição de Inteligência Artificial. De forma simplificada, pode-se entendê-la como a capacidade das máquinas de tomar decisões inteligentes com base em dados, adaptando-se e aprimorando seu próprio funcionamento.

Conforme Spiegeleire (2017) a Inteligência Artificial é um campo da ciência da computação que busca desenvolver sistemas capazes de executar tarefas anteriormente reservadas à inteligência humana, como reconhecimento de fala, tomada de decisões, tradução de idiomas, diagnósticos médicos, entre outras. As ciências tecnológicas relacionadas à IA têm o potencial de transformar radicalmente todos os setores da sociedade, abrindo portas para inovações que anteriormente eram vistas apenas na ficção científica. Em linhas gerais, a IA não visa recriar a inteligência humana, mas sim foca em sistemas que se aproximam de um desempenho ideal característico.

Considerando isso, embora a tecnologia atualmente amplie as capacidades humanas, é crucial levar em conta o aspecto humano nessas adaptações. É fundamental evitar uma substituição excessiva pela máquina em contextos onde a capacidade humana de interpretação é essencial para a resolução de problemas. Isso se deve ao fato de que o cérebro humano, embora possa ser comparado a um computador, é substancialmente diferente da inteligência artificial.

Churchland e Churchland (2015) demonstram que o cérebro humano pode ser comparado a um tipo de computador, embora muitas de suas características permaneçam ainda desconhecidas. Essa comparação não deve ser encarada de forma trivial nem leviana. O cérebro realiza operações computacionais, muitas delas de grande complexidade, porém de uma maneira distinta da inteligência artificial clássica. Quando afirmamos que os cérebros são como computadores, não estamos sugerindo que funcionam da mesma forma que os computadores digitais em série, que são programados, que possuem uma distinção clara entre hardware e software, ou que seguem regras e manipulam símbolos da mesma forma. Os cérebros são computadores, mas em um estilo radicalmente diferente.

Russell e Norvig (2013) definem a Inteligência Artificial como o estudo de agentes que recebem percepções do ambiente e realizam ações. Cada agente implementa uma função que mapeia sequências de percepções em ações.

Nessa senda, as tecnologias, especialmente a Inteligência Artificial,

desempenham um papel indispensável na sociedade contemporânea, pois têm o poder de moldar e transformar o mundo. Elas simbolizam a interseção entre o conhecimento científico e sua aplicação prática, com o objetivo de proporcionar benefícios tangíveis para a humanidade. Contudo, é elementar introduzi-las no cotidiano de forma responsável, considerando cuidadosamente tanto os benefícios quanto os desafios que elas apresentam.

Sousa et al. (2022) decorrem que a IA é um tópico frequente em diversas discussões científicas e em previsões sobre as áreas que terão um grande impacto na sociedade em um futuro próximo. Apesar de sua associação lógica com as áreas de tecnologia e informática, o desenvolvimento científico da inteligência artificial requer uma abordagem inserida em um contexto mais amplo, que leve em consideração os aspectos éticos e humanos do avanço e da implementação dessa tecnologia. Enquanto a inteligência artificial oferece vastas oportunidades para o futuro, também existem riscos que precisam ser cuidadosamente avaliados do ponto de vista das ciências sociais, humanas e, naturalmente, do Direito.

Nesse escopo, embora as tecnologias exerçam influência e moldem a sociedade, a responsabilidade última pelo uso e integração delas no dia a dia recai sobre os seres humanos. É essencial manter um diálogo constante sobre essas questões para assegurar um futuro em que a tecnologia beneficie a humanidade de forma ética¹ e preservando os direitos fundamentais dos cidadãos.

Conforme Padhy e Padhy (2019) a Inteligência Artificial tem sido um anseio da humanidade por séculos, tanto na ficção quanto na filosofia. No entanto, com os avanços exponenciais da tecnologia nas últimas décadas, esse sonho se

¹ Desde a Grécia Antiga, a ética está relacionada à análise da vida prática, especialmente no que diz respeito às ações humanas. Agir eticamente implica tomar decisões conscientes e fundamentadas, que exigem reflexão e justificativas consistentes. No entanto, essas exigências podem entrar em conflito com as dinâmicas da vida pública. Um exemplo clássico é o dilema enfrentado por Antígona, personagem da tragédia de Sófocles, que precisa decidir entre cumprir as leis estabelecidas por Creonte, que garantem a ordem política, ou seguir os costumes religiosos tradicionais, mesmo que isso lhe custe a própria vida. A situação levanta questões como: seguir a própria consciência ou obedecer às normas do Estado? Como equilibrar legalidade e legitimidade? Ou, ainda, como conciliar o que é ético com o que é político? Em meio a interesses contraditórios, qual critério deve orientar a decisão? A palavra ética vem do grego *éthos*, que remete tanto ao comportamento humano diante das dificuldades da vida e da morte, quanto ao modo como o homem habita o mundo. Nesse sentido, *éthos* pode ser compreendido, por um lado, como o caráter individual de cada pessoa e, por outro, como sua forma de se relacionar com o ambiente, com os outros e com a sociedade. Essas duas dimensões se entrelaçam, pois o modo como cada um é em sua essência influencia diretamente a maneira como convive e interage socialmente. Dessa forma, viver eticamente é também viver em comunidade, exercendo uma forma de ser que respeita a vida coletiva e contribui para o bem comum. (Santos, 2021).

tornou realidade. Atualmente, a dependência humana da inteligência artificial aumentou significativamente. Desde carros automatizados até drones, da ciência da computação à medicina, e de assistentes de voz em telefones a sistemas de IA auxiliando advogados, é difícil encontrar alguma área da vida cotidiana que não tenha sido impactada por isso. A IA tem contribuído para tornar a vida humana mais fácil, célere e mais eficiente, poupando tempo e energia preciosos.

Considerando essa perspectiva voltada para o futuro, as tecnologias são fundamentadas no princípio do bem-estar humano. Uma gama diversificada de cenários se apresenta, dependendo das direções que se opta seguir e dos obstáculos que se enfrenta. Algumas previsões otimistas incluem melhorias significativas na saúde, uma vez que os avanços na medicina personalizada, genética e robótica podem resultar em tratamentos mais eficazes, aumento da expectativa de vida e até mesmo a possibilidade de cura para doenças anteriormente consideradas incuráveis.

Apesar das vantagens, é importante reconhecer os malefícios associados à tecnologia, os quais podem resultar em sérios prejuízos para a sociedade. Um desses malefícios é a desigualdade, pois a tecnologia pode agravá-la se não for distribuída ou acessível de forma equitativa. Isso pode contribuir para uma ampliação da divisão entre os mais ricos e os menos favorecidos economicamente.

Vilaça e Araújo (2016) comentam que desta forma, ao refletir sobre a popularização de tecnologias (como os computadores pessoais e o acesso à internet), é crucial ter em mente que sua disseminação não é verdadeiramente universal. Diversos fatores, especialmente de natureza econômica e social, continuam a dificultar ou mesmo impedir o acesso de muitas pessoas a essas tecnologias. É um equívoco pensar que todos se beneficiam desses produtos ou que tal benefício ocorre de maneira uniforme. Portanto, não se deve interpretar erroneamente que o acesso a essas tecnologias seja plenamente garantido para todos. A redução de custos, a estabilidade econômica, o aumento da oferta de crédito e as políticas de inclusão digital são, sem dúvida, alguns dos fatores que contribuíram para ampliar o acesso a esses recursos.

Outra questão contemporânea de grande relevância é a dependência tecnológica, um problema emergente que pode levar a humanidade a se tornar excessivamente dependente das tecnologias, tornando-se assim mais vulnerável a

diversos tipos de crimes digitais, além de possíveis impactos negativos na saúde mental, entre outros malefícios.

Diante desse cenário, o futuro da humanidade em relação à tecnologia será determinado pelas escolhas feitas pela sociedade. Embora a tecnologia tenha o potencial de trazer benefícios incríveis, os riscos associados não podem ser subestimados. É fundamental abordar esses desafios de maneira consciente, ainda mais, se considerar a esfera das decisões judiciais.

Os progressos na área da Inteligência Artificial continuam em um ritmo acelerado, suscitando questões éticas, sociais e filosóficas sobre o papel da IA na sociedade. Temas como vieses em algoritmos, privacidade, automação de empregos e responsabilidade nas decisões tomadas por máquinas têm sido amplamente discutidos.

Brochado (2023) afirma que o potencial impacto desestabilizador que o aumento de casos decididos por máquinas pode ter na simbologia social associada ao papel e à importância do Direito levanta questionamentos sobre a necessidade contínua do Poder Judiciário como instituição de poder, caso seja percebido simplesmente como um conjunto de programas computacionais. Isso representaria um grande prejuízo em diversas áreas, especialmente nas relações jurídicas e contratuais, onde a presença humana genuína é essencial para lidar com outros seres humanos.

Nesse contexto, a distinção entre IA forte e IA fraca faz-se importante. A IA fraca refere-se a sistemas inteligentes projetados para desempenhar tarefas específicas, incapazes de tomar decisões fora do escopo para o qual foram treinados. Por exemplo, algoritmos desenvolvidos para diagnosticar doenças não têm habilidade para jogar xadrez. A "inteligência" desses sistemas é muito limitada em comparação com a inteligência humana abrangente. Por outro lado, a IA forte envolve a criação de sistemas capazes de tomar decisões inteligentes em uma variedade de áreas. Esta representa sistemas de IA com a capacidade de entender, aprender, raciocinar e resolver problemas em vários domínios, de forma semelhante à inteligência humana, podendo generalizar o conhecimento e adaptar-se a novas situações. Atualmente, considera-se que o conceito de inteligência artificial forte ainda é teórico, estando-se mais próximo da inteligência artificial fraca ou estreita, a qual se concentra na execução de tarefas específicas e limitadas. (Ferrari, 2018).

Sistemas Inteligentes, por sua vez, dizem respeito aos sistemas

especialistas, também conhecidos como sistemas expert, que são sistemas computacionais projetados para imitar a capacidade de tomada de decisão de especialistas humanos em um campo específico. Esses sistemas são desenvolvidos para resolver problemas complexos usando um conjunto específico de conhecimentos ou regras. Com isso, Sistemas de Decisões Algorítmicas podem ser definidos como algoritmos inteligentes empregados para tomada ou o apoio a decisões. (Ferrari, 2018).

É essencial compreender as bases do funcionamento dos algoritmos inteligentes para perceber as potencialidades e os riscos associados aos chamados sistemas de decisões algorítmicas, demonstrando como eles são capazes de encontrar soluções para os problemas apresentados.

2.2 Os Algoritmos e seu funcionamento

Sichman (2021) destaca que, para compreender melhor a definição de Inteligência Artificial, é essencial entender o conceito de algoritmo, frequentemente mencionado na mídia, muitas vezes de forma imprecisa. Em essência, um algoritmo é uma sequência finita de instruções projetadas para resolver um problema específico. Ele serve como a base de inúmeros sistemas computacionais, sendo responsável por diversos processos do cotidiano digital.

Os algoritmos são conjuntos de instruções ou regras definidas para serem seguidas na resolução de um problema ou na execução de uma tarefa. Eles são a espinha dorsal de muitos sistemas computacionais e estão por trás de uma variedade de processos que encontramos em nosso dia a dia digital.

Um algoritmo pode ser definido como uma sequência finita de passos (instruções) para resolver um determinado problema. Sempre que desenvolvemos um algoritmo estamos estabelecendo um padrão de comportamento que deverá ser seguido (uma norma de execução de ações) para alcançar o resultado de um problema (Ferrari; Cechinel, s.d., p. 15).

Cumprido esclarecer que o termo "algoritmo" tem origem no nome do matemático do século IX, Muḥammad ibn Mūsā al-Khwārizmī, cuja versão latinizada

tornou-se "*Algoritmi*", mas, no entanto, o aludido conceito já era utilizado muito antes, como, por exemplo, pelos matemáticos gregos, que empregavam procedimentos sistemáticos para encontrar números primos ou determinar o maior divisor comum entre dois inteiros (Teixeira, 2019).

É importante mencionar que o algoritmo, a partir do momento que é programado, constitui-se de ao menos três partes, a saber: entrada de dados, processamento de dados e saída de dados.

Desse modo, o funcionamento dos algoritmos envolve uma série de etapas. Primeiramente, eles recebem dados de entrada, que podem ser números, textos, imagens ou qualquer outra forma de informação. Em seguida, esses dados são processados de acordo com as instruções do algoritmo. Durante esse processamento, o algoritmo pode realizar operações matemáticas, comparações, seleções ou repetições, dependendo da natureza do problema a ser resolvido.

Uma das características fundamentais dos algoritmos é sua capacidade de tomar decisões com base nos dados de entrada e nas instruções fornecidas. Isso permite que os algoritmos sejam adaptáveis e capazes de lidar com uma variedade de situações. É importante notar que os algoritmos podem variar em complexidade, desde simples cálculos matemáticos até sistemas altamente sofisticados de inteligência artificial. Além disso, eles desempenham um papel crucial em uma ampla gama de aplicações, incluindo motores de busca na internet, sistemas de recomendação de produtos, carros autônomos, entre outros.

Há diversos tipos de algoritmos, cada um com características e aplicações específicas. Entre eles, destacam-se os algoritmos de ordenação, busca e recursão. Os algoritmos de ordenação, por exemplo, são projetados para organizar elementos em uma determinada sequência. Alguns dos mais conhecidos incluem o algoritmo de ordenação por bolha, seleção, inserção e o rápido (quicksort), entre outros. (Souza; Oliveira ; Pinto, 2019)

Os algoritmos de Busca são utilizados para encontrar um elemento específico em uma coleção de dados. Exemplos incluem o algoritmo de busca binária. Dado um conjunto de elementos com identificadores únicos e uma chave de busca, o objetivo é encontrar o elemento na coleção que possui o mesmo identificador que a chave de busca, ou determinar se não há nenhum elemento com a chave fornecida. A busca binária é um método mais eficiente para esse propósito, mas exige que a lista esteja ordenada pelos valores da chave de busca. A lógica

subjacente a esse algoritmo é a seguinte: presume-se que a lista esteja organizada em ordem crescente com base nos valores da chave de busca. (Dias, 2020).

Algoritmos de recursão são algoritmos que chamam a si mesmos para resolver um problema. Um exemplo clássico é o algoritmo de recursão para calcular o fatorial de um número. Segundo Dias (2023), por meio do método de recursão ou indução, a solução de um problema pode ser articulada da seguinte forma: primeiramente, estabelece-se a solução para os casos base; em seguida, determina-se como resolver o problema em um caso geral, utilizando-se, como fundamento, as soluções de instâncias menores do mesmo problema.

Ferrari (2018) discorre que no que diz respeito ao seu funcionamento, os algoritmos podem ser classificados em duas categorias: programados e não programados. Algoritmos programados seguem as operações e o caminho definidos pelo programador. Nesse caso, a informação entra no sistema, o algoritmo processa essa informação e o resultado, ou output, é gerado pelo sistema. O programador controla todas as etapas operacionais do algoritmo. Dado que o processamento é determinado pelo programador, que tem conhecimento completo do seu itinerário, quaisquer problemas nos resultados dos algoritmos programados podem ser atribuídos à fonte: seja a programação ou os dados utilizados para alimentá-los. Mesmo que decisões inteligentes sejam tomadas autonomamente pela máquina, escolhendo entre diferentes cursos de ação, é importante observar que o algoritmo programado não adquiriu conhecimento por si só; ele foi ensinado sobre o que fazer. Portanto, a definição de sua operação, embutida no código, torna o processo compreensível. Essa categoria de algoritmos não é relevante para o presente trabalho.

Por outro lado, conforme já mencionado, Alan Turing, em 1950, propôs uma abordagem diferente para a operação de algoritmos. Em vez de imitar o cérebro de um adulto, programando todas as operações a serem realizadas, Turing sugeriu que seria mais produtivo simular o cérebro de uma criança, com capacidade de aprendizado aleatório. Alan Turing explorou a questão fundamental de saber se as máquinas poderiam pensar. Ele também levantou a questão sobre a viabilidade de programar computadores de forma a imitar o cérebro de uma criança, dotando-os de capacidade autônoma de aprendizado, em vez de simplesmente inserir todas as informações disponíveis diretamente nos sistemas computacionais. Esse conceito

deu origem à ideia dos algoritmos não programados, que utilizam técnicas como a aprendizagem de máquina, ou machine learning (ML). (Ferrari,2018).

Este trabalho tratará das ferramentas pertencentes ao segundo grupo mencionado, o dos algoritmos não programados, que utilizam predominantemente machine learning. Os algoritmos de aprendizado de máquina são tipos de algoritmos não programados. Assim, em certa medida, seguem essa lógica de programação autônoma. Para compreender suas capacidades e os riscos relacionados aos sistemas de decisões algorítmicas, é necessário ter uma compreensão dos princípios de funcionamento desses algoritmos.

2.2.1 Machine Learning, Deep Learning e Redes Neurais Artificiais

A capacidade de adquirir conhecimento de maneira artificial, permitindo que os sistemas aprendam a agir com base em seu conjunto de dados, sem serem explicitamente programados para tarefas específicas, é conhecida como aprendizado de máquina (machine learning). Em resumo, por meio do aprendizado de máquina, quando dados são fornecidos ao algoritmo do sistema, ele analisa essas informações e aprende como executar tarefas por conta própria, sem a necessidade de intervenção humana direta, apenas com base em experiências anteriores e treinamento, eliminando a necessidade de criar códigos com instruções específicas. (Ferrari, 2018).

No progresso da Inteligência Artificial, o aprendizado de máquina é um subcampo dedicado ao desenvolvimento de algoritmos e modelos que capacitam computadores a aprimorar seu desempenho em tarefas específicas por meio da experiência, geralmente representada por dados. Em vez de serem instruídos explicitamente sobre como executar uma tarefa, os computadores aprendem com os dados fornecidos. (Alcolea, 2009).

Devido à abordagem iterativa do aprendizado de máquina, o sistema utiliza repetições do processo para aprender com os dados, o que possibilita sua automação. Assim, ele executa a atividade de forma contínua até alcançar uma solução precisa. Isso permite que o software se ajuste a novas circunstâncias, antecipando respostas e resolvendo problemas que não foram previamente

encontrados. Essa capacidade é baseada exclusivamente nos dados fornecidos durante o treinamento. (Ferrari,2018)

A técnica de aprendizado de máquina é uma abordagem antiga, que teve suas primeiras aplicações na década de 1950. Embora tenha sido difundida desde os anos 1970, somente recentemente ganhou destaque devido às numerosas inovações que apoiou, como sistemas de recomendação de compras, conexões entre pessoas em redes sociais ou aplicativos de relacionamento, pontuação de crédito, e sistemas de otimização da força de trabalho. O aumento significativo na relevância das técnicas de aprendizado de máquina é atribuído a duas condições essenciais para sua operação: o surgimento do big data e o crescimento exponencial na capacidade computacional. (Ferrari, 2018)

Big data refere-se a conjuntos de dados volumosos, complexos e variados, que são gerados continuamente a partir de diversas fontes, como transações comerciais, interações em redes sociais, dispositivos móveis e sensores. Esses dados apresentam desafios em termos de captura, armazenamento, processamento e análise. O crescimento do big data na atualidade é impulsionado pela expansão da digitalização, pela proliferação de dispositivos conectados à internet e pela crescente demanda por análises de dados para obter insights e tomar decisões informadas em diversos setores. (Ferrari,2018).

A profusão de dados estruturados e não estruturados², conhecida com o big data, é a matéria-prima essencial para os sistemas que utilizam machine learning aprenderem. Essas informações são processadas por computadores, que se tornaram cada vez mais poderosos e acessíveis ao longo do tempo. Como resultado, a combinação de técnicas de machine learning, alimentadas por grandes volumes de dados e executadas por computadores poderosos, tem impulsionado transformações significativas na sociedade humana, e espera-se que essas mudanças se aprofundem nos próximos anos. (Ferrari,2018).

Assim, a machine learning tem uma ampla gama de aplicações, que incluem reconhecimento de voz, assistentes virtuais, sistemas de recomendação,

² Dados estruturados são aqueles organizados em formatos padronizados (por exemplo, tabelas em bancos de dados relacionais), permitindo fácil indexação, consulta e análise. Já os Dados não estruturados, por sua vez, não seguem uma estrutura fixa e podem incluir textos livres, vídeos, áudios, imagens e outros formatos mais complexos. Sua análise demanda técnicas avançadas, como processamento de linguagem natural ou sistemas de reconhecimento de imagem, para extrair insights de forma efetiva (Enap, 2020).

entre outros. Seu uso está transformando significativamente o mundo em que vivemos, impulsionando inovações em diversos setores e facilitando o desenvolvimento de sistemas capazes de se adaptar e melhorar continuamente ao longo do tempo.

Destacam-se os padrões de aprendizagem supervisionado, não supervisionado e por reforço. Fontana (2020) apresenta que os algoritmos de aprendizagem supervisionada relacionam uma saída com uma entrada com base em dados rotulados. Nesse contexto, o usuário fornece ao algoritmo pares de entradas e saídas conhecidas, geralmente na forma de vetores. Cada saída é associada a um rótulo, que pode ser um valor numérico ou uma classe. O algoritmo então é treinado para prever o rótulo de saída com base em uma entrada fornecida. Além disso, os dados são previamente incluídos no sistema juntamente com o resultado esperado.

Ao carregar os dados refinados junto com os rótulos dos resultados desejados, o algoritmo pode identificar diversas correlações entre as características desses resultados. Isso permite que o algoritmo “aprenda” quais informações estão relacionadas a cada resultado. Após o treinamento, o algoritmo pode então fornecer probabilisticamente as chances do resultado pretendido com base em novos inputs. (Ferrari,2018).

Desse modo, nos algoritmos supervisionados, em que há disponibilidade de informações sobre as saídas desejadas, os conjuntos de dados e as saídas desejadas são carregados no sistema. Durante o treinamento, o modelo adapta suas variáveis para correlacionar as entradas com as saídas correspondentes. (Ferrari,2018).

Os algoritmos de aprendizagem não supervisionada ,por sua vez, não recebem rótulos para os dados de saída. Eles exploram um grande conjunto de dados em busca de padrões e similaridades entre os dados, o que permite identificar grupos de itens semelhantes ou encontrar similaridades entre novos itens e grupos já existentes. Esses algoritmos são capazes de agrupar amostras sem a necessidade de classes predefinidas (clusterização) .Na aprendizagem não supervisionada, a distribuição dos dados é modelada para compreender os relacionamentos entre as entradas. (Ferrari,2018).

Esses algoritmos podem ser categorizados como algoritmos de transformação, que buscam representar os dados de forma mais compacta, e algoritmos de agrupamento, que procuram dividir os dados em clusters ou grupos.

Nos algoritmos de reforço, um agente realiza ações em um ambiente e recebe uma recompensa com base no resultado dessas ações. O objetivo do algoritmo é aprender a realizar ações que maximizem a recompensa recebida. Esses sistemas são compostos por três componentes principais: o agente, que toma as decisões, o ambiente em que as ações são realizadas e a interação entre o agente e o ambiente. (Fontana, 2020).

Após abordar o funcionamento de cada tipo de aprendizagem, faz-se relevante fazer uma distinção entre algoritmos e Modelos.

Como observado, a inteligência artificial é capaz de criar sua própria programação. Para isso, além dos dados, que são a matéria-prima da aprendizagem, é essencial selecionar um algoritmo como base para o treinamento. Esses algoritmos podem assumir diferentes formas, como redes neurais, árvores de decisão e florestas randômicas.

O algoritmo de treinamento é simplesmente a base matemática utilizada, e portanto não tem a capacidade de carregar preconceitos ou vieses, ou gerar efeitos discriminatórios por si só. Os modelos algorítmicos, por sua vez, podem carregar vieses, especialmente se forem treinados com conjuntos de dados que contenham vieses inerentes.

Um algoritmo é uma sequência de instruções lógicas e passos definidos que orientam a resolução de problemas ou a execução de tarefas específicas. Ele representa o método ou processo utilizado para alcançar um objetivo particular, sem necessariamente lidar com dados específicos. Algoritmos são mais abstratos e genéricos, e podem ser aplicados a uma variedade de situações.

Por outro lado, um modelo algorítmico é uma representação específica e aplicada de um algoritmo em um contexto particular, geralmente relacionado à análise de dados. Este modelo é construído usando um algoritmo específico de aprendizado de máquina ou estatística, aplicado a um conjunto específico de dados. O modelo algorítmico captura padrões e relações nos dados e é usado para fazer previsões, classificações ou inferências sobre novos dados. Assim, enquanto um algoritmo é uma abstração geral de um método ou processo, um modelo algorítmico

é uma aplicação específica e concreta desse método ou processo em um conjunto de dados específico.(Ferrari,2018).

Desse modo, para que um modelo de Aprendizado de Máquina (ML) possa aprender, é necessário fornecer-lhe dados. Os algoritmos de aprendizado desempenham um papel fundamental nesse processo, pois são responsáveis por guiar o modelo no aprendizado a partir dos dados disponíveis. Esses dados são utilizados para treinar o modelo, fornecendo indicações sobre como ele deve se comportar, por exemplo.

Após a seleção do algoritmo de treinamento e a análise dos dados disponíveis, é comum que os programadores dividam esses dados em dois conjuntos distintos: o conjunto de dados de treinamento e o conjunto de dados de teste. O conjunto de dados de treinamento é utilizado para ensinar o algoritmo de treinamento. Durante essa interação entre dados e algoritmo, o algoritmo é capaz de criar múltiplas correlações a partir dos dados fornecidos. Esse processo resulta em um segundo algoritmo, conhecido como algoritmo treinado ou modelo algorítmico.(Ferrari,2018).

Conforme descrito, os algoritmos de treinamento são considerados neutros, pois são simplesmente bases matemáticas que consistem em um conjunto de regras sem qualquer julgamento de valor ou viés. Por outro lado, os modelos, que também são algoritmos, podem ter incorporado vieses devido a escolhas humanas feitas durante o processo de treinamento, relacionadas aos dados utilizados ou à sua programação, seja de forma consciente ou inconsciente. Portanto, percebe-se que o que é empregado para a tomada de decisões é o modelo .

Após ser treinado no primeiro conjunto de dados, o modelo passa por avaliação em um conjunto de teste para verificar sua capacidade de generalização, medida em termos de acurácia. Esse procedimento, chamado de validação cruzada, permite estimar o quão bem o algoritmo se adapta a novos dados (Ferrari, 2018).

É comum, entretanto, sustentar que o modelo atende às finalidades para as quais foi criado apenas por apresentar um alto índice de acurácia, geralmente superior a 95%. Contudo, essa métrica se baseia estritamente na comparação entre o conjunto de treinamento e o de teste, não capturando possíveis vieses ou falhas existentes em ambos. Em outras palavras, mesmo que o algoritmo alcance resultados expressivos em termos de acurácia com base nos dados disponíveis, isso não garante que ele seja igualmente eficaz – ou justo – ao lidar com situações reais,

fora daquele contexto (Ferrari, 2018).

Assim, diferentemente do que possa parecer à primeira vista, a acurácia por si só não é suficiente para uma avaliação abrangente do desempenho do modelo, pois este envolve aspectos que vão além da simples validação interna. Para contemplar a efetiva capacidade do algoritmo de responder aos problemas do mundo real, seria necessário adotar métricas mais amplas de desempenho. Nesse sentido, Cortiz (2020, p. 45) frisa a importância de “definir novas métricas de avaliação para os modelos de IA que levem em consideração não apenas requisitos técnicos, mas aspectos sociais”.

Sobre isso, Ferreira et al (2022) afirmam que para avaliar a eficácia de um modelo, é comum realizar fases de testes para avaliar sua performance em dados que não foram utilizados durante o treinamento. No entanto, o processo enfrenta desafios complexos, como o overfitting e o underfitting, que são comuns no aprendizado de máquina. O overfitting ocorre quando um modelo é excessivamente complexo e memoriza os dados de treinamento, mas falha em generalizar para novos dados. Por outro lado, o underfitting ocorre quando o modelo é muito simples para capturar os padrões subjacentes nos dados. É importante observar que, ao utilizar os mesmos dados para treinamento e teste, o algoritmo pode apenas reproduzir os resultados já observados e otimizados durante o treinamento. O overfitting também pode ocorrer quando o modelo é muito complexo e tenta capturar todas as nuances dos dados, como no caso de utilizar uma função polinomial de alto grau para ajustar os dados. Por outro lado, o sub-ajuste (underfitting) acontece quando o modelo é incapaz de capturar os padrões dos dados, seja por sua simplicidade, como utilizar uma função linear para descrever dados não lineares, ou devido à falta de dados suficientes para o modelo aprender de forma adequada.

Martins e Schor (2021) apontam que muitos pesquisadores ainda não conseguem elucidar completamente os processos que levaram os algoritmos inteligentes a determinadas conclusões, o que gera certa desconfiança no uso da inteligência artificial. A falta de transparência nessas situações, em que não compreendemos completamente o funcionamento interno dos algoritmos, é comumente referida como "caixas-pretas". Esse conceito implica a ausência de acesso às informações detalhadas sobre o design e a implementação dos algoritmos, o que pode minar a confiança dos profissionais da área.

Diante disso, algumas questões merecem ser levantadas, a privacidade tem sido um tema constante de discussão, especialmente com o avanço das redes sociais e tecnologias de vigilância. A privacidade das pessoas está cada vez mais comprometida de maneiras antes inimagináveis. Sem uma regulamentação adequada e práticas éticas por parte das empresas, informações pessoais podem ser exploradas para ganhos financeiros ou até mesmo propósitos maliciosos. No entanto, quando se trata de investigações criminais, as mesmas empresas muitas vezes encontram obstáculos em fornecer dados que podem ser vitais para tais investigações.

Rossetti e Angeluci (2021) afirmam que os proprietários dos algoritmos mantêm segredo de seu funcionamento, seja por questões de competitividade, de segurança nacional e de privacidade informacional. Quando a relação entre os dados e seus resultados não é acessível ou inteligível para escrutínio, a opacidade conduz a uma espécie de algoritmo “caixa-preta”. Tem-se aqui um conflito ético entre o princípio da transparência e os princípios da privacidade e da autonomia. A opacidade do algoritmo do aplicativo Whatsapp foi a principal razão de diversos bloqueios temporários ao seu serviço no território brasileiro em 2015. Na época, o Poder Judiciário não conseguia obter do Facebook, proprietário do aplicativo, dados e informações trocados entre criminosos e que poderiam auxiliar na resolução de crimes. O Facebook justificava que os dados eram criptografados e não eram armazenados pela empresa.

Neste contexto, a complexidade de alguns modelos de IA dificulta a compreensão de como determinadas decisões são alcançadas. Essa falta de transparência pode representar um desafio na responsabilização dos sistemas e de seus desenvolvedores por eventuais erros ou decisões prejudiciais.

Sousa et al. (2022) afirmam que alguns algoritmos operam como verdadeiras “caixas-pretas”, processando dados de entrada e gerando resultados de forma opaca. Em outras palavras, a complexidade da estrutura algorítmica dificulta a compreensão do processo que leva a um resultado específico, mesmo quando o modelo apresenta alta precisão na geração de novas informações. Quanto mais avançados e sofisticados forem os modelos de aprendizado de máquina, maior será essa opacidade. Em certos casos, a tecnologia atual ainda não permite uma compreensão completa sobre como um resultado foi alcançado, seja ele uma previsão, recomendação ou decisão.

Os algoritmos de machine learning, especialmente aqueles que utilizam deep learning e redes neurais artificiais, são conhecidos por sua natureza de "caixa-preta".

Redes neurais artificiais são um subconjunto do machine learning, inspiradas no funcionamento dos neurônios biológicos, e são compostas por camadas de neurônios artificiais que processam informações de forma hierárquica. Essas redes são organizadas em camadas e são capazes de processar informações e aprender com exemplos fornecidos durante o treinamento. Eles são amplamente utilizados para tarefas de reconhecimento de padrões, classificação, regressão e outras aplicações em áreas como reconhecimento de fala, processamento de imagem e muitos outros.(Ferrari,2018).

O aspecto distintivo das redes neurais é que o peso atribuído às conexões entre os neurônios é ajustável, o que significa que pode ser modificado com base na experiência, ou seja, no processo de aprendizagem. Ou seja, os trajetos que, seguindo uma lógica de tentativa e erro, conduziram aos resultados corretos adquirem maior importância no cálculo matemático subjacente a essas aplicações. (Ferrari,2018).

Com isso, diz-se que o algoritmo cria sua própria programação e ,independentemente do método de aprendizado utilizado, seja supervisionado ou não supervisionado, a programação em redes neurais ocorre através da formação de diversas correlações, que são desenvolvidas nas camadas internas do software. Essas correlações são estabelecidas de forma autônoma pelo próprio sistema, sem intervenção direta do programador nessa atividade específica. (Ferrari,2018).

Por sua vez, Deep Learning (Aprendizado Profundo) é um subcampo do Machine Learning que utiliza redes neurais artificiais com múltiplas camadas para analisar e extrair padrões complexos de grandes volumes de dados. Ou seja, essas redes neurais profundas simulam, em certa medida, a estrutura do cérebro humano, permitindo que os algoritmos aprendam representações hierárquicas e progressivamente mais abstratas das informações.

A deep learning segundo Leite (2022) é uma técnica no campo do aprendizado de máquina que tem alcançado resultados significativos em diversas tarefas, superando outras técnicas na área. Entre essas tarefas, a detecção e classificação de objetos em imagens se destacam como exemplos notáveis de sucesso. Geralmente, nesse tipo de aplicação, uma única rede neural convolucional

é empregada para realizar o processo de detecção das regiões de interesse, delimitando assim a área que contém os objetos a serem identificados, além de classificar essa região em uma classe específica.

Hartmann e Zumblick (2019,p .32) afirmam que “a profundidade imensa na formação dos conceitos vem com o custo de interpretabilidade do modelo: quanto mais camadas, mais abstrata é a representação do modelo e mais ininteligíveis são as features” .

Mesmo ao utilizar algoritmos mais simples e independentemente do paradigma de aprendizado de máquinas empregado, observar a dinâmica operacional descrita acima torna evidente a perda de controle, em certo grau, sobre os processos decisórios desses algoritmos. Conseqüentemente, a autonomia dos algoritmos que fazem uso de aprendizado de máquina torna as tarefas por eles realizadas difíceis de prever e, mesmo após a tomada de decisão, difícil de explicar. Além disso, à medida que os algoritmos se tornam mais complexos e começam a interagir entre si, a tendência é que esse desafio se torne ainda mais acentuado. (Ferrari,2018).

Um dos principais riscos decorrentes do uso da inteligência artificial no processo de tomada de decisão judicial é a possibilidade de viés algorítmico. Os algoritmos utilizados nos sistemas de inteligência artificial são desenvolvidos com base em dados históricos, que podem refletir preconceitos e desigualdades presentes na sociedade. Isso significa que as decisões tomadas por esses algoritmos podem reproduzir e amplificar esses vieses, resultando em discriminação e injustiça (Ferreira, 2020).

Os algoritmos podem perpetuar preconceitos e desigualdades existentes na sociedade. Isso pode resultar em decisões injustas e discriminatórias, especialmente em áreas como justiça criminal, contratação e crédito. Outra preocupação é a falta de controle humano sobre as decisões dos algoritmos.

À medida que confiamos cada vez mais em algoritmos para automatizar processos e tomar decisões importantes, corremos o risco de perder a capacidade de questionar e contestar essas decisões. Isso pode minar a autonomia e a liberdade das pessoas, criando uma sociedade onde as escolhas individuais são cada vez mais influenciadas por algoritmos opacos sem supervisão. Por isso a questão merece ser mais debatida. “Uma das principais fontes de preconceito está no conjunto de dados utilizados para gerar modelos, como é o caso das aplicações

que geram modelos por meio da aplicação de algoritmos aprendizados.” (Carvalho, 2021, p. 29).

Os algoritmos de inteligência artificial são treinados com conjuntos de dados do mundo real, os quais podem conter vieses culturais, sociais ou preconceitos incorporados. De acordo com Higasi (2020) essa situação poderia levar a insultos e comentários discriminatórios, como machismo, homofobia ou racismo. Isso ocorre devido à interação potencial com indivíduos que expressam agressividade verbal e atitudes discriminatórias. O bot poderia ser fortemente influenciado por essa interação, reproduzindo tais comportamentos como uma forma de se identificar com as pessoas com as quais interage regularmente.

A criação e o desenvolvimento de um software de inteligência artificial se baseia e apoia na utilização de bancos de dados. Após a criação do modelo, são disponibilizados dados para o sistema, permitindo assim o processo de machine learning (aprendizado de máquina). Nesse processo, a máquina examina as informações fornecidas, seguindo as instruções definidas pelo algoritmo, a fim de identificar padrões e, com base nisso, realizar previsões. A qualidade dos dados fornecidos aos sistemas de inteligência artificial também terá um impacto significativo nos resultados. Isso ocorre porque os dados são provenientes da sociedade, que por sua vez é caracterizada por desigualdades, exclusões e discriminações. (Mendonça Junior ; Nunes, 2023)

O aprendizado de máquina tem a capacidade de validar padrões discriminatórios; se esses padrões estiverem presentes nos dados, o sistema de classificação, conseqüentemente, os replicará de forma precisa. Um exemplo emblemático que ilustra essa questão é o caso do robô Tay, um sistema de inteligência artificial desenvolvido pela Microsoft com o propósito de interagir com os usuários do Twitter. A ideia era que o programa fosse alimentado com as mensagens enviadas pelos usuários da plataforma. No entanto, em questão de horas, o robô começou a emitir mensagens xenofóbicas, racistas e antisemitas. (Bahia; Nunes; Pedron, 2021)

Na era das tecnologias avançadas, é crucial examinar se o uso da IA promove o direito à proteção de dados ou, pelo contrário, contribui para práticas discriminatórias, afetando os direitos individuais ou coletivos. A reflexão sobre a discriminação é atual e pertinente, uma vez que os sistemas de IA estão sendo amplamente empregados em diversos países com uma variedade de propósitos. Por

exemplo, o policiamento preditivo utiliza análise de dados disponíveis para tentar prever onde crimes podem ocorrer. (Lima ;Sá, 2020).

Além do mais, a rápida evolução da tecnologia de inteligência artificial e aprendizado de máquina levanta preocupações sobre o desemprego e a desigualdade econômica. À medida que mais empregos são automatizados por algoritmos, há o risco de uma maior polarização econômica e social, com consequências significativas para a estabilidade e coesão da sociedade.

Embora os algoritmos ofereçam muitos benefícios e oportunidades, é imprescindível abordar seus aspectos negativos e mitigar os riscos associados à sua utilização. Isso requer uma abordagem cautelosa e responsável, que leve em consideração não apenas a eficiência e a conveniência, mas também os princípios éticos, a equidade e o bem-estar humano, e ,como será tratado neste trabalho, os riscos a igualdade e imparcialidade devido a existência de vieses algorítmicos.

Portanto, é essencial garantir a transparência, responsabilidade e ética no desenvolvimento e na implementação de algoritmos para garantir que eles funcionem de maneira justa e imparcial para todos, ainda mais, levando em conta para essa pesquisa, decisões jurídicas, que podem prejudicar ou lesar os indivíduos de forma substancial.

Diante desses riscos, torna-se necessário um debate ético acerca do uso da inteligência artificial no processo judicial. É preciso considerar os possíveis vieses e discriminações presentes nos algoritmos utilizados, bem como as consequências dessas decisões para os indivíduos envolvidos no processo judicial. Além disso, é fundamental garantir que as decisões tomadas por máquinas sejam transparentes e passíveis de revisão, evitando assim a perpetuação de injustiças (Damilano, 2021).

2.2.2 Inteligência Artificial Generativa

A Inteligência Artificial Generativa (IA Gen) é baseada no Deep Learning e utiliza modelos e algoritmos capazes de criar novos dados que se assemelham aos dados de treinamento, mas sem replicá-los exatamente. Para isso, o modelo aprende a distribuição probabilística dos dados de entrada, permitindo a geração de novas instâncias que seguem essa mesma estrutura. Em essência, a IA generativa

identifica padrões, características e a organização dos dados fornecidos e, a partir desse conhecimento, produz conteúdo inédito e original.

Segundo Grossman et al. (2023) a Inteligência Artificial Generativa (IA Gen) é um ramo específico da inteligência artificial projetado para criar novos conteúdos a partir do treinamento em grandes volumes de dados, geralmente extraídos da Internet, em resposta a comandos de usuários. Tanto o prompt quanto o conteúdo gerado podem incluir texto, imagens, áudio ou vídeo. O rápido avanço da IA generativa surpreendeu o público devido à sua habilidade em executar tarefas criativas, como escrever poesias e produzir ilustrações, além de sua capacidade de gerar conteúdos sintéticos que se assemelham a interações humanas reais. Outra mudança significativa foi a evolução na fluência da linguagem dos modelos de IA. Até poucos anos atrás, os modelos de linguagem frequentemente perdiam o contexto de diálogos em andamento. No entanto, as tecnologias atuais demonstram uma compreensão mais sofisticada da linguagem, sendo capazes de produzir frases quase indistinguíveis das de um ser humano e de manter a coerência ao longo de uma conversa, "lembrando" informações previamente mencionadas com maior precisão.

Duque-Pereira e Moura (2023, p.11) complementam que:

A IA Gen representa uma evolução notável no campo da aprendizagem profunda. Essencialmente, ela opera por meio de redes neurais que processam tanto dados rotulados quanto não rotulados, abrangendo abordagens supervisionadas, não supervisionadas e semi-supervisionadas. O cerne da IA Gen é sua capacidade de gerar novos conteúdos, fundamentados em aprendizados de conjuntos de dados anteriores. Esta capacidade não surge aleatoriamente, mas é moldada por sofisticados Modelos estatísticos e equações matemáticas.

Ao interagir com uma IA Generativa por meio de um comando (*prompt*), sua resposta não se restringe à simples reprodução de informações previamente armazenadas. Em vez disso, a IA utiliza modelos estatísticos para prever e elaborar um output coerente com o contexto e a sequência de palavras fornecidas. Essa capacidade preditiva evidencia a sofisticação e a abrangência do treinamento pelo qual o modelo passou, permitindo-lhe gerar respostas contextualmente adequadas (Duque-Pereira; Moura, 2023).

Segundo SAP (2023), os principais modelos de IA generativa seguem

etapas semelhantes às do Machine Learning, embora produzam resultados distintos. O processo inicia-se com a coleta de dados, na qual se reúne um grande volume de informações que ilustre o tipo de conteúdo almejado (por exemplo, uma base de imagens para gerar figuras realistas ou textos para elaborar sentenças coerentes). Em seguida, ocorre o treinamento do modelo, conduzido por redes neurais, para que ele assimile padrões e estruturas fundamentais dos dados coletados. Após esse período de aprendizagem, inicia-se a geração de conteúdo: o modelo, a partir de amostras do espaço latente ou de uma rede geradora (dependendo do tipo de algoritmo empregado), cria novos materiais como síntese do que foi absorvido durante o treinamento. Por fim, há o refinamento, fase em que o conteúdo produzido pode ser aperfeiçoado ou receber ajustes específicos, visando aprimorar sua qualidade ou adequá-lo a exigências particulares.

A IA Generativa pode ser dividida em diferentes categorias, cada uma com características específicas e aplicações mais adequadas. Entre os modelos mais utilizados estão os Autoencodificadores Variacionais (VAEs), Redes Adversariais Gerativas (GANs), Máquinas de Boltzmann Restritas (RBMs) e os Modelos de Linguagem Baseados em Transformadores (Transformers) . No contexto da geração de textos, os modelos baseados em transformadores, como o GPT-3 e o GPT-4, desempenham um papel essencial. Sua arquitetura permite considerar o contexto completo das informações de entrada, possibilitando a produção de textos altamente coerentes e contextualmente apropriados (SAP, 2023).

Mesmo que os modelos baseados em transformadores, bem como os modelos VAEs e GANs, sejam os mais amplamente usados na atualidade para aplicações de IA generativa, há outros modelos que também se destacam. Entre eles estão os modelos autorregressivos, que estimam pontos de dados futuros tomando como base os valores anteriores, e os modelos de fluxo normalizante, que utilizam uma sucessão de transformações para representar distribuições de dados mais complexas (SAP, 2023).

O Processamento de Linguagem Natural (PLN) permitiu avanços significativos no desenvolvimento dos Modelos de Linguagem, que aprimoram a capacidade dos sistemas computacionais de prever palavras dentro de um contexto. Esses modelos são treinados com grandes volumes de dados, utilizando métodos probabilísticos e cálculos matemáticos para gerar previsões mais precisas. Com o progresso da tecnologia, surgiram os Large Language Models (LLMs), que possuem

bilhões ou até trilhões de parâmetros, aumentando sua capacidade de processamento e precisão (Carle, 2024).

Embora PLN e LLMs sejam termos frequentemente tratados como sinônimos, é importante ressaltar que os LLMs podem ser vistos como uma evolução dos modelos de processamento de linguagem natural, apresentando tanto semelhanças quanto distinções em relação a esses. O PLN abrange métodos e técnicas voltados à interação entre computadores e a linguagem humana, objetivando que as máquinas sejam capazes de entender, interpretar e responder de forma próxima à compreensão humana (Duque-Pereira e Moura, 2023). Já os LLMs, por sua vez, são grandes modelos de linguagem que, em geral, utilizam redes neurais profundas com um elevado número de parâmetros. Esses modelos são treinados com grandes quantidades de texto não rotulado, por meio de aprendizado de máquina autossupervisionado (Digiampietri, 2024).

Os Large Language Models (LLMs) são essenciais para a IA Generativa, pois vão além das funções de classificação, sugestão e reconhecimento de padrões dos modelos mais simples. Sua principal característica é a capacidade de criar conteúdos a partir do conhecimento adquirido durante o treinamento. Além de replicar informações, esses modelos realizam inferências, integram dados de diferentes fontes e produzem respostas originais e contextualmente coerentes (Carle, 2024).

O ano de 2024 tem sido marcado por uma série de inovações e aprimoramentos em modelos de IA generativa. Entre os exemplos mais difundidos baseados em Large Language Models (LLMs) estão o GPT (Generative Pre-trained Transformer) da OpenAI, o Gemini do Google, o Claude da Anthropic, o Cohere, o Phi-3 da Microsoft e o LLAMA da Meta. Esses modelos se destacam por aprimorar a interação por meio de interfaces de linguagem natural, como chats, tornando a tecnologia mais acessível e permitindo que usuários comuns aproveitem seus benefícios sem a necessidade de conhecimentos especializados em inteligência artificial ou áreas correlatas.

A Inteligência Artificial Generativa tem se destacado por sua capacidade inovadora de compreender e produzir linguagem natural, sendo o ChatGPT (Chat Generative Pre-Trained Transformer) um dos principais exemplos desse avanço. Desenvolvido pela OpenAI, uma empresa especializada em pesquisa e desenvolvimento de IA, esse chatbot foi projetado para interagir de maneira fluida e

natural, simulando conversas humanas. Diferente dos chatbots tradicionais, que costumam ser repetitivos e limitados, e das assistentes pessoais automatizadas, o ChatGPT apresenta uma abordagem mais sofisticada ao gerar respostas contextualizadas e coerentes (Santos, 2022). Uma das características mais notáveis desse modelo de linguagem (LLM) é sua capacidade de não apenas recuperar informações, mas também criar respostas originais, demonstrando uma espécie de "raciocínio" e "reflexão". Em vez de depender de respostas pré-programadas, ele gera conteúdos dinamicamente com base em seu treinamento. Além disso, embora tenha sido previamente alimentado com um grande volume de textos, o ChatGPT pode ajustar-se progressivamente conforme interage com os usuários, tornando-se cada vez mais preciso e adaptável (Duque-pereira ; Moura, 2023). Os modelos de IA generativa, como o ChatGPT, possuem uma ampla gama de aplicações. Eles podem ser utilizados para escrever e-mails, preencher formulários, estruturar e organizar textos, criar resumos e sínteses, traduzir conteúdos, revisar literatura acadêmica de forma sistemática, transcrever áudios e até mesmo desenvolver e corrigir códigos de programação. Todas essas tarefas podem ser realizadas de maneira eficiente a partir de comandos simples, tornando essa tecnologia uma ferramenta versátil e poderosa para diversas áreas do conhecimento (Alexopoulos, 2023).

A base estrutural do ChatGPT é a arquitetura Transformer, que utiliza mecanismos de atenção para examinar diferentes partes de uma entrada, permitindo que o modelo considere o contexto global ao formular uma resposta. Essa capacidade de produzir conteúdo coeso e relevante, seja ao completar uma sentença ou ao elaborar um texto mais extenso, evidencia o avanço da IA Generativa. Os Transformers conseguem direcionar a atenção para distintas seções de um texto, possibilitando que o ChatGPT compreenda e desenvolva respostas adequadas ao contexto apresentado (Duque-pereira ; Moura, 2023). Dessa maneira, o ChatGPT vai além da simples identificação de padrões linguísticos, demonstrando a habilidade de interpretar o contexto, captar metáforas e lidar com ambiguidades. Além disso, ele consegue diferenciar palavras polissêmicas—termos com múltiplos significados—determinando seu sentido mais adequado com base na forma como são empregadas no discurso.

Duque-Pereira e Moura(2023) destacam que:

O chatGPT introduz uma inovação significativa no campo da Inteligência

Artificial: a capacidade de incorporar feedback humano diretamente em seu processo de aprendizado. Este modelo opera por meio de um sistema de aprendizagem por reforço, onde os usuários desempenham um papel crucial ao indicar as respostas mais adequadas e pertinentes às suas consultas. Esse mecanismo de feedback contínuo não apenas otimiza a performance do chatGPT, mas também aprimora sua capacidade de interagir de forma mais natural e coerente com os usuários. Assim, o chatGPT não é apenas um chatbot convencional, mas representa um avanço na simulação da língua humana, estabelecendo-se como uma ferramenta que se aproxima cada vez mais da complexidade e fluidez da comunicação natural.

Os modelos de linguagem utilizam mecanismos para prever quais palavras ou símbolos podem suceder em uma sentença. Embora esse princípio pareça simples, o processo em si é bastante complexo, demandando conhecimento aprofundado tanto da linguagem quanto do contexto geral. Para alcançar tais previsões, os modelos são treinados em grandes volumes de textos, valendo-se de modelos generativos que consistem em conjuntos de regras e instruções capazes de aprender como as palavras normalmente se combinam. Em seguida, esses padrões são empregados para projetar como novos textos podem ser produzidos, buscando imitar ao máximo a forma como as palavras são usadas na prática, mesmo que seja impossível capturar todas as nuances e complexidades da linguagem (Almeida, 2023).

Apesar do alto desempenho dos modelos de linguagem de grande porte (LLMs) em diversas tarefas, eles ainda apresentam certas limitações, como falhas de raciocínio e restrições em sua base de conhecimento. Além disso, algumas dessas IAs podem gerar respostas imprevisíveis, até mesmo para seus desenvolvedores, o que levanta questionamentos sobre a legalidade e a precisão dos conteúdos que produzem (Ganguli, 2022). Isso ocorre porque os modelos dificilmente repetem uma resposta de forma idêntica e, em algumas situações, os textos gerados podem ser incoerentes, gramaticalmente falhos ou até mesmo incorretos. Em determinados casos, esses modelos afirmam com convicção informações completamente inventadas, podendo incluir declarações falsas e potencialmente difamatórias sobre indivíduos reais. Em maio de 2024, pesquisadores da Universidade de Purdue divulgaram, na “CHI Conference on Human Factors in Computing System”, um estudo apontando que 52% das respostas fornecidas pelo ChatGPT estavam incorretas. Além disso, 77% das respostas geradas eram mais detalhadas do que as produzidas por humanos e 78% apresentavam algum grau de inconsistência em

relação às respostas humanas. O estudo identificou quatro principais tipos de erro nas respostas do ChatGPT: conceituais (54%), factuais (36%), relacionados a códigos (28%) e terminológicos (12%) (Kabir, 2023).

Esse fenômeno tem sido denominado pela comunidade técnica como “alucinações” (Hacker, 2024), caracterizando-se por situações em que o modelo gera uma resposta que, embora estruturalmente coerente, é incorreta em termos de fatos e dados (Alkaissi, 2023).

Diversos fatores podem contribuir para essas inconsistências, como o treinamento do modelo com informações insuficientes, de baixa qualidade ou com viés. Além disso, a falta de um contexto adequado ou restrições na formulação do comando podem levar a IA a produzir respostas imprecisas ou completamente fictícias. O impacto dessas informações geradas erroneamente não é irrelevante, pois pode resultar na disseminação de dados enganosos e imprecisos (Alkaissi, 2023).

Em novembro de 2023, um caso envolvendo um juiz federal da 1ª Região que utilizou o ChatGPT para redigir parte de uma sentença ganhou destaque na mídia. A ferramenta de inteligência artificial generativa, porém, fundamentou-se em uma jurisprudência inexistente do Superior Tribunal de Justiça, levando o advogado da parte derrotada a perceber o problema ao ler a decisão, o que resultou em uma manifestação à Corregedoria Regional da Justiça Federal da 1ª Região (Brave, 2023).

Em decorrência desse episódio, o desembargador federal Néviton Guedes, Corregedor Regional da 1ª Região, expediu uma circular recomendando que não se usem soluções de IA generativa abertas e não homologadas pelos órgãos de controle do Poder Judiciário para fins de pesquisa de precedentes jurisprudenciais. Tal orientação decorre tanto da necessidade de cumprimento da Resolução CNJ nº 332 de 2020 quanto dos deveres de prudência, supervisão e divulgação responsável de informações processuais (TRF1, 2023).

O juiz que recorreu ao ChatGPT alegou ter sido um engano motivado por excesso de trabalho, apontando que a parte da sentença em questão teria sido elaborada por um servidor. A Corregedoria Regional da Justiça Federal da 1ª Região arquivou o caso, mas o Conselho Nacional de Justiça ainda avaliará o assunto (Brave, 2023).

Vale ressaltar que a empresa chinesa DeepSeek recentemente ganhou

projeção global com o lançamento de seu modelo de linguagem avançado, o DeepSeek-R1, em janeiro de 2025. Esse modelo se diferencia por apresentar um desempenho semelhante ao de outras IAs de ponta, como o GPT-4 da OpenAI, porém com um custo de desenvolvimento consideravelmente reduzido. (Zhang,2025)

É fundamental observar que a Inteligência Artificial se desenvolve de forma acelerada, trazendo à tona questões éticas, sociais e filosóficas sobre seu papel na sociedade. Pontos como vieses em algoritmos, privacidade, automação de empregos e responsabilidade em decisões tomadas por máquinas vêm sendo amplamente debatidos.

Diante desse cenário, o presente estudo se propõe a investigar o problema dos vieses algorítmicos na tomada de decisões no âmbito do Poder Judiciário. Portanto, este capítulo torna-se indispensável para aprofundar o entendimento do estado da arte da tecnologia, em especial sobre o funcionamento dos sistemas e algoritmos inteligentes.

3 INTELIGÊNCIA ARTIFICIAL NO PODER JUDICIÁRIO

A evolução tecnológica tem transformado diversos setores da sociedade, e o Poder Judiciário não é exceção, e a aplicação da Inteligência Artificial (IA) no sistema judicial representa uma revolução na forma como os processos são conduzidos, trazendo avanços significativos na automação de tarefas, na análise de dados e na tomada de decisões, todavia, essa inovação também suscita desafios éticos, jurídicos e técnicos que precisam ser cuidadosamente analisados.

A adoção de soluções de Inteligência Artificial no Judiciário não é um fenômeno recente. Já ao final do século XX, Sartor e Branting (1998) abordaram várias maneiras de empregar a IA na tomada de decisões judiciais, evidenciando estudos que examinavam tanto rotinas forenses baseadas na aplicação direta de regras objetivas a fatos indiscutíveis, quanto análises direcionadas a aspectos mais intrincados do Direito, em especial no que se refere ao processo decisório dos magistrados.

Conforme apontam Staats e Morais (2020), o impacto das novas tecnologias no cotidiano e no âmbito jurídico começou a se tornar mais evidente na década de 1960, com o surgimento da jurisprudência mecânica e da jurimetria. Essas iniciativas representaram os primeiros esforços sistematizados de aplicar a Inteligência Artificial ao Direito, destacando-se ainda a adoção de soluções propostas por lawtechs para aprimorar serviços como a litigância de massa, tema que até então não recebia tanta atenção.

Segundo Dabass e Dabass (2018), a Inteligência Artificial vem desempenhando um papel cada vez mais relevante em serviços jurídicos, por meio de soluções como pesquisa de dados, modelos preditivos, interfaces inteligentes, sistemas de triagem e bots legais. Nesse contexto, a pesquisa de dados jurídicos inclui o uso de algoritmos de previsão, que estimam a probabilidade de êxito de um processo judicial com base em temas específicos e em análises de ações correlatas, fundamentando-se em conclusões e esquemas de conhecimento validados em distintas áreas de atuação legal.

McGinnis e Pearce (2014) destacam que a tecnologia vem transformando o Direito em diversas dimensões. Inicialmente, enfatizam a descoberta legal, evidenciada pela capacidade de realizar buscas e examinar documentos jurídicos. Em seguida, apontam para a utilidade na pesquisa jurídica, na qual algoritmos identificam informações essenciais contidas na doutrina e na jurisprudência. Além disso, ressaltam a automação na criação de documentos — como formulários, memorandos e relatórios — e, por fim, mencionam a aplicação da tecnologia na previsão de casos judiciais por meio da integração e análise de informações.

Silva (2017) aponta que as tecnologias de informação e comunicação estão cada vez mais inseridas no cotidiano das pessoas, impactando tanto a esfera pessoal quanto a profissional, no contexto do Poder Judiciário brasileiro, essa realidade se tornou ainda mais evidente com a implantação do Processo Judicial Eletrônico (PJe), e a modernização da prestação jurisdicional implica, portanto, na informatização e no uso de ferramentas tecnológicas, o que traz benefícios como aumento da produtividade, redução do tempo dedicado a tarefas burocráticas e economia de recursos financeiros e ambientais, ao eliminar a necessidade de uso de papel.

Nesse sentido, a IA tem se firmado como um dos avanços tecnológicos mais impactantes no sistema judiciário, contribuindo para a maior eficiência na tramitação processual e na tomada de decisões, de acordo com Coelho (2017), diversas iniciativas já fazem uso dessa tecnologia no campo jurídico, buscando otimizar diferentes áreas de atuação, todavia, como ainda se encontra em uma fase inicial de desenvolvimento e testes, a aplicação da IA enfrenta limitações tanto horizontais, relacionadas às áreas do direito em que pode ser empregada, quanto verticais, que dizem respeito às funções que essa tecnologia é capaz de desempenhar no ambiente jurídico.

Dessa forma, atividades de maior complexidade, que demandam altos níveis de empatia, criatividade, sensibilidade ou interação direta com o ambiente, ainda estão longe de serem plenamente reproduzidas por sistemas computacionais, não representando, portanto, uma ameaça iminente da tecnologia, entre essas atividades, destacam-se a comunicação e a negociação entre as partes, a presença física em audiências e o aconselhamento jurídico aos clientes, aspectos que dependem fortemente da atuação humana e do julgamento subjetivo do profissional do direito (Coelho, 2018).

Nesse senda Matias e Araújo (2023) destacam que as questões relacionadas à Inteligência Artificial foram concebidas e idealizadas muito antes de sua implementação nos tempos modernos, especialmente quando relacionadas à vida, comportamento e sociedade, a evolução tecnológica acelerou a Revolução Digital (Indústria 4.0), impactando diversas áreas, como economia, meios de comunicação pública e privada, consultorias, medicina e direito, gerando transformações sociais e empresariais inesperadas, nesse escopo de disrupção digital global, surge a necessidade de adaptação aos novos paradigmas no campo jurídico.

3.1 Aplicações de IA em Sistemas do Poder Judiciário Brasileiro

O uso da inteligência artificial no Poder Judiciário brasileiro tem apresentado crescimento significativo nos últimos anos, sendo adotado por quatorze tribunais nacionais – entre eles o STF e o STJ – com o intuito de aprimorar a prestação jurisdicional. Segundo levantamento realizado em 2023 pelo Conselho Nacional de Justiça (CNJ), existem atualmente cerca de 140 projetos de IA desenvolvidos ou em desenvolvimento nos tribunais e conselhos de Justiça, representando um aumento de 26% em comparação com a pesquisa de 2022. Dentre essas soluções, 63 já estão operacionais ou aptas a serem utilizadas, 46 encontram-se em fase final de desenvolvimento, 17 estão em estágio inicial e 3 ainda não foram iniciadas. Além disso, 11 projetos foram concluídos, mas ainda não foram implementados (CNJ, 2023).

De acordo com a Organização para a Cooperação e Desenvolvimento Econômico (OCDE) e o Relatório da Fundação Getúlio Vargas, que integra a segunda fase da pesquisa sobre inteligência artificial nos tribunais brasileiros, o Judiciário do Brasil conta com um número maior de ferramentas de IA em comparação com os sistemas judiciais da União Europeia, dos Estados Unidos, do Reino Unido, da China e da Índia (Salomão, 2022).

As possibilidades de aplicação da inteligência artificial no campo do Direito são diversas, conforme sistematizado por Junquilho(2022): modelos preditivos, organização de documentos, busca textual, automação de documentos e

decisões automatizadas. Essas categorias refletem os diferentes papéis que a IA pode desempenhar nos tribunais brasileiros, desde a inferência de informações em documentos específicos até a análise preditiva para auxiliar em possíveis decisões.

Os sistemas de IA desenvolvidos para o Poder Judiciário visam, entre outros objetivos, reduzir o tempo médio de tramitação processual, viabilizar análises preditivas, identificar e analisar padrões na jurisprudência e empregar chatbots para atender e esclarecer dúvidas dos usuários.

Uma das aplicações mais conhecidas de inteligência artificial no Poder Judiciário brasileiro é o robô VICTOR. Desenvolvido em parceria com a Universidade de Brasília e implantado no Supremo Tribunal Federal (STF) em 2019, o VICTOR tem como principal objetivo facilitar a identificação de padrões em documentos jurídicos, geralmente no formato PDF, submetidos à Corte. A ferramenta é capaz de reconhecer recursos relacionados a 27 temas de repercussão geral, encaminhando-os automaticamente aos tribunais de origem, além de identificar e separar as cinco principais peças processuais: sentença, acórdão, petição de recurso extraordinário, juízo prévio de admissibilidade e petição de agravo. Futuramente, pretende-se incorporar ao sistema a funcionalidade de agrupamento por similaridade. A adoção do VICTOR trouxe ganhos significativos de eficiência, reduzindo o tempo de execução de determinadas tarefas de 44 minutos para apenas cinco segundos (CNJ, 2021).

Destacando que, segundo Valle, Gasó e Ajus (2023) no âmbito jurídico, uma das principais preocupações dos especialistas é que o uso de inteligência artificial por órgãos julgadores leve, eventualmente, à substituição dos magistrados por essas tecnologias, no Brasil, destaca-se o Projeto Victor, desenvolvido pelo Supremo Tribunal Federal em parceria com a Universidade de Brasília, com a colaboração de estudantes de Direito, Engenharia de Software, Informática, entre outras áreas.

O Projeto Victor tem como funções centrais a identificação dos principais documentos nos processos encaminhados ao STF e a categorização dos recursos conforme os Temas de Repercussão Geral. Apesar de sua sofisticação e potencial para agilizar o andamento processual com maior precisão e velocidade em comparação ao trabalho humano, a ferramenta apresenta limitações relevantes, ademais, entre os desafios identificados está a possibilidade de vieses na programação, algo comum em sistemas de inteligência artificial, outro ponto de

atenção é o fenômeno da hipernormatização artificial, que já é uma característica presente no sistema de Repercussão Geral e tende a ser acentuado com o uso da IA. Assim, é essencial que o Projeto Victor seja analisado com responsabilidade e criticidade, de modo a garantir que sua implementação não comprometa a imparcialidade e a justiça no processo decisório (Valle; Gasó; Ajus, 2023).

Diversos outros robôs de inteligência artificial estão em desenvolvimento ou já foram implementados nos tribunais do Poder Judiciário brasileiro. Projetos como o *Athos* (STJ), *Bem-Te-Vi* (TST), *Berna* (TJGO) e *Análise Legal Inteligente* (TRF da 1ª Região) têm como foco principal a consolidação de processos para julgamento conjunto (CNJ, 2023). Um exemplo notável é o sistema *Berna*, utilizado pelo Tribunal de Justiça de Goiás (TJGO), que contribuiu significativamente para a redução de decisões conflitantes e de anulações baseadas no princípio do juiz natural. Atualmente, a plataforma é capaz de reconhecer mais de 16 mil grupos de processos similares. Além disso, o *Berna* automatiza movimentações processuais ao gerar documentos a partir dos dados extraídos dos sistemas do TJGO e encaminhá-los diretamente aos magistrados, realizando leitura e comparação automática de todas as petições recebidas com as informações previamente armazenadas. O sucesso do modelo levou à sua adoção por outros sete tribunais estaduais, consolidando-se como referência no Judiciário (Ortega, 2023).

Outro exemplo recente é o *Bastião*, desenvolvido pelo Tribunal de Justiça de Pernambuco (TJPE) com o objetivo de identificar e mitigar demandas predatórias e repetitivas. O sistema opera com uma rede social interna composta por magistrados, servidores e instituições parceiras – como a OAB e o Ministério Público – por meio da qual emite alertas sobre o potencial predatório das ações, realiza classificações e etiquetamentos em massa, gera relatórios sobre demandas semelhantes, identifica documentos reutilizados em diferentes processos, agrupa ações para atuação coletiva e facilita a troca de estratégias e experiências entre os usuários (CNJ, 2023).

Antes do *Bastião*, o TJPE já utilizava a IA *Elis*, voltada à identificação e análise inicial de processos de execução fiscal. Na primeira unidade em que foi implantada, o tempo necessário para analisar 80 mil processos e dar andamento aos casos foi reduzido de um ano e meio para apenas 15 dias, representando uma economia de mais de um ano na tramitação processual (Ortega, 2023).

Apesar dos avanços, Maranhão et al. (2023) chamam atenção para os riscos associados ao uso dessas tecnologias, especialmente no que diz respeito à possibilidade de erros nos agrupamentos automáticos. Agrupamentos inadequados podem levar à inclusão de casos em conjuntos cujo julgamento diverge do precedente aplicável, gerando prejuízos às partes envolvidas.

Os chatbots constituem uma aplicação amplamente adotada, exemplificada pelo Peticionamento Inteligente (TJ/RO), Chatbot DIGEP (TJ/RS) e JUDI (TJ/SP), e são utilizados para interagir com servidores, jurisdicionados e cidadãos. Entretanto, os principais desafios desses sistemas estão relacionados aos vieses discriminatórios e à desinformação que podem ocorrer na comunicação com os usuários, impactando negativamente as interações pré-processuais com as partes e o público jurisdicionado (Maranhão et al., 2023).

O Projeto OMNI (TJ/MT) e a Jurimetria com IA (TJ/MS) utilizam sistemas preditivos para gerar alertas, métricas e indicadores. No entanto, o uso da IA para fazer previsões apresenta riscos, como generalizações inadequadas e previsões que não se ajustam ao caso concreto, o que torna necessária a revisão humana para assegurar a conformidade dos precedentes (Maranhão et al., 2023).

Em novembro de 2023, o Tribunal de Justiça do Paraná (TJPR) lançou o *NatJusGPT*, considerada a primeira ferramenta de inteligência artificial generativa voltada ao apoio de magistrados em processos de judicialização da saúde. Desenvolvido pelo Departamento de Tecnologia da Informação e Comunicação (Dtic), em parceria com o Ateliê de Inovação do Gabinete da Presidência, e implementado com o suporte de uma empresa contratada para serviços e licenciamentos, o *NatJusGPT* funciona como um chatbot capaz de interagir com os usuários e fornecer, de forma ágil, as notas técnicas mais relevantes ao caso analisado. Para isso, utiliza informações provenientes de diversas bases de dados confiáveis. A ferramenta opera exclusivamente no ambiente seguro e privado da nuvem Microsoft do TJPR, sem qualquer compartilhamento externo de dados. Um dos principais diferenciais do sistema é sua limitação deliberada às notas técnicas que compõem seu banco de treinamento, o que reduz significativamente o risco de "alucinações" – respostas incorretas ou inventadas, comuns em IAs generativas (Schwartz, 2023).

Como evolução do *NatJusGPT*, o Tribunal de Justiça do Paraná (TJPR) lançou, em janeiro de 2024, a *JurisprudênciaGPT*, considerada a primeira

ferramenta funcional de pesquisa jurisprudencial baseada em inteligência artificial generativa voltada ao uso por todos os tribunais brasileiros. Desenvolvido com o objetivo de modernizar e agilizar a busca por precedentes judiciais, o sistema permite consultas em linguagem natural, retornando respostas detalhadas, contextualizadas e juridicamente fundamentadas.(TJPR, 2024)

Treinada com uma ampla base de dados que inclui acórdãos, súmulas e precedentes de diversos tribunais, a *JurisprudênciaGPT* viabiliza a identificação rápida de informações relevantes para cada caso analisado, reduzindo o tempo despendido com pesquisas manuais. Além de localizar precedentes específicos, a ferramenta é capaz de reconhecer padrões e tendências na jurisprudência, oferecendo insights valiosos que auxiliam magistrados e operadores do Direito na construção de decisões mais fundamentadas e alinhadas com os entendimentos predominantes nos tribunais superiores (Schwartz, 2023).

Após o lançamento do *NatJusGPT* pelo Tribunal de Justiça do Paraná (TJPR), o Tribunal de Justiça de Minas Gerais (TJMG) criou um grupo de trabalho com o objetivo de avaliar a eficiência dessa ferramenta e explorar sua aplicabilidade no âmbito mineiro. Paralelamente, o TJMG passou a testar duas soluções próprias baseadas em inteligência artificial generativa: *SAVIA* e *SOFIA*. Atualmente em fase experimental, a *SAVIA* utiliza o modelo GPT-3.5 da OpenAI e é equipada com técnicas avançadas de Processamento de Linguagem Natural. Alimentada por um vasto banco de dados disponível na internet, a ferramenta permite aos usuários obter sugestões de textos em poucos segundos. As entradas podem ser feitas por voz ou texto, e a IA busca gerar conteúdos compatíveis com a demanda apresentada, cabendo ao usuário revisar e ajustar o material conforme necessário. (TJMG,2023)

A expectativa do TJMG é aplicar a *SAVIA* em diversas atividades administrativas, como elaboração de e-mails, portarias, resoluções, relatórios e demais documentos institucionais, otimizando o tempo e aumentando a produtividade dos setores envolvidos (Schwartz, 2023)

Em abril de 2024, a Universidade de São Paulo (USP), em colaboração com a Escola Superior da Magistratura Tocantinense (Esmat) e o Tribunal de Justiça do Estado do Tocantins (TJTO), apresentou a "SARA" (Síntese Adaptativa de Requisitos de Admissibilidade). Essa ferramenta de IA generativa foi desenvolvida especificamente para apoiar a análise de recursos constitucionais, especiais e

extraordinários, sendo capaz de gerenciar todas as etapas da avaliação da admissibilidade recursal – desde o início do processo até a verificação de jurisprudência correlata nos tribunais superiores. Além disso, a SARA automatiza a criação de minutas e documentos que seguem os padrões do tribunal e se ajustam ao estágio e encaminhamento de cada recurso, permitindo um monitoramento completo das sínteses produzidas e possibilitando a intervenção humana em qualquer fase. Essa abordagem resulta em uma análise e tramitação de recursos mais eficiente e precisa. O desenvolvimento da SARA contou com o apoio da startup Taqui, integrada por pesquisadores do grupo SmartCitiesBR da USP (Sousa,2024).

Em todas as regiões do Brasil, estão sendo desenvolvidas ou implementadas soluções baseadas em IA. Alguns tribunais que ainda não possuem projetos próprios em andamento ou operacionais informam estar avaliando iniciativas já existentes, embora ainda não tenham formalizado parcerias.(Silva; Filho, 2020)

3.2 Benefícios do uso da IA na tomada de decisão judicial

O crescente volume de informações jurídicas atualmente ultrapassa as capacidades humanas de memória, pesquisa e análise, e a utilização da Inteligência Artificial (IA) permitirá que os profissionais do Direito naveguem por esse vasto oceano de dados com maior agilidade, rapidez e eficiência, empresas já estão incorporando o potencial "artificial" em suas cadeias de produção ou na prestação de serviços, superando limitações biológicas e expandindo suas operações para níveis além das capacidades humanas, ao delegar à máquina tarefas e funções que os seres humanos não conseguem ou não estão dispostos a realizar, esses benefícios resultam em uma redução de custos e tempo, traduzindo-se em serviços mais rápidos, acessíveis e de melhor qualidade (Hogemann, 2018).

A introdução da IA no judiciário é também vista de maneira positiva, conforme destacam Faraco e Todesco (2018), no contexto do Direito, uma das abordagens mais eficazes para apoiar o raciocínio jurídico é o raciocínio baseado em casos (RBC), ou *case-based reasoning*, que, quando combinado com outras técnicas de inferência, tem se mostrado promissor, normalmente, essa metodologia

é estruturada em um modelo de duas camadas: uma camada voltada para as regras (leis) e outra destinada ao raciocínio fundamentado em casos, permitindo que a IA apoie os profissionais do direito de forma mais eficiente e contextualizada.

Conforme afirma Rover (1997), mais do que em qualquer outro momento da história do Direito, é fundamental enfrentar a complexidade administrativa e técnica do Sistema Jurídico, buscando respostas eficazes às necessidades da sociedade, e a solução imediata para esses desafios está no desenvolvimento, aprimoramento e implementação das tecnologias de informação, que podem ser divididas em tecnologias de comunicação e de conhecimento, nesse contexto, essas tecnologias se referem aos mecanismos e programas que facilitam o acesso universal aos dados, ou seja, sem impor restrições, exceto aquelas necessárias para garantir a segurança e a integridade dos sistemas.

Além do mais, Hogemann (2018) enfatiza que se está vivendo em uma era de avanços tecnológicos extraordinários, mas também de grandes incertezas para diversas profissões, incluindo a jurídica, entretanto, esse cenário também é repleto de oportunidades, uma dessas oportunidades é a possibilidade de redefinir o perfil do "advogado do futuro", por outro lado, é essencial reconhecer que chegou o momento de transformar a figura do advogado de ficção científica em uma realidade, para entender o papel e as funções que os advogados desempenharão no futuro próximo, é necessário desenvolver uma compreensão mais profunda de como o crescimento acelerado das tecnologias disruptivas está moldando o mundo e, conseqüentemente, a prática jurídica.

Nesse diálogo, Maia Filho e Junquilha (2018) destacam que as inovações tecnológicas têm o potencial de impactar de forma decisiva as análises de dados no âmbito jurídico, permitindo a coleta de informações e a combinação de sentenças e dados relacionados a diversos casos que não poderiam ser adequadamente confrontados com os métodos tradicionais de pesquisa jurisprudencial, e a gestão, o controle e a categorização dos processos judiciais podem ocorrer de maneira mais eficiente e racional, possibilitando uma análise de dados mais qualificada e abrangente, além do que, a tecnologia permite o reconhecimento de padrões, o cruzamento de informações e a geração de insights a partir de diferentes perspectivas e contextos.

A identificação de correlações e a construção de conexões entre diversos processos e casos têm o poder de proporcionar uma compreensão mais

detalhada dos dados jurisprudenciais, das leis e da doutrina consultada, isso permite uma análise mais clara de como os julgamentos anteriores foram conduzidos e, principalmente, das motivações que sustentaram as decisões, em vez de apresentar apenas um conjunto de informações, o produto dessa análise será um modelo estatístico que possibilitará a realização de avaliações mais precisas e preditivas (Maia Filho; Junquilha, 2018).

As tecnologias fundamentadas em inteligência artificial têm sido aplicadas em diversas áreas do conhecimento e estão começando a ser incorporadas nos sistemas jurídicos, especialmente em países que adotam o Common Law, um exemplo citado por Valentini (2017) é o ROSS, um sistema de IA criado para auxiliar advogados e tribunais na pesquisa e análise de precedentes jurídicos, ilustrando o potencial dessas ferramentas para melhorar a eficiência no setor judiciário.

Valentini (2017) também destaca que o principal desafio do Poder Judiciário, no contexto atual, não reside na criação de um sistema totalmente automatizado para os julgamentos, mas sim no desenvolvimento de uma tecnologia capaz de se adaptar e evoluir de forma contínua frente às diversas situações jurídicas que surgem, nesse sentido, o sistema deve ser aprimorado por meio de um aprendizado constante, com intervenção e supervisão humana, aprimorando sua base de dados a cada nova decisão tomada, e o objetivo final, conforme ele, é alcançar um nível de desenvolvimento que permita a elaboração de minutas de decisões com qualidade técnica suficiente para auxiliar no julgamento de casos concretos, promovendo maior eficiência e precisão na aplicação da Justiça.

Outrossim, conforme Carvalho (2021), os países que estão na vanguarda da adoção da Inteligência Artificial podem expandir seus ganhos econômicos em até 25%, enquanto as nações que ficam atrás nesse processo tendem a registrar um crescimento em torno de 10%, previsões semelhantes são observadas no setor empresarial, onde as empresas que lideram a implementação da IA têm potencial para praticamente dobrar seus lucros, ao passo que aquelas que adiam essa adoção podem enfrentar dificuldades no cenário competitivo, além do mais, o mercado de trabalho se mostrará mais promissor para os profissionais capazes de integrar a IA em suas funções, oferecendo salários mais elevados em comparação com aqueles que desempenham tarefas repetitivas e suscetíveis à automação.

Desta feira, nações que associam a liderança em Inteligência Artificial ao fortalecimento de sua soberania nacional têm realizado investimentos significativos na atração, retenção e capacitação de talentos na área de IA, além de estruturarem uma base sólida de pesquisa, tanto fundamental quanto aplicada, esse desenvolvimento possibilita a geração e a transferência de conhecimento para o setor produtivo, agregando valor à produção nacional e incentivando a criação de startups inovadoras, como resultado, esses países não só atendem às suas próprias necessidades internas, mas também fortalecem sua balança comercial, atraem novos investimentos e geram empregos qualificados e bem remunerados (Carvalho, 2021).

Nesse diálogo, Rocha e Kissimoto (2022) destacam que a trajetória da Inteligência Artificial se desenvolveu ao longo da história em duas principais vertentes: a abordagem centrada no ser humano, que utiliza métodos empíricos baseados em experimentação e validação de hipóteses, e a abordagem racionalista, que integra modelos matemáticos e princípios de engenharia, as pesquisas iniciais sobre IA se concentraram na capacidade das máquinas de aprender, evidenciando avanços promissores em processos de tomada de decisão e resolução de problemas.

Embora o progresso tenha sido mais lento do que o inicialmente esperado, a partir da década de 1990 houve uma retomada significativa dos investimentos por parte de governos e empresas para impulsionar o desenvolvimento da Inteligência Artificial, e o aprimoramento dos métodos e tecnologias, a maior eficiência das tecnologias de informação no processo de captura e armazenamento de dados organizacionais, a crescente acessibilidade a ferramentas tecnológicas e a expansão dos serviços em nuvem são fatores que têm impulsionado a ampla adoção da IA no setor corporativo (Rocha; Kissimoto, 2022).

Segundo Toledo e Pessoa (2023) no contexto jurídico, eles destacam que a informatização e a integração da Inteligência Artificial no Judiciário brasileiro oferecem vantagens evidentes, especialmente em relação ao aumento da celeridade e eficiência na prestação jurisdicional, contudo, esses fatores não devem ser considerados isoladamente nem como objetivos finais, na prática, a rapidez e a eficiência devem ser vistas como instrumentos para promover a verdadeira missão da Justiça, que é a realização do Direito, esse valor essencial está consagrado internacionalmente nos direitos humanos e, no âmbito nacional, nos direitos

fundamentais, cuja plena efetivação só pode ser garantida em um regime democrático.

O Poder Judiciário brasileiro enfrenta, há décadas, uma realidade marcada pela alta carga processual. Com milhões de ações em tramitação em todo o país, o sistema judiciário se vê sobrecarregado, o que compromete sua capacidade de oferecer respostas rápidas e eficazes à sociedade. Essa demanda excessiva resulta em longas filas de espera, aumento do tempo médio de tramitação e acúmulo de processos nos gabinetes dos magistrados, dificultando o cumprimento do princípio constitucional da duração razoável do processo. A morosidade da justiça, portanto, tornou-se um dos principais desafios institucionais do Brasil (CNJ, 2023).

Nesse contexto, a incorporação de tecnologias baseadas em inteligência artificial (IA) surge como uma estratégia promissora para aumentar a celeridade e a eficiência do Judiciário. Ferramentas de IA têm sido utilizadas para automatizar tarefas repetitivas, como triagem de petições, organização de documentos, identificação de precedentes e elaboração de minutas. Ao assumir essas funções operacionais, os sistemas inteligentes liberam tempo dos magistrados e servidores, permitindo que se dediquem à análise jurídica mais aprofundada dos casos. Além disso, a IA pode reduzir falhas humanas, padronizar procedimentos e tornar a tramitação processual mais fluida e coerente (Junquilha, 2022).

Os benefícios da inteligência artificial no sistema de justiça brasileiro já são visíveis em diversas iniciativas implementadas em tribunais do país. Projetos como o VICTOR (STF), SINAPSE (TJRO) e BERNA (TJGO) demonstram como a tecnologia pode contribuir para acelerar o julgamento de processos, organizar demandas de forma inteligente e auxiliar na tomada de decisões. Quando aplicada com responsabilidade, supervisão humana e critérios éticos, a IA tem o potencial de transformar o Judiciário em uma instituição mais ágil, eficaz e acessível, sem comprometer a qualidade das decisões e o respeito aos direitos fundamentais.

Ademais, Valentini (2017), em sua pesquisa, argumenta que diversas fases do processo decisório, como a análise dos argumentos jurídicos pertinentes a cada caso e atividades que exigem algum tipo de julgamento por parte de um sistema automatizado, poderão ser realizadas por meio do desenvolvimento de mecanismos de busca sofisticados, esses sistemas serão capazes de consultar

bases de jurisprudência e identificar casos semelhantes, proporcionando suporte na fundamentação das decisões.

Portanto, a trajetória da IA, que avança por meio de abordagens empíricas e técnicas, tem progredido de forma crescente no setor jurídico, a informatização e a integração da IA ao Judiciário brasileiro têm trazido melhorias em termos de celeridade e eficiência, todavia, é elementar ressaltar que esses avanços não devem ser vistos como objetivos finais, mas sim como ferramentas para alcançar o valor essencial do Direito: a justiça, que deve ser assegurada dentro de um regime democrático, com total respeito aos direitos fundamentais.

3.3 Riscos e Desafios do uso da IA no cenário judiciário

Apesar das vantagens expressivas, o sistema processual eletrônico ainda enfrenta críticas, especialmente no que se refere ao tempo necessário para a conclusão dos processos e à elaboração de documentos, nessas etapas, o processo eletrônico continua apresentando uma lentidão comparável à do processo judicial tradicional em papel, o que evidencia a necessidade de aprimoramento nas ferramentas tecnológicas e na infraestrutura do sistema (Silva, 2017).

Nesse cenário, os desafios e riscos associados ao uso da inteligência artificial como ferramenta para a resolução de demandas no sistema judiciário têm se intensificado, e o avanço tecnológico possibilitou a tomada de decisões automatizadas no âmbito judicial, o que representa um progresso significativo, mas também envolve riscos e potenciais falhas, diante disso, a criação de legislações e resoluções específicas torna-se essencial para mitigar esses problemas e garantir o uso responsável da IA, essas regulamentações podem abranger desde a definição de princípios éticos e diretrizes de transparência até a instituição de órgãos reguladores especializados na supervisão da aplicação da IA no sistema jurídico (Silva; Souza; Souza, 2024).

A regulamentação pode impor critérios rigorosos para a criação e aplicação de algoritmos de inteligência artificial, incluindo a realização de testes para garantir imparcialidade e equidade, auditorias periódicas e um acompanhamento constante para identificar e corrigir possíveis vieses e discriminações, além do mais, a transparência na utilização da IA no sistema judiciário é essencial, exigindo que os

processos decisórios automatizados sejam acessíveis e passíveis de auditoria, garantindo que todas as partes envolvidas possam compreendê-los e questioná-los quando necessário (Silva; Souza; Souza, 2024).

Nesta senda, Coelho (2018) destaca que o principal desafio no desenvolvimento de sistemas de inteligência artificial está na capacidade de gerar criatividade útil, embora os computadores sejam capazes de combinar diversos elementos preexistentes e criar novas ideias ou conceitos, essas produções nem sempre são inovadoras ou têm aplicação substanciais.

Conforme o Parlamento Europeu (2024) os sistemas de IA de uso geral, assim como os modelos nos quais esses sistemas se baseiam, devem atender a certos requisitos de transparência, incluindo o cumprimento das legislações da UE sobre direitos autorais e a divulgação de informações detalhadas sobre os conteúdos utilizados no treinamento, para os modelos de IA de uso geral mais avançados, que possam representar riscos sistêmicos, serão exigidos requisitos adicionais, como a realização de avaliações dos modelos, a identificação e mitigação dos riscos sistêmicos e a comunicação de incidentes.

Ainda nesse debate crítico, Coelho (2018) aponta que a limitação da inteligência artificial nesse aspecto está ligada à ausência de empatia e de uma visão abrangente do contexto humano, o que dificulta a identificação de conceitos que realmente sejam relevantes ou impactantes para as pessoas, um exemplo disso é a capacidade da IA de produzir textos rimados, mas sem a profundidade emocional necessária para criar um poema que desperte sentimentos genuínos, assim, embora a inteligência artificial possa replicar padrões e organizar informações, ela ainda enfrenta dificuldades para criar com verdadeira originalidade e sensibilidade humana.

A regulamentação da IA no setor judicial tem se tornado uma preocupação crescente em diversos países, o Parlamento Europeu (2024) aponta que o AI ACT da União Europeia, por exemplo, classifica o uso da IA na Justiça como de "alto risco", impondo a necessidade de transparência, supervisão humana e explicação das decisões tomadas pelos sistemas.

Nesse sentido, o Parlamento Europeu (2024) expõe que também estão previstas obrigações claras para outros sistemas de IA considerados de alto risco, devido aos potenciais danos significativos que podem causar à saúde, segurança, direitos fundamentais, meio ambiente, democracia e Estado de direito, exemplos de

áreas com IA de alto risco incluem infraestruturas críticas, educação e treinamento profissional, emprego, serviços públicos e privados essenciais (como saúde e setor bancário), certos sistemas de aplicação da lei, migração e gestão de fronteiras, justiça e processos democráticos (como influenciar eleições), esses sistemas devem ser avaliados quanto aos riscos e ter medidas para mitigá-los, manter registros de uso, garantir transparência e precisão, além de assegurar supervisão humana, os cidadãos terão o direito de apresentar queixas sobre sistemas de IA e obter explicações sobre decisões tomadas por esses sistemas que impactem seus direitos.

No Brasil, a recente Resolução nº 615/2023 do Conselho Nacional de Justiça (CNJ) e o Projeto de Lei nº 2338/2023, em tramitação no Senado Federal, estabelecem diretrizes voltadas ao uso ético, transparente e responsável da inteligência artificial no Poder Judiciário. Essas normativas visam mitigar a ocorrência de vieses algorítmicos, garantir a auditabilidade dos sistemas utilizados e assegurar que as decisões automatizadas respeitem os direitos fundamentais. Contudo, apesar dos avanços normativos, a aplicação efetiva dessas regras ainda enfrenta desafios significativos, especialmente no que se refere ao controle e fiscalização dos modelos de IA em larga escala e à capacitação técnica dos operadores do Direito para lidar com essas novas tecnologias.

O Projeto de Lei nº 2.338, de 2023, conforme o Senado, de autoria do Senador Rodrigo Pacheco, tem como objetivo estabelecer diretrizes para o desenvolvimento, implementação e uso responsável de sistemas de inteligência artificial (IA) no Brasil, e a proposta visa proteger os direitos fundamentais, garantindo que os sistemas de IA sejam seguros e confiáveis, em benefício da pessoa humana, do regime democrático e do avanço científico e tecnológico.

O projeto também aborda a necessidade de regulamentação específica para sistemas de IA considerados de alto risco, devido aos potenciais danos que podem causar à saúde, segurança, direitos fundamentais, meio ambiente, democracia e Estado de direito, exemplos desses sistemas incluem infraestruturas críticas, educação, emprego, serviços públicos essenciais, aplicação da lei, migração, gestão de fronteiras, justiça e processos democráticos.

A proposta ressalta a importância de avaliar e mitigar os riscos associados a esses sistemas, mantendo registros de utilização, garantindo transparência e precisão, além de assegurar supervisão humana, os cidadãos terão

o direito de apresentar queixas sobre sistemas de IA e de receber explicações sobre decisões baseadas nesses sistemas que afetem seus direitos.

Atualmente, o Projeto de Lei nº 2.338 de 2023 está em tramitação no Senado Federal, e, em junho de 2024, foi realizada uma sessão temática para debater a regulamentação do uso de inteligência artificial no país, com a presença de especialistas e autoridades no assunto, e o relator da proposta, Senador Eduardo Gomes (PL-TO), apresentou um novo parecer sobre o projeto, que pode ser votado no Plenário da Casa em breve.

A aprovação desse projeto representa um avanço significativo na criação de um marco regulatório para a inteligência artificial no Brasil, alinhando o país às melhores práticas internacionais e garantindo que o desenvolvimento tecnológico seja realizado de forma ética, transparente e em respeito aos direitos fundamentais dos cidadãos.

No que se refere aos impactos na atuação dos profissionais do Direito, Silva (2017) destaca que as tecnologias de informação e comunicação se tornaram parte integrante do cotidiano das pessoas em diversas áreas do conhecimento, tanto no âmbito pessoal quanto profissional, no caso do Poder Judiciário brasileiro, essa realidade não é diferente, especialmente com a adoção do Processo Judicial Eletrônico, e a modernização dos serviços judiciais, por meio da informatização e do uso crescente de ferramentas tecnológicas, tem gerado benefícios significativos, como o aumento da eficiência e da produtividade, graças à automação de tarefas burocráticas que antes exigiam atuação humana, além do mais, essa transformação resulta em economia de recursos financeiros e ambientais, com a redução do uso de papel, entre outros efeitos positivos.

Segundo Sperandio (2018) a prática jurídica pode ser profundamente transformada por tecnologias, especialmente as preditivas, os legisladores criam normas que são, então, comunicadas aos cidadãos através de pequenos comandos, dessa maneira, o papel dos advogados sofrerá mudanças significativas, e as funções relacionadas ao contencioso e à conformidade (compliance) provavelmente perderão relevância, enquanto a consultoria (advisory) se tornará mais central. Isso ocorre porque parte da função de compliance, que envolve prever a sanção para uma determinada ação, pode ser facilmente realizada por uma máquina equipada com um banco de dados abrangente sobre as sanções previstas para diversas condutas.

Do ponto de vista ético, a questão da responsabilidade pelas decisões tomadas com o auxílio da IA é um dos principais pontos de discussão, segundo Valentini (2017), a automação não substituirá a necessidade de juristas criativos, que são capazes de desenvolver novas teses ou reinterpretar as antigas para aplicá-las em casos futuros, no entanto, ela reduzirá a demanda por técnicos jurídicos, profissionais que atualmente desempenham o papel de "computação humana", ajustando os dados fáticos e jurídicos necessários para a elaboração de peças processuais.

Assim, segundo Valentini (2017) para lidar com as questões éticas, filosóficas e políticas envolvidas na administração da Justiça, é fundamental repensar a estrutura do sistema processual, de modo a valorizar o papel dos juristas e dos profissionais humanos que desempenham funções essenciais nesse processo.

Como já demonstrado, a IA tem se mostrado uma ferramenta valiosa para auxiliar os juízes em tarefas repetitivas e burocráticas, como a triagem de processos e a identificação de precedentes relevantes, contudo, surge uma preocupação legítima: a possibilidade de que os magistrados sejam completamente substituídos por "juízes robôs".

O papel do juiz transformou-se de maneira significativa, deixando de ser um mero executor das leis. A ideia tradicional de que o juiz é apenas a "boca da lei" foi superada; hoje, sua função vai além de aplicar automaticamente os fatos às normas. O magistrado é responsável por assegurar a efetividade da jurisdição e a realização concreta dos direitos, adotando métodos de interpretação que combinam racionalidade e subjetividade. Esse papel, que exige a consideração das nuances e complexidades de cada caso decorrentes das interações humanas, não pode ser desempenhado integralmente por máquinas ou robôs, pois o Direito é, fundamentalmente, uma ciência humana (Barroso, 2006).

A maioria dos especialistas e das normas de regulação em IA sustenta que seu uso deve ser centrado no ser humano. Assim, embora os sistemas eletrônicos de inteligência artificial possam ser um recurso valioso para aprimorar a administração da justiça, eles não podem substituir o discernimento prudente do juiz – a máquina deve funcionar apenas como um apoio, e não como substituta da decisão, que precisa ser tomada por pessoas a serviço da humanidade (Pereira, 2017).

Além disso, o ato de julgar está intimamente relacionado a uma dimensão humana, com seus aspectos sociais, que nenhum sensor consegue captar, pois são características intrínsecas aos seres humanos (Lee, 2019). A capacidade de sentir emoções – como dor, alegria, amor ou ira – é exclusiva da consciência humana (Andrighi; Bianchi, 2020). Nesse contexto, destacam-se três dimensões fundamentais que separam o homem da máquina: o *logos*, que se refere à inteligência ou razão humana; o *phatos*, que abrange sentimentos e paixões; e o *virtus*, que representa os valores éticos e morais das ações humanas. Portanto, embora a inteligência artificial possa ser útil em questões jurídicas de natureza mais simples, ela não é adequada para lidar com as complexidades que envolvem o raciocínio pragmático, as emoções e os comportamentos éticos – atributos que permanecem exclusivos dos seres humanos (Hasselmann, 2019).

De acordo com Andrighi e Bianchi (2020), há situações em que as máquinas não conseguem decidir, especialmente quando a análise probabilística é inviável ou irrelevante – como ocorre na mudança de entendimento jurisprudencial, que requer um "julgamento de valores" não derivado de cálculos matemáticos, ou quando há dilemas morais e emocionais envolvidos. Nesse sentido, Spaulding (2020) argumenta que tais decisões devem ser tomadas por humanos, como se observa em questões relacionadas a direitos fundamentais, incluindo a liberdade de expressão, de imprensa e o livre exercício da religião.

O ato de julgar vai muito além de aplicar a lei de forma mecânica. Ele exige uma interpretação cuidadosa, a ponderação de valores e princípios, além de uma compreensão profunda do contexto social e humano que permeia cada caso. Sentenciar é, na verdade, sentir. Conforme De Plácido et al. (2002), a palavra "sentença" deriva do latim *sententia*, que significa "modo de ver, parecer, decisão". Da mesma forma, Dinamarco (2001) sustenta que esse termo carrega a ideia de que o juiz manifesta, ao proferir uma decisão, o modo como percebe a causa e as pretensões dos litigantes, aplicando não apenas a razão, mas também seus sentimentos.

Para compreender a fundo a causa e identificar a solução justa, é essencial que o juiz seja dotado de inteligência e cultura. A dimensão emocional presente no ato de julgar é algo que nenhuma máquina conseguirá replicar. Dessa forma, a inteligência artificial foi concebida para apoiar o magistrado em tarefas repetitivas e desgastantes, como a classificação de demandas, o monitoramento do

tempo de tramitação dos processos no gabinete e a pesquisa de dados das partes em plataformas públicas. Esse sistema inteligente e integrado permite que o juiz se concentre em sua função principal: julgar. Embora a IA possa elaborar um rascunho de decisão, é improvável que um magistrado consiga analisar centenas ou milhares de rascunhos gerados por tecnologia em poucos minutos, enquanto os assessores, com sua capacidade limitada, podem identificar problemas específicos que exigem uma análise aprofundada, mesmo em casos inicialmente considerados simples (Possa ; Azevedo, 2023). Isso ficou evidenciado em novembro de 2023, quando um juiz federal do TRF-1 utilizou o ChatGPT para gerar uma decisão que continha jurisprudência inexistente do Superior Tribunal de Justiça (Ale, 2023).

Além disso, no campo processual, existe uma preocupação com o que Junquillo (2022) chama de "decisionismo tecnológico", ou seja, um empobrecimento das decisões decorrente de uma parametrização excessiva que gera sentenças em massa, sem considerar as particularidades de cada caso. Em outras palavras, quando um algoritmo agrupa situações específicas como parte de uma categoria geral – sem o conhecimento das partes envolvidas – isso impede o pleno exercício do direito à ampla defesa e ao contraditório, dificultando também a auditoria de decisões convencionais.

A adoção de tecnologias de inteligência artificial – especialmente as de natureza generativa – no sistema judiciário representa uma oportunidade inédita para aprimorar a eficiência na administração da justiça. Contudo, para que esses benefícios sejam efetivados, é fundamental que magistrados e servidores recebam capacitação adequada para compreender as potencialidades e limitações dessas ferramentas tecnológicas (Avelar, 2024). Isso implica dominar o funcionamento dos sistemas de IA, desde seus princípios básicos até os algoritmos e processos de tomada de decisão.

A formação dos profissionais do Judiciário deve incluir treinamentos voltados à identificação e mitigação de possíveis vieses nos sistemas de inteligência artificial, com o objetivo de evitar que essas tecnologias reforcem ou ampliem desigualdades já existentes. É fundamental que magistrados e servidores estejam preparados para avaliar criticamente as decisões automatizadas, garantindo sua imparcialidade e conformidade com os princípios da justiça. Com essa finalidade, o Conselho Nacional de Justiça (CNJ) lançou o curso "Introdução à Inteligência Artificial para o Poder Judiciário", voltado a agentes interessados em entender os

conceitos fundamentais e o funcionamento dos modelos de IA, com o intuito de capacitá-los a reconhecer e propor formas de uso adequadas dessas ferramentas nos tribunais. Tal capacitação é imprescindível para assegurar que a IA seja utilizada como instrumento de apoio, e não como substituto da atuação humana no processo decisório (CNJ, 2023).

Esses desafios não devem ser vistos como obstáculos ao avanço tecnológico, mas como elementos cruciais para refletir sobre a governança algorítmica. Dessa forma, é imprescindível estabelecer limites éticos, considerando que, assim como qualquer outra funcionalidade da justiça, a IA deve estar subordinada aos princípios do Estado Democrático de Direito (Junquilho, 2022).

Apesar dos riscos e desafios abordados até o momento, os dois principais problemas relacionados à utilização de sistemas de inteligência artificial são a falta de transparência e interpretabilidade dos algoritmos, bem como a discriminação algorítmica. Esses tópicos serão explorados com mais profundidade a seguir.

3.3.1 Falta de transparência e explicabilidade nas decisões algorítmicas(opacidade)

Moreira (2017) argumenta que qualquer pessoa que se declare especialista em questões jurídicas relacionadas à IA e à Robótica estaria sendo desonesta, embora a Inteligência Artificial e os robôs já façam parte do cotidiano, o mundo jurídico ainda está apenas começando a despertar para analisar os fenômenos relacionados à adoção dessas tecnologias.

A opacidade algorítmica, também conhecida como o fenômeno "caixa-preta", tem emergido como uma das questões mais críticas no debate sobre o uso da inteligência artificial (IA) em diversas áreas, especialmente na moderação de conteúdo na internet, de acordo com Ferraz et al. (2024), a opacidade está diretamente relacionada à dificuldade de compreender como os sistemas de IA tomam decisões, o que leva a preocupações éticas e jurídicas sobre a transparência, a responsabilidade e os potenciais vieses nos resultados gerados.

Ferraz et al. (2024) explicam que os modelos de IA, especialmente os baseados em aprendizado profundo, são frequentemente tratados como caixas-pretas devido à sua complexidade técnica, e o funcionamento interno desses

sistemas – que envolve milhões de cálculos e interações entre camadas de redes neurais – torna a interpretação de seus processos de decisão inacessível para a maioria das pessoas, incluindo os próprios desenvolvedores, essa opacidade pode dificultar a detecção de erros e vieses, além de limitar a capacidade de auditar ou contestar decisões automatizadas.

A opacidade algorítmica é particularmente problemática no contexto da moderação de conteúdo online, onde algoritmos são amplamente utilizados para identificar e remover conteúdos que violam as diretrizes das plataformas, de acordo com Ferraz et al. (2024), os sistemas utilizados por plataformas como YouTube e Facebook removem automaticamente grande parte do conteúdo prejudicial, mas a falta de clareza sobre os critérios utilizados pelos algoritmos levanta preocupações sobre possíveis erros e discriminações, por exemplo, conteúdos relacionados a grupos minoritários, como a comunidade LGBTQIA+, podem ser removidos de forma desproporcional devido a falhas nos modelos de IA em interpretar nuances culturais e sociais.

Os impactos da opacidade algorítmica vão além da esfera técnica, afetando diretamente os direitos humanos, como a liberdade de expressão e a igualdade, Ferraz et al. (2024) destacam que a natureza opaca dos algoritmos pode levar à discriminação algorítmica, um fenômeno em que decisões automatizadas prejudicam desproporcionalmente determinados grupos sociais. Isso ocorre porque os algoritmos frequentemente reproduzem vieses presentes nos dados de treinamento, o que pode resultar em decisões injustas ou excludentes.

Além do que, a opacidade dos sistemas de IA compromete o direito à explicação, que é essencial para garantir a transparência e a responsabilidade nos processos de decisão automatizada, Ferraz et al. (2024) apontam que a ausência de explicações claras sobre as decisões algorítmicas dificulta a identificação de falhas e limita a capacidade de corrigir injustiças, criando um ambiente de desconfiança em relação à tecnologia.

Pode-se citar também a influência dos algoritmos na formação da opinião pública. Conforme observado por Ferraz et al. (2024), os algoritmos de moderação de conteúdo não apenas removem conteúdos considerados inadequados, mas também moldam o que os usuários veem e interagem nas plataformas digitais, essa curadoria algorítmica, quando opaca, pode reforçar câmaras de eco e polarizações, comprometendo o debate público saudável e a diversidade de opiniões.

Deste modo, para enfrentar os desafios impostos pela opacidade algorítmica, Ferraz et al. (2024) defendem a adoção de modelos de inteligência artificial explicável (Explainable AI ou XAI), essa abordagem busca desenvolver sistemas que sejam interpretáveis, permitindo que usuários e reguladores compreendam os critérios utilizados nas decisões automatizadas, a XAI tem o potencial de tornar os processos algorítmicos mais transparentes, mitigando os riscos de discriminação e aumentando a confiança nas tecnologias.

A explicabilidade, segundo Ferraz et al. (2024), pode ser alcançada de duas formas principais, por meio do desenvolvimento de modelos intrinsecamente interpretáveis, que são projetados desde o início para serem compreensíveis; e por meio de ferramentas pós-hoc, que analisam modelos existentes e fornecem explicações sobre suas decisões, ambas as abordagens são essenciais para garantir que os sistemas de IA operem de maneira ética e responsável.

Dessa forma, a transparência algorítmica deve ser acompanhada de regulamentações que exijam a explicação das decisões automatizadas, a exemplo do Regulamento Geral de Proteção de Dados (GDPR) da União Europeia. Ferraz et al. (2024) destacam que a criação de marcos legais claros é fundamental para equilibrar direitos como a privacidade, a liberdade de expressão e a proteção contra abusos tecnológicos, contribuindo para a construção de um ambiente digital mais justo, seguro e inclusivo.

Ferraz et al. (2024) sugerem que a adoção da XAI pode transformar o paradigma atual de moderação de conteúdo na internet, tornando-o mais transparente, ético e inclusivo, eles defendem que as plataformas digitais devem implementar sistemas de IA explicável que permitam a auditoria de suas decisões e ofereçam explicações claras aos usuários sobre porque determinados conteúdos foram removidos ou sinalizados.

Outrossim, Ferraz et al. (2024) enfatizam a importância de envolver a sociedade no processo de definição dos critérios de moderação. Isso inclui a criação de comitês de supervisão independentes e a transparência sobre as políticas de moderação adotadas pelas plataformas, e a participação democrática no desenvolvimento e na aplicação de algoritmos é essencial para garantir que as decisões automatizadas respeitem os valores e os direitos fundamentais, os autores destacam também a necessidade de investir em pesquisas interdisciplinares que

integrem conhecimentos de áreas como ciência da computação, direito e sociologia para abordar os desafios éticos e técnicos da opacidade algorítmica.

A opacidade algorítmica representa um dos maiores desafios éticos e técnicos da era digital, especialmente no contexto da moderação de conteúdo online, conforme argumentado por Ferraz et al. (2024), a falta de transparência nos sistemas de IA compromete a justiça, a responsabilidade e a confiança nas tecnologias, perpetuando desigualdades e prejudicando direitos fundamentais.

Para enfrentar esses desafios, é essencial adotar a IA explicável como um novo padrão de desenvolvimento e uso de algoritmos, e a transparência e a participação social devem estar no centro desse processo, garantindo que as decisões automatizadas sejam compreensíveis, auditáveis e alinhadas aos valores democráticos, assim, será possível construir um ambiente digital mais ético, inclusivo e confiável, no qual a tecnologia seja uma aliada na promoção da justiça e da equidade.

Ademais, essa nova realidade traz à tona diversas questões éticas, como a necessidade de transparência nos dados e nos processos, a responsabilização por eventuais danos, os impactos discriminatórios, a proteção da privacidade informacional, a autonomia tanto das máquinas quanto dos usuários e a possibilidade de geração de resultados tendenciosos, de modo que, a ética desempenha um papel fundamental na garantia de que os algoritmos sejam utilizados de maneira responsável, maximizando seus benefícios enquanto minimizam os riscos associados (Rossetti; Angeluci, 2021).

Outrossim, a opacidade algorítmica é um dos principais desafios éticos da atual sociedade da informação, fortemente mediada por sistemas de inteligência artificial (IA), conforme, Rossetti e Angeluci (2021), os algoritmos estão no cerne desse contexto, sendo responsáveis por processar vastas quantidades de dados e tomar decisões que impactam diretamente a vida cotidiana, todavia, a falta de transparência nesses sistemas, muitas vezes descritos como "caixas-pretas", levanta questões éticas e políticas sobre a rastreabilidade das decisões, a responsabilidade pelos erros e as consequências sociais das escolhas algorítmicas.

A opacidade algorítmica, conforme Rossetti e Angeluci (2021), refere-se à dificuldade em compreender os processos internos dos algoritmos e suas decisões, essa característica está ligada tanto à complexidade técnica dos sistemas quanto à falta de acesso aos dados e aos códigos que os compõem, mesmo quando os

algoritmos são projetados de forma ética, sua linguagem e arquitetura frequentemente tornam-se inacessíveis para aqueles que não possuem conhecimentos especializados em programação, assim, isso cria uma barreira que impede a análise crítica e a supervisão, dificultando a identificação de falhas e vieses.

Os autores citam que um dos exemplos dessa opacidade ocorreu com o algoritmo do aplicativo WhatsApp, que foi bloqueado temporariamente no Brasil em 2015 por não fornecer informações de mensagens trocadas entre suspeitos de crimes, eles explicam que a empresa alegou que os dados eram criptografados e, portanto, inacessíveis até mesmo para os próprios desenvolvedores, esse caso demonstra como a opacidade algorítmica pode gerar tensões entre os princípios de privacidade e as demandas por transparência e segurança pública (Rossetti; Angeluci, 2021).

Dessarte, a opacidade não se restringe ao acesso aos dados, mas também à compreensão sobre como as decisões são tomadas, como apontam Rossetti e Angeluci (2021), mesmo que os dados estejam disponíveis, os processos algorítmicos baseados em aprendizado de máquina são tão complexos que os próprios desenvolvedores podem não ser capazes de explicar como o sistema chegou a uma conclusão específica, esse fenômeno, conhecido como "algoritmos caixa-preta", representa um risco significativo, pois impossibilita a auditoria e a responsabilização.

Os impactos da opacidade algorítmica vão além do campo técnico, afetando diretamente as relações sociais e os direitos fundamentais, deste modo, Rossetti e Angeluci (2021) destacam que, em muitos casos, a falta de transparência nos processos algorítmicos resulta em decisões injustas e discriminatórias, e um exemplo dado por esses autores é o caso do algoritmo COMPAS, utilizado no sistema de justiça criminal dos Estados Unidos para prever reincidência, e o sistema classificou de forma desproporcional réus negros como de maior risco, mesmo em situações com histórico criminal semelhante ao de réus brancos, esse caso ilustra como a opacidade dos algoritmos pode perpetuar desigualdades e preconceitos sociais.

Outra questão relativa é a perda da autonomia dos indivíduos. Rossetti e Angeluci (2021) argumentam que os algoritmos, ao filtrarem e selecionarem informações, podem influenciar o comportamento dos usuários de maneira invisível,

plataformas como Facebook e YouTube, por exemplo, utilizam algoritmos para personalizar conteúdos, criando "câmaras de eco" que reforçam as crenças e preferências dos indivíduos, essa dinâmica não apenas limita a diversidade de perspectivas, mas também compromete a capacidade dos indivíduos de tomar decisões informadas e autônomas.

Outrossim, a opacidade algorítmica compromete a confiança nos sistemas que utilizam inteligência artificial, como apontado por Rossetti e Angeluci (2021), a incerteza sobre como os algoritmos operam gera desconfiança entre os usuários, especialmente em áreas sensíveis como saúde, justiça e segurança, e a ausência de mecanismos claros de explicabilidade e auditoria agrava essa situação, reforçando a percepção de que os algoritmos são entidades insondáveis e incontroláveis.

Para tentar sanar os problemas decorrentes da opacidade algorítmica, Rossetti e Angeluci (2021) defendem a necessidade de aumentar a transparência nos processos de desenvolvimento e operação dos algoritmos, isso inclui o uso de mecanismos de explicabilidade, que permitam aos usuários e reguladores compreender como os sistemas tomam decisões. É elementar, pois, implementar regulamentações que exijam auditorias periódicas dos sistemas algorítmicos, garantindo que eles atendam a padrões éticos e de justiça.

Outra questão ressaltada pelos autores é a importância de desenvolver uma governança ética para os algoritmos, que inclua a participação de diversos atores sociais, como programadores, usuários e especialistas em ética, Rossetti e Angeluci (2021) afirmam que a transparência não deve ser uma responsabilidade exclusiva das empresas de tecnologia, mas um compromisso coletivo que envolva governos, organizações da sociedade civil e o público em geral.

No mais, os autores ressaltam que a educação digital é fundamental para capacitar os cidadãos a lidar com os desafios da opacidade algorítmica, incluindo não apenas o entendimento básico sobre como os algoritmos operam, mas também o desenvolvimento de uma consciência crítica sobre os impactos sociais e éticos das tecnologias, somente por meio de uma abordagem integrada e participativa será possível enfrentar os desafios impostos pela opacidade algorítmica e garantir que a inteligência artificial seja utilizada de maneira justa e responsável.

Nesse sentido, a opacidade algorítmica representa um dos maiores desafios éticos da sociedade contemporânea, afetando tanto o processamento de

dados quanto a tomada de decisões em áreas essenciais, e como mostram Rossetti e Angeluci (2021), a falta de transparência nos sistemas algorítmicos compromete a justiça, a autonomia e a confiança, perpetuando desigualdades e prejudicando os direitos fundamentais, para superar esses desafios, é essencial promover a transparência, implementar mecanismos de explicabilidade e desenvolver uma governança ética robusta que envolva todos os atores sociais.

Embora os algoritmos sejam frequentemente vistos como ferramentas neutras e objetivas, é elementar reconhecer que eles são produtos de escolhas humanas e, portanto, carregam consigo os valores e preconceitos de seus criadores, como destacam Rossetti e Angeluci (2021), enfrentar a opacidade algorítmica exige não apenas avanços técnicos, mas também um compromisso ético e político com a construção de tecnologias mais justas e inclusivas. Somente assim será possível garantir que a inteligência artificial contribua para o bem-estar coletivo, em vez de reforçar desigualdades e exclusões.

O uso de inteligência artificial no Judiciário brasileiro tem o potencial de agilizar processos e aumentar a eficiência, mas também levanta sérias preocupações quanto ao respeito ao devido processo legal. A opacidade dos algoritmos pode comprometer o contraditório e a ampla defesa, dificultando que as partes compreendam ou contestem decisões automatizadas. Além disso, há o risco de reprodução de vieses discriminatórios e a indefinição sobre quem deve ser responsabilizado por eventuais erros. Assim, é fundamental que o uso da IA seja regulado com base em princípios de transparência, supervisão humana e respeito aos direitos fundamentais (Ferraz et al, 2024).

3.3.2 Viés algorítmico e risco de discriminação nas decisões

Um dos principais desafios enfrentados atualmente no uso de sistemas de inteligência artificial é a presença de vieses algorítmicos, que ocorrem quando há tratamento desigual ou discriminatório nas decisões automatizadas. Esses vieses podem comprometer a imparcialidade e a justiça, especialmente em contextos sensíveis como o Poder Judiciário. Uma das formas mais eficazes de enfrentá-los é por meio da transparência dos sistemas, permitindo que os critérios utilizados pelas máquinas sejam compreendidos e analisados. Dessa forma, torna-se possível

identificar decisões discriminatórias e corrigi-las, promovendo maior equidade e responsabilidade no uso da tecnologia (Cavalcante, 2023).

As principais causas identificadas para o viés algorítmico, segundo Cavalcante (2023), incluem a estrutura social existente, dados tendenciosos, a concepção da ferramenta e/ou a interpretação dos resultados, independentemente da origem do resultado discriminatório, é um fato que, mesmo com toda a confiança depositada no Big Data, sistemas inteligentes podem ser discriminatórios, e o viés cognitivo por si só pode causar danos, mas quando combinado com a falta de transparência ou justificativa para a decisão, esses danos são ampliados, como será discutido ao longo deste trabalho, com a análise de um caso emblemático em que isso foi observado.

O viés algorítmico é um dos maiores desafios éticos no desenvolvimento e uso de sistemas de IA, e de acordo com Cavalcante (2023), o viés ocorre quando algoritmos reproduzem ou ampliam preconceitos existentes na sociedade, no contexto do Poder Judiciário, onde a IA tem sido adotada para dar celeridade aos processos e melhorar a eficiência, os riscos do viés algorítmico se tornam ainda mais preocupantes, pois decisões enviesadas podem comprometer direitos fundamentais, como a igualdade e a dignidade humana.

Cavalcante (2023) descreve o viés algorítmico como um reflexo direto das estruturas sociais que alimentam os sistemas de IA, os algoritmos, desenvolvidos com base em grandes volumes de dados, acabam absorvendo discriminações historicamente arraigadas, como preconceitos de raça, gênero e classe social, e isso ocorre tanto pela inadequação dos dados utilizados no treinamento dos sistemas quanto por falhas nos modelos matemáticos que os estruturam, assim, ao contrário da percepção de neutralidade e imparcialidade frequentemente atribuída à tecnologia, os sistemas inteligentes podem reforçar desigualdades e injustiças.

Deste modo, o viés algorítmico representa um desafio ético central no uso de inteligência artificial, especialmente no Poder Judiciário, onde decisões enviesadas podem comprometer direitos fundamentais e perpetuar desigualdades, e como observado por Cavalcante (2023), enfrentar esse problema exige a adoção de diretrizes éticas claras, transparência nos processos algorítmicos e a promoção da diversidade nas equipes de desenvolvimento, apenas com essas medidas será possível garantir que a IA contribua para a justiça e a igualdade, em vez de reforçar preconceitos e exclusões.

Nesse diálogo, segundo Sainz, Gabardo e Ongaratto (2024) a discriminação algorítmica surge como um fenômeno que reflete desigualdades estruturais da sociedade em um contexto tecnológico, levantando questões éticas e jurídicas que exigem atenção crescente, e essa discriminação ocorre quando algoritmos reproduzem, amplificam ou criam preconceitos ao processar dados, muitas vezes de forma não intencional, mas com impactos significativos nos direitos fundamentais e no acesso à justiça, no Brasil, o debate sobre o tema está em ascensão, com foco em regulamentação, transparência e mitigação de vieses.

Sainz, Gabardo e Ongaratto (2024) explicam que a discriminação algorítmica pode ocorrer de diferentes formas, incluindo falhas técnicas, vieses nos dados utilizados para o treinamento do algoritmo ou na própria programação do sistema, de modo relevante é o caso do algoritmo COMPAS, nos Estados Unidos, que superestimava a probabilidade de reincidência criminal de réus negros em comparação com réus brancos, demonstrando como sistemas baseados em estatísticas podem perpetuar desigualdades raciais, esse tipo de discriminação não se limita ao contexto norte-americano, sendo observado também em plataformas digitais e sistemas automatizados no Brasil.

Os impactos da discriminação algorítmica afetam diretamente os direitos humanos, como igualdade, dignidade e liberdade, e reforçam desigualdades históricas. Sainz, Gabardo e Ongaratto (2024) apontam que grupos minoritários, como mulheres, negros e pessoas LGBTQIA+, são frequentemente os mais prejudicados por vieses algorítmicos, a exemplo disso, casos de discriminação no reconhecimento facial têm demonstrado maior propensão a erros ao identificar indivíduos negros, reforçando estigmas raciais e ampliando desigualdades.

No próximo capítulo, serão abordados em detalhe o viés algorítmico e a discriminação automatizada, com a análise de seus conceitos, origens e dos mecanismos de mitigação e controle desse relevante problema no uso da inteligência artificial pelo Poder Judiciário. Também será examinando o impacto dessas questões na confiança institucional e na segurança jurídica no âmbito judicial e propostos alguns mecanismos de mitigação e controle da problemática em estudo.

4 VIESES ALGORÍTMICOS

A crescente utilização de tecnologias baseadas em inteligência artificial tem trazido importantes avanços para diversos setores, inclusive o jurídico, todavia, essa modernização também levanta preocupações relevantes, como a presença de viés algorítmico e a ocorrência de discriminação automatizada, esses fenômenos ocorrem quando os sistemas reproduzem ou intensificam desigualdades sociais, seja por falhas nos dados utilizados para treiná-los ou pelas escolhas humanas em sua programação, assim, compreender essas questões é essencial para garantir o uso ético e justo dessas ferramentas.

As fontes de viés nos sistemas de IA são, em grande parte, atribuídas à qualidade e à representatividade dos dados utilizados para treiná-los, dados enviesados ou incompletos podem levar a decisões injustas, uma vez que a IA aprende a partir das informações que recebe, no contexto judicial, isso pode resultar em decisões que não apenas falham em atender aos princípios de equidade e justiça, mas que também reforçam estereótipos prejudiciais, e a análise dos dados deve, portanto, ser uma prioridade, com a implementação de políticas de governança que garantam a curadoria adequada e a diversidade das informações utilizadas nos sistemas de IA (Toledo; Pessoa, 2023).

Os impactos do viés algorítmico na equidade e na confiança no Judiciário são profundos, e a utilização de IA na tomada de decisões judiciais pode minar a confiança do público no sistema, especialmente se as partes afetadas não têm acesso a informações claras sobre como suas decisões foram tomadas, a falta de auditabilidade e a dificuldade de contestar decisões automatizadas podem levar a um sentimento de impotência entre os cidadãos, que se veem à mercê de sistemas que não compreendem, assim, é essencial que o Judiciário adote medidas que promovam a transparência e a responsabilização, garantindo que as decisões tomadas por IA sejam passíveis de revisão humana (Toledo; Pessoa, 2023).

Destaca-se, conforme Gomes, Vaz e Dias (2022), a utilização de algoritmos com a justificativa de promover maior celeridade processual não pode, em hipótese alguma, substituir a função jurisdicional exercida pelo juiz humano, a atividade de interpretar e aplicar o Direito a casos concretos exige sensibilidade, ponderação e análise contextual, características próprias da atuação humana,

embora encontrar a solução jurídica mais adequada seja um desafio, o julgador, ao identificar corretamente os pontos controvertidos da demanda, considerar os argumentos apresentados pelas partes e aplicar a norma pertinente ao caso, mesmo que discorde de seu conteúdo, tem condições de proferir uma decisão bem fundamentada e, muito provavelmente, alinhada à melhor interpretação jurídica.

Alem disso, Cavalcante (2023) defende a necessidade de uma governança ética robusta no desenvolvimento e uso de sistemas de IA, incluindo assim, a implementação de avaliações de impacto algorítmico (AIAs), que permitam identificar preconceitos nos dados e nos modelos antes de sua aplicação prática, além do mais, é essencial que os sistemas sejam auditáveis e explicáveis, garantindo que as decisões algorítmicas possam ser compreendidas e contestadas.

Outro ponto destacado por Cavalcante (2023) é a importância da diversidade nas equipes que desenvolvem sistemas de IA, e a ausência de perspectivas diversas – seja de gênero, raça ou classe social – contribui para a reprodução de vieses nos algoritmos, assim, promover a inclusão nos processos de desenvolvimento tecnológico é fundamental para reduzir desigualdades e tornar os sistemas mais justos.

Um dos principais desafios para lidar com a discriminação algorítmica é a falta de transparência nos sistemas de inteligência artificial, conhecida como opacidade algorítmica, Sainz, Gabardo e Ongaratto (2024) explicam que essa opacidade impede que usuários, desenvolvedores e reguladores compreendam os critérios utilizados pelos algoritmos para tomar decisões, dificultando a identificação de vieses e a responsabilização por seus efeitos, e a falta de explicabilidade, ou seja, a capacidade de rastrear e compreender como o algoritmo chegou a determinado resultado, é um obstáculo significativo para a promoção de justiça e equidade.

No Brasil, a regulamentação sobre o uso de inteligência artificial ainda está em estágio inicial, mas medidas como a Lei Geral de Proteção de Dados (LGPD) e a Resolução nº 332/2020 do Conselho Nacional de Justiça (CNJ) representam passos importantes, e a LGPD, por exemplo, prevê o direito dos titulares de dados a solicitar explicações sobre decisões automatizadas, enquanto a Resolução nº 332/2020 estabelece diretrizes para o uso ético de algoritmos no sistema judicial, no entanto, Sainz, Gabardo e Ongaratto (2024) alertam que essas

regulamentações ainda carecem de mecanismos robustos para lidar com a opacidade e a discriminação algorítmica de forma eficaz.

A Resolução nº 332/2020 do Conselho Nacional de Justiça (CNJ) representou um marco na regulamentação do uso da Inteligência Artificial (IA) no Poder Judiciário brasileiro, publicada em 21 de agosto de 2020, a norma estabelece diretrizes para o desenvolvimento, aplicação e monitoramento de sistemas de IA dentro dos tribunais, assegurando que sua implementação ocorra de forma ética, transparente e alinhada aos princípios fundamentais da Justiça.

Destarte, a motivação para a criação dessa resolução surgiu da crescente digitalização do Judiciário e da adoção de ferramentas tecnológicas para otimizar a prestação jurisdicional, com o avanço da inteligência artificial, os tribunais passaram a utilizar algoritmos para automatizar tarefas repetitivas, analisar grandes volumes de dados e até mesmo auxiliar na formulação de decisões, o que gerou a necessidade de diretrizes claras para garantir a segurança, transparência e confiabilidade dessas tecnologias.

Nesse sentido, a Resolução nº 332/2020 do Conselho Nacional de Justiça estabelece diretrizes fundamentais para o uso da inteligência artificial no Poder Judiciário, sendo um passo importante para que a sua implementação ocorra de maneira ética, transparente e segura. Entre as principais diretrizes destacam-se a transparência e a explicabilidade, assegurando que os sistemas de IA utilizados no Judiciário sejam compreensíveis e auditáveis, isso significa que deve haver clareza sobre o funcionamento dos algoritmos, os dados utilizados e o processamento das decisões, além do mais, os cidadãos têm o direito de saber quando uma decisão judicial ou administrativa foi influenciada por um sistema de IA, garantindo maior confiança no uso dessas tecnologias.

Outro destaque é o da não discriminação e imparcialidade, que impõe o desenvolvimento e monitoramento das tecnologias para evitar vieses algorítmicos que possam levar a decisões discriminatórias, para isso, devem ser adotadas medidas que garantam a justiça e a equidade dos sistemas utilizados, e a segurança e a privacidade também são prioridades, alinhando o uso da inteligência artificial à Lei Geral de Proteção de Dados (LGPD – Lei nº 13.709/2018), assim, a resolução exige a adoção de mecanismos para proteger as informações pessoais dos usuários e evitar qualquer vazamento de dados.

A supervisão humana e a responsabilidade são aspectos indispensáveis para o uso adequado da IA no Judiciário, e as decisões geradas ou assistidas por inteligência artificial não podem ser completamente autônomas, devendo sempre contar com a supervisão de juízes ou servidores, também é essencial que a responsabilidade pelo uso desses sistemas esteja claramente definida, garantindo a possibilidade de correção de erros ou falhas, e a eficiência e a melhoria contínua são igualmente destacadas, enfatizando que a tecnologia deve ser utilizada para aprimorar a celeridade processual e otimizar a gestão de recursos no Judiciário, para tanto, o uso da IA precisa ser constantemente avaliado e aprimorado com base nos resultados obtidos.

Os impactos dessa resolução no Judiciário brasileiro são amplos e relevantes, entre os benefícios, destaca-se a automatização de tarefas repetitivas, permitindo que sistemas de IA classifiquem petições, organizem processos, analisem jurisprudências e até sugiram minutas de decisões, liberando magistrados e servidores para atividades mais complexas, com isso, há uma redução da morosidade processual, aumentando a agilidade na tramitação das ações judiciais.

A transparência e a explicabilidade das decisões automatizadas tornam o sistema mais acessível, permitindo que cidadãos e advogados compreendam melhor os critérios adotados pelos algoritmos, outro impacto positivo é a padronização das decisões judiciais, contribuindo para maior uniformidade na aplicação do Direito e reduzindo divergências interpretativas em casos semelhantes, dessa maneira, a Resolução nº 332/2020 representa um avanço significativo na modernização do Judiciário brasileiro, equilibrando inovação tecnológica e proteção aos direitos fundamentais.

A Resolução nº 332/2020 do CNJ é um passo essencial para garantir que a inteligência artificial seja utilizada no Poder Judiciário de maneira ética, eficiente e transparente, ao estabelecer diretrizes claras para o uso dessas tecnologias, a norma busca equilibrar inovação e segurança jurídica, garantindo que o avanço tecnológico não comprometa direitos fundamentais dos cidadãos.

Contudo, sua efetiva implementação exige investimentos em pesquisa, treinamento de profissionais, monitoramento contínuo dos algoritmos e ajustes regulatórios conforme novas demandas surgirem, e a devida utilidade dessa regulamentação dependerá da capacidade do Judiciário de adaptar-se às transformações tecnológicas sem abrir mão da imparcialidade, da equidade e da

transparência na aplicação da Justiça. Recentemente a Resolução nº 332/2020 do CNJ foi atualizada pela Resolução 615/2025, que estabelece criterios mais objetivos e claros para o uso da IA no Poder Judiciario, sendo abordada mais detalhadamente nos tópicos a seguir .

4.1 Definição de viés algorítmico e discriminação automatizada

Inicialmente, ao proceder com a análise acerca da definição de viés algorítmico, é importante mencionar que o mesmo ocorre nos casos em que os algoritmos de aprendizado de máquina geram resultados tendenciosos ou desiguais, muitas vezes reproduzindo ou intensificando desigualdades sociais, econômicas, raciais e de gênero presentes na sociedade (Jonker; Rogers, 2024).

Viés, pois, se refere a julgamentos falhos que resultam em análises incorretas, o qual pode ser cognitivo, quando realizado por seres humanos, ou algorítmico, na hipótese em que é gerado por máquinas, cujo estudo é inerente para compreender como as percepções e decisões são formadas, bem como a maneira como diversas áreas da atividade humana restam afetadas:

Por viés entendem-se os fenômenos de julgamento falho que levam ao que se consideram análises falsas. Podem ser cognitivos (realizado por seres humanos) ou algorítmicos (realizado por máquinas. O estudo dos vieses cognitivos é essencial para entendermos como são construídas as percepções e as tomadas decisões, o que tem impacto em todas as áreas da atividade humana, inclusive com influência direta nos vieses algorítmicos (Ávila; Corazza, 2022, p. 187).

A inteligência artificial busca replicar funções cognitivas humanas por meio de máquinas e, mais especificamente no âmbito dessa área, o aprendizado de máquina utiliza algoritmos para analisar dados e aprender de forma autônoma. Tome-se como exemplo o fato de ao treinar um sistema para identificar livros adequados para crianças, é importante considerar que, se os dados usados para treinamento forem imprecisos, limitados ou enviesados, o sistema pode gerar resultados falhos, o que demonstra como a qualidade dos dados impacta diretamente nas decisões tomadas por sistemas baseados em IA (UFSM, 2020).

Nesse passo, “O viés algorítmico é especialmente preocupante quando presente em sistemas de IA que apoiam decisões que afetam significativamente a vida das pessoas, como nas áreas de saúde, aplicação da lei e recursos humanos” (Jonker; Rogers, 2024, p. 1). Cabe expor que o viés é capaz de adentrar nos algoritmos de maneiras diferenciadas, como ocorre, por exemplo, nos casos em que os treinamentos de dados sejam distorcidos ou limitados, bem como na hipótese em que forem evidenciadas decisões subjetivas.

Logo, ideia de viés algorítmico se encontra intimamente vinculado a um reflexo de desigualdades sociais, pondo em evidência, neste particular, como algoritmos podem perpetuar preconceitos de raça, gênero e classe, incorporando-se à discriminação e/ou preconceito nos algoritmos de inteligência artificial, muitas vezes decorrente de dados enviesados ou de decisões subjetivas no design dos sistemas (Simões-Gomes; Roberto; Mendonça, 2020).

Não se pode olvidar, ademais, que o viés pode ser dividido da seguinte forma: o viés histórico, que envolve estruturas hierárquicas da sociedade que são reforçadas e naturalizadas pelos modelos de aprendizado de máquina, bem como o viés de base de dados que ocorre durante a criação das bases de dados, podendo refletir tendências ou desequilíbrios, seja por intervenção humana ou processos automatizados. Por fim, tem-se o viés algorítmico, que amplifica esses dois tipos anteriores através do próprio modelo de aprendizado de máquina (Gonçalves, 2022).

Oportuno ainda lembrar que o viés é apresentado em quatro categorias, sendo que, a primeira delas é evidenciado o fato de que o algoritmo precisa escolher quais dados serão utilizados, o que pode gerar exclusões e distorções, o que é comum em softwares que utilizam inteligência artificial, como programas de navegação. Um exemplo simples é o sistema de navegação que prioriza rotas rápidas para motoristas, mas pode desconsiderar opções para ciclistas ou pedestres, prejudicando quem não tem carro, originando-se, pois, quando certos dados são considerados mais relevantes do que outros, afetando a equidade das decisões do algoritmo (Carvalho, 2020).

Aliado a isso, viés que de igual forma é frequente consiste no fato dos dados apresentarem algum tipo de erro, ou, na hipótese em que se encontram desatualizados ou descritos de forma incompleta. Nesse passo, “No caso do trajeto descrito, ainda que o sistema oferecesse a rota utilizando o transporte público, de nada adiantaria se as informações de horários não estivessem atualizadas,

novamente prejudicando usuários que não utilizam carros” (Carvalho, 2020, p. 42).

Ademais, existe uma terceira categoria de viés algorítmico, atrelada aos dados de entrada, chamada viés de seleção, estabelecendo-se nos casos em que os dados inseridos no sistema não refletem de forma precisa a realidade. Por exemplo, se um sistema de navegação for treinado majoritariamente com dados de telefones de última geração, a precisão pode ser comprometida em áreas mais carentes, já que a população de menor poder aquisitivo não utilizaria os serviços de forma igual, recebendo, assim, um serviço de qualidade inferior:

Há ainda uma terceira categoria de viés algorítmico, relacionado aos dados de entrada, que compreende o viés de seleção (que talvez seja o mais frequente na prática), ocorrendo quando os dados inseridos no sistema não correspondem à realidade. No exemplo escolhido acima, se os dados forem majoritariamente obtidos a partir de telefones de última geração, é provável que a precisão do sistema seja prejudicada nas áreas mais pobres da região, que receberiam um serviço de menor qualidade, já que essa população de menor poder aquisitivo não participaria do serviço da mesma forma (Carvalho, 2020, p. 42).

Salienta-se que um tipo comum de viés envolve a perpetuação de vieses históricos por meio de ciclos de retroalimentação (*feedback loops*) e, quando um modelo de aprendizado é enviesado, ele tende a replicar esses vieses indefinidamente, uma vez que a inteligência artificial aprende com os padrões detectados, como no caso dos algoritmos de seleção de candidatos a emprego, onde um sistema pode desconsiderar candidatos qualificados com base em características, como faixa etária, que não correspondem ao perfil usualmente contratado pela empresa (Carvalho, 2020).

Noutro ponto, ao promover uma análise acerca da discriminação automatizada, enfatiza-se, inicialmente, que ao refletir sobre o conceito de discriminação, frequentemente se pensa em situações em que uma pessoa é excluída por uma característica específica, mas, no entanto, o escopo pretendido é averiguar os resultados discriminatórios baseados no fato de uma pessoa pertencer a um grupo, sendo julgada pelas características desse grupo, em vez de ser avaliada de forma individual. Nesse contexto, a pessoa é vista apenas como parte de um conjunto, sem consideração pelas suas qualidades pessoais (Mendes; Mattiuzzo, 2019).

Um dos principais argumentos a favor das decisões automatizadas é que elas poderiam superar as falhas e vieses dos julgamentos humanos, usando critérios

objetivos e a linguagem matemática, mas, no entanto, a realidade tem mostrado que a inteligência artificial e os processos algorítmicos não apenas falham em corrigir erros subjetivos, mas podem até replicar e intensificar preconceitos sociais, resultando em discriminação algorítmica, que prejudica a igualdade de tratamento, especialmente para grupos vulneráveis (Requião; Costa, 2022).

Nesse passo, “pode decorrer tanto do resultado da programação efetuada para o software quanto ser uma reprodução de preconceitos e discriminações já existentes na sociedade, e que adquirem um potencial de perpetuação de vieses por meio de diferentes tratamentos de dados” (Sainz; Gabardo; Ongaratto, 2024, p. 260).

A nomenclatura discriminação automatizada é usada para se referir tanto a cenários com afirmações estatisticamente inconsistentes quanto àqueles em que as afirmações, embora lógicas do ponto de vista estatístico, tratam os indivíduos de maneira generalizada, sem consideração pelas suas características individuais, tratando-os apenas como parte de um grupo. Desta feita, mesmo classificações estatisticamente válidas podem ser injustas, dependendo do contexto (Mendes; Mattiuzzo, 2019).

Observe que as discriminações automatizadas podem ser classificadas em quatro tipos principais, a saber: discriminação por erro estatístico; discriminação por generalização; discriminação pelo erro de informações sensíveis; e, discriminação limitadora do exercício de direitos.

Os autores identificam quatro tipos principais de discriminação algorítmica, por erro estatístico, quando há falhas na coleta ou processamento de dados; por generalização, quando o algoritmo não considera a diversidade de características individuais; por uso de informações sensíveis, envolvendo variáveis como raça ou gênero; e por impacto limitador do exercício de direitos, em que os resultados algorítmicos restringem direitos e oportunidades, todos esses tipos estão presentes em casos amplamente documentados e discutidos em estudos acadêmicos, revelando a necessidade urgente de regulamentação e mitigação (Sainz; Gabardo; Ongaratto, 2024).

Esclarece-se que a discriminação por erro estatístico ocorre quando há falhas genuínas nos dados ou no código do algoritmo, incluindo-se, aqui, dados incorretamente coletados, problemas no processamento dos dados ou erros no cálculo. Essencialmente, é um tipo de discriminação causado por enganos cometidos pelos profissionais responsáveis pela criação do algoritmo:

Discriminação por erro estatístico – todo e qualquer erro que seja genuinamente estatístico, abrangendo desde dados incorretamente coletados, até problemas no código do algoritmo, de forma que ele falhe em contabilizar parte dos dados disponíveis, contabilize-os de forma incorreta, etc. Basicamente, é o tipo de discriminação que decorre de um erro cometido pelos engenheiros ou cientistas de dados responsáveis pelo desenho do algoritmo (Mendes; Mattiuzzo, 2019, p. 51-52).

Por sua vez, a discriminação por generalização ocorre quando, apesar de um modelo funcionar corretamente e ser estatisticamente válido, ele classifica incorretamente indivíduos em determinados grupos. Isso acontece, por exemplo, quando o modelo usa apenas o endereço de uma pessoa para decidir sua elegibilidade para um empréstimo, associando-a a um grupo de acordo com a vizinhança, mesmo que ela tenha características atípicas, como uma renda diferente da média local, sendo comum em modelos probabilísticos, onde exceções à média podem ser ignoradas (Mendes; Mattiuzzo, 2019).

Quanto à discriminação pela utilização de informações sensíveis, “a razão pela qual consideramos esta categoria como discriminatória, embora muitas vezes seja estatisticamente correta, é porque ela se baseia em dados ou proxies legalmente protegidos” (Mendes; Mattiuzzo, 2019, p. 52). Exemplo disso é no caso em que a discriminação ocorre quando um algoritmo utiliza informações sobre a religião de um indivíduo para definir seu crédito, o que é proibido pela Lei do Cadastro Positivo no Brasil. Nesse passo, tem-se que duas características tornam esse perfilamento discriminatório: o uso de dados sensíveis e a basear-se em características endógenas ou em grupos historicamente discriminados.

Derradeiramente, salienta-se que a discriminação limitadora do exercício de direitos acontece no caso em que apesar de resultados estatisticamente corretos, o uso de informações por um algoritmo restringe o acesso a um direito fundamental. Logo, o problema não está nos dados em si, mas na conexão entre a informação utilizada e o impacto significativo no exercício de um direito, tornando o uso potencialmente discriminatório se esse direito for excessivamente afetado (Mendes; Mattiuzzo, 2019).

Sucintamente, considera-se que a discriminação por erro estatístico ocorre devido a falhas na coleta de dados ou no código do algoritmo, como erros

amostrais ou desequilíbrios nos cálculos, ao passo que a discriminação por generalização surge quando os modelos estatísticos não conseguem considerar a diversidade individual, ignorando a margem de erro. Já a discriminação por uso de informações sensíveis envolve a utilização de características intrínsecas de grupos discriminados, podendo violar a proteção de dados ou desconsiderar variáveis essenciais e, já a discriminação limitadora de direitos decorre do impacto negativo que o algoritmo tem sobre os direitos do indivíduo (Sainz; Gabardo; Ongaratto, 2024).

Há entendimento no sentido de que as discriminações automatizadas podem ocorrer de duas formas, isto é, na hipótese em que refletem preconceitos humanos, bem como nos casos em que entram em contato com dados que comportam vieses preconceituosos.

Quanto ao fato de refletirem em preconceitos humanos, os vieses algorítmicos podem surgir quando as máquinas refletem os valores humanos implícitos na programação, mesmo que de forma não intencional e, como os vieses fazem parte do pensamento humano, algoritmos criados por programadores tendenciosos provavelmente também os terão. Como exemplo, pode ser citado o que ocorreu com Joy Buolamwini, do MIT Media Lab, que percebeu que um software de reconhecimento facial não conseguia identificar seu rosto devido à falta de treinamento do algoritmo para reconhecer uma ampla variedade de tons de pele e estruturas faciais, especialmente os de pessoas negras (Requião; Costa, 2022).

No segundo caso, o viés algorítmico pode surgir da interação entre o software de aprendizado de máquina e os dados com os quais ele é alimentado e, mesmo sem a intervenção direta do programador, o viés pode se infiltrar se os dados forem de baixa qualidade ou pouco confiáveis, já que esses dados são coletados da sociedade (como em redes sociais), o que pode fazer com que o sistema reforce padrões discriminatórios presentes no banco de dados (Requião; Costa, 2022).

Um exemplo emblemático é o caso do software COMPAS, utilizado no sistema judicial norte-americano para prever a reincidência criminal. O sistema foi criticado por classificar de forma desproporcional réus negros como de maior risco de reincidência, mesmo em situações em que réus brancos com histórico criminal mais grave eram avaliados como de menor risco, segundo Cavalcante (2023), este tipo de viés reflete a falta de transparência nos algoritmos, que não permitem

compreender como as decisões são tomadas, dificultando a identificação e a correção de erros.

Cavalcante (2023) alerta que aplicação de IA sem diretrizes éticas claras pode acarretar graves violações de direitos fundamentais, e a falta de transparência e explicabilidade nos resultados algorítmicos é um dos principais desafios apontados, pois impede que advogados e partes compreendam os critérios que levaram a determinada decisão judicial.

Além do que Cavalcante (2023) enfatiza que a discriminação algorítmica não se limita a casos criminais, em áreas como concessão de benefícios previdenciários ou decisões trabalhistas, sistemas enviesados podem prejudicar grupos mais vulneráveis, como mulheres, negros e pessoas de baixa renda, esse risco é agravado pela ausência de mecanismos eficazes de fiscalização e auditoria dos sistemas utilizados nos tribunais brasileiros.

Ademais, os autores destacam que a discriminação algorítmica não se limita a contextos criminais, mas também ocorre em áreas como recrutamento de emprego, acesso ao crédito e moderação de conteúdo em plataformas digitais, em casos como o do sistema de seleção da Amazon, que favorecia candidatos masculinos devido à predominância de currículos de homens no banco de dados, evidencia-se como os algoritmos podem perpetuar preconceitos estruturais, no âmbito brasileiro, sistemas de pontuação de crédito (credit scoring) também têm sido criticados por desconsiderar variáveis socioeconômicas que afetam populações marginalizadas (Sainz; Gabardo ; Ongaratto, 2024).

Continuando na ótica de Sainz, Gabardo e Ongaratto (2024), segundo esses autores, para enfrentar os impactos da discriminação algorítmica, ele defendem a adoção de estratégias que promovam transparência, diversidade e governança ética no desenvolvimento e uso de algoritmos, e uma perspectiva sugerida é o desenviesamento algorítmico, que busca identificar e corrigir preconceitos nos dados e nos modelos utilizados pelos sistemas de IA, de modo que, a implementação de inteligência artificial explicável (Explainable AI ou XAI) é essencial para aumentar a transparência e a confiabilidade dos sistemas, permitindo que seus processos sejam auditados e compreendidos.

Os autores também enfatizam a importância da diversidade nas equipes que desenvolvem algoritmos, argumentando que a inclusão de perspectivas variadas pode reduzir a reprodução de preconceitos, no contexto brasileiro,

iniciativas como o grupo de estudo Diversidade e Tecnologia, criado pelo Ministério Público do Trabalho, exemplificam esforços para promover a conscientização sobre discriminação algorítmica e desenvolver soluções inclusivas (Sainz; Gabardo; Ongaratto, 2024).

A regulamentação precisa ser fortalecida para garantir que os algoritmos sejam desenvolvidos e utilizados de forma ética e responsável, Sainz, Gabardo e Ongaratto (2024) sugerem a criação de marcos legais específicos para a inteligência artificial, com foco na transparência, no direito à explicação e na responsabilização por danos causados por decisões automatizadas, e a adoção de avaliações de impacto algorítmico (AIAs) também é recomendada como uma forma de identificar riscos de discriminação antes da implementação dos sistemas.

A discriminação algorítmica representa um desafio ético e jurídico significativo na era digital, exigindo atenção especial no Brasil e no mundo, conforme argumentam Sainz, Gabardo e Ongaratto (2024), a falta de transparência e os vieses presentes nos algoritmos têm o potencial de perpetuar desigualdades e comprometer direitos fundamentais, especialmente de grupos historicamente marginalizados, para mitigar esses impactos, é essencial promover a transparência algorítmica, adotar práticas de desviesamento e fortalecer a regulamentação sobre o uso ético de inteligência artificial.

Cambi e Amaral (2023) destacam que os algoritmos(ou melhor, modelos algorítmicos) não são neutros, pois são programados por seres humanos que, conscientemente ou não, carregam vieses cognitivos e culturais, esses vieses acabam sendo incorporados aos sistemas, perpetuando discriminações que deveriam ser combatidas, um exemplo citado pelos autores é a aplicação da IA em recrutamentos, onde algoritmos podem favorecer candidatos de um determinado gênero ou grupo social, dependendo das características predominantes nos dados históricos utilizados para treinamento.

Cambi e Amaral (2023) afirmam que os sistemas de aprendizado de máquina (machine learning) possuem a capacidade de identificar padrões em grandes volumes de dados, mas essa habilidade também os torna suscetíveis à repetição de padrões discriminatórios, mesmo quando a programação inicial não inclui preconceitos explícitos, o uso de bases de dados desbalanceadas ou enviesadas leva à reprodução de desigualdades, ocorrendo assim, por exemplo, em algoritmos utilizados para avaliar a concessão de crédito, que podem negar

empréstimos a determinados grupos minoritários com base em características associadas ao histórico econômico de suas comunidades.

Os autores relatam que esse tipo de discriminação algorítmica é especialmente preocupante em contextos judiciais, onde decisões automatizadas podem impactar diretamente a liberdade e a dignidade de indivíduos, outro exemplo citado por Cambi e Amaral (2023) é o uso do software COMPAS nos Estados Unidos, já citado anteriormente, este caso ilustra como a IA pode reforçar preconceitos raciais, comprometendo a imparcialidade e a justiça.

Ademais, a relação entre discriminação algorítmica e direitos humanos fundamentais é um dos aspectos mais críticos do debate sobre a inteligência artificial, Cambi e Amaral (2023) explicam que as decisões algorítmicas podem violar direitos como a igualdade, a dignidade e o acesso à justiça, eles ressaltam que grupos historicamente marginalizados, como negros, mulheres e pessoas LGBTQIA+, são os mais afetados por sistemas discriminatórios, que muitas vezes os excluem de oportunidades ou os tratam de forma injusta.

Os autores destacam que a Constituição Federal brasileira protege esses direitos por meio de dispositivos como o artigo 5º, que garante a igualdade de todos perante a lei, todavia, a aplicação de algoritmos enviesados pode comprometer essa proteção, criando novos desafios jurídicos e éticos, Cambi e Amaral (2023) apontam que, no contexto do Judiciário, a discriminação algorítmica pode influenciar decisões judiciais, prejudicando o devido processo legal e a ampla defesa.

Os autores também ressaltam que a opacidade dos algoritmos – ou seja, a dificuldade de compreender como tomam decisões, agrava o problema da discriminação, pois impede que as partes afetadas identifiquem e contestem prejuízos, essa falta de transparência compromete não apenas os direitos individuais, mas também a confiança pública nas instituições que utilizam IA (Cambi; Amaral, 2023).

Uma medida essencial apontada pelos autores é o desenviesamento das bases de dados utilizadas para treinar algoritmos. Isso significa revisar e corrigir dados que contenham padrões históricos de discriminação, garantindo que os sistemas de IA sejam treinados com informações equilibradas e representativas, Cambi e Amaral (2023) também defendem a criação de equipes interdisciplinares para o desenvolvimento de algoritmos, incluindo especialistas em direitos humanos e diversidade, para reduzir o risco de vieses.

Esses autores também sustentam que a discriminação algorítmica é um dos maiores desafios éticos e jurídicos relacionados à inteligência artificial, especialmente no que diz respeito à proteção dos direitos humanos fundamentais, assim, como mostram eles, a opacidade e os vieses presentes nos algoritmos podem perpetuar desigualdades e comprometer princípios básicos de justiça e igualdade, para mitigar esses riscos, é fundamental promover a transparência, corrigir vieses nos dados e adotar regulamentações que assegurem o uso ético e responsável da IA, outrossim, os autores destacam a importância de incluir a pauta de direitos humanos nos debates sobre a regulamentação da IA, garantindo que as soluções desenvolvidas reflitam os valores democráticos e a diversidade da sociedade brasileira (Cambi; Amaral, 2023),

Embora avanços importantes já tenham sido feitos, como a LGPD e a Resolução nº 332/2020 do CNJ, ainda há um longo caminho a percorrer para garantir que os algoritmos sejam ferramentas de inclusão e justiça, em vez de perpetuadores de desigualdades, e a pesquisa acadêmica e a formulação de políticas públicas devem continuar a evoluir em resposta aos desafios impostos pela discriminação algorítmica.

4.2 Fontes de viés nos sistemas de IA: Dados

Os sistemas de inteligência artificial são altamente dependentes dos dados utilizados em sua construção e operação, todavia, quando esses dados carregam distorções, omissões ou representações desiguais da realidade social, tornam-se fontes significativas de viés, essa problemática compromete a neutralidade e a justiça dos resultados produzidos pela IA, especialmente em áreas sensíveis como segurança pública, saúde e relações de consumo.

Outrossim, os sistemas de inteligência artificial dependem da qualidade e representatividade dos dados utilizados para treiná-los, todavia, esses dados frequentemente carregam consigo preconceitos e desigualdades históricas, sociais e culturais, que acabam sendo reproduzidos ou até amplificados pelos algoritmos, assim o viés nos dados é uma das principais causas de discriminação em sistemas

de IA, afetando desde a imparcialidade de decisões automatizadas até a confiabilidade das aplicações tecnológicas em diferentes contextos, desta maneira, compreender as fontes de viés nos dados é fundamental para mitigar os impactos negativos associados ao uso de IA.

Segundo Crawford (2021), os dados frequentemente refletem desigualdades sociais e preconceitos históricos, o que faz com que os sistemas de IA reproduzam essas mesmas distorções, a exemplo disso, em sistemas de recrutamento automatizados, dados históricos que demonstram discriminação contra mulheres em posições de liderança podem levar os algoritmos a priorizar candidatos masculinos, perpetuando desigualdades de gênero no mercado de trabalho, e a autora destaca que, enquanto os dados usados para treinar os modelos permanecerem contaminados por preconceitos históricos, os sistemas de IA continuarão a replicar esses padrões.

Neste cenário, os dados são elementares no desenvolvimento e no desempenho de sistemas da IA, contudo, quando os dados utilizados para treinar algoritmos apresentam falhas, lacunas ou preconceitos, esses problemas são incorporados e amplificados pelas tecnologias, resultando em decisões enviesadas e prejudiciais, e a análise das fontes de viés relacionadas aos dados é essencial para compreender como essas tecnologias podem reforçar desigualdades sociais e econômicas, além de propor estratégias para mitigar tais problemas.

Uma das principais fontes de viés nos dados é a representação inadequada de grupos sociais, deste modo, como explica O'Neil (2016), os dados frequentemente refletem as desigualdades históricas e estruturais da sociedade, como discriminação racial, de gênero ou socioeconômica, assim, em sistemas de IA aplicados a processos seletivos, os dados de treinamento podem priorizar perfis de candidatos que historicamente foram contratados, geralmente homens brancos, excluindo automaticamente mulheres e minorias de oportunidades, esse tipo de viés ocorre porque os dados utilizados são um "espelho" das práticas passadas, e os algoritmos acabam aprendendo a replicar esses padrões.

Outra questão relevante é o viés introduzido pela coleta e curadoria de dados, segundo Benjamin (2019), os métodos de coleta de dados podem ser enviesados pela forma como são projetados, por vezes os dados são coletados de populações específicas, ignorando a diversidade de experiências e contextos de outros grupos, por exemplo, em sistemas médicos baseados em IA, os dados de

treinamento frequentemente são coletados de populações urbanas e economicamente favorecidas, resultando em diagnósticos menos precisos para pessoas de áreas rurais ou em condições de pobreza, assim, a falta de representatividade nos dados perpetua desigualdades no acesso e na qualidade dos serviços oferecidos.

Além disso, o viés pode surgir durante os processos de pré-processamento e limpeza de dados, em que decisões humanas são questões substanciais, conforme argumentado por Broussard (2018), a exclusão de registros incompletos ou a normalização de variáveis pode criar vieses involuntários, ademais, cita-se que se registros de determinados grupos forem mais propensos a conter informações ausentes (devido a barreiras sociais ou econômicas), esses grupos podem ser desproporcionalmente excluídos do conjunto de dados, esse tipo de viés afeta diretamente a equidade dos sistemas de IA, pois os algoritmos não conseguem lidar adequadamente com a diversidade da população real.

Destarte, O'Neil (2016) destaca que a transparência nos processos de coleta e a inclusão de perspectivas diversas são medidas essenciais para minimizar os impactos negativos dos vieses, de modo que, Benjamin (2019) argumenta que as equipes responsáveis pelo desenvolvimento de sistemas de IA devem ser compostas por profissionais de diferentes áreas e origens, garantindo uma abordagem mais ampla e ética, assim, a auditoria contínua dos dados e dos sistemas é indispensável para garantir que os algoritmos reflitam valores de justiça e ética.

Outra fonte de viés está relacionada às escolhas feitas durante o design e desenvolvimento dos algoritmos, conforme aponta Eubanks (2018), as decisões tomadas pelos desenvolvedores, como a escolha de métricas de desempenho e a definição de parâmetros do modelo, podem introduzir vieses involuntários, por vezes, os desenvolvedores priorizam a eficiência e o desempenho técnico dos sistemas, sem considerar adequadamente os impactos sociais e éticos das decisões automatizadas, isso é evidente em sistemas de avaliação de risco usados na justiça criminal, que tendem a penalizar desproporcionalmente indivíduos pertencentes a minorias raciais devido a decisões algorítmicas que não foram projetadas para lidar com essas complexidades sociais.

O viés de exclusão ocorre quando determinados grupos ou informações relevantes são deixados de fora do conjunto de dados ou do processo de

modelagem, Noble (2018) destaca que a exclusão de vozes e perspectivas de grupos marginalizados nos dados de treinamento pode levar a resultados discriminatórios, nesse escopo, ela cita que em sistemas de busca na internet, algoritmos podem apresentar resultados enviesados que refletem estereótipos prejudiciais, pois os dados utilizados para treinar esses sistemas não são neutros, ademais, a autora argumenta que os vieses nos sistemas de IA não são apenas questões técnicas, mas também estão profundamente conectados a dinâmicas de poder e desigualdade estrutural.

Nesse sentido, Crawford (2021) enfatiza a importância de incluir especialistas em ciências sociais e representantes de comunidades impactadas no desenvolvimento de sistemas de IA, enquanto Eubanks (2018) sugere que as comunidades afetadas devem ter voz ativa nas decisões relacionadas à implementação dessas tecnologias, além do que, a transparência nos processos de design e tomada de decisão é fundamental para garantir que os sistemas de IA sejam auditáveis e responsáveis.

Uma das principais fontes de viés nos dados é a sub-representação de grupos específicos, especialmente aqueles que já enfrentam desigualdades estruturais, de acordo com Barocas, Hardt e Narayanan (2019), quando determinados grupos sociais, como minorias raciais ou mulheres, não estão adequadamente representados nos dados de treinamento, os algoritmos tendem a apresentar um desempenho inferior para esses grupos, esse problema é agravado quando os dados refletem padrões históricos de discriminação, perpetuando desigualdades já existentes, por exemplo, sistemas de reconhecimento facial treinados com dados predominantemente compostos por rostos de pessoas brancas apresentam taxas mais altas de erros ao identificar pessoas negras.

Os dados podem ser enviesados devido à forma como são coletados, conforme Obermeyer et al. (2019) destacam, métodos de coleta inadequados ou focados apenas em populações específicas podem introduzir distorções nos sistemas de IA, a exemplo disso dado por esses autores é o uso de dados de saúde para prever necessidades médicas futuras, eles identificaram em sua pesquisa que estudos revelaram que sistemas de IA projetados para prever cuidados médicos priorizaram pacientes brancos em detrimento de pacientes negros, porque os dados de treinamento consideraram custos médicos como indicador de saúde, ignorando que a população negra historicamente tem menos acesso ao sistema de saúde e,

portanto, apresenta custos menores nos registros hospitalares.

Para mitigar o viés nos dados e, conseqüentemente, nos sistemas de IA, é essencial adotar práticas rigorosas de curadoria e governança de dados. Isso inclui a coleta de dados representativos e diversificados, a análise contínua de possíveis distorções e a implementação de auditorias regulares nos sistemas de IA, de modo que, Barocas, Hardt e Narayanan (2019) enfatizam a importância de incorporar princípios éticos na construção e no uso de algoritmos, garantindo que as decisões automatizadas sejam justas e inclusivas, tais medidas são indispensáveis para promover a equidade e a confiança no uso da inteligência artificial em diferentes contextos.

Neste cenário, segundo Meireles (2021) nos últimos anos, a humanidade passou a produzir e armazenar volumes inéditos de informações, em uma diversidade de formatos como textos, imagens, vídeos e áudios, esses dados são captados por uma ampla gama de tecnologias — desde drones e sistemas de reconhecimento facial até cartões de crédito e sensores variados, como os de temperatura e movimento, e o ponto central dessa transformação está na mudança do registro analógico para o digital, o que altera significativamente a maneira como as informações são tratadas, nesse novo cenário, o direito à privacidade, por si só, já não é suficiente para enfrentar os desafios impostos pela enorme quantidade de dados pessoais gerados diariamente, surge, então, a necessidade de garantir o direito à proteção de dados pessoais, como forma de proteger os indivíduos contra o uso indevido de suas informações.

Dessarte, segundo Meireles (2021) ,tal proteção não se limita à vigilância estatal, mas abrange também práticas como a manipulação de preferências de consumo, influências políticas e impactos diretos sobre o exercício da cidadania e a autonomia informacional, a discussão sobre a regulação da propriedade dos dados representa um dos desafios mais complexos enfrentados pela sociedade atual, especialmente diante do tratamento massivo e, muitas vezes, indiscriminado de informações pessoais, no cotidiano digital, exemplos como a publicidade direcionada, os sistemas de recomendação e os filtros de conteúdo em plataformas online evidenciam como os algoritmos são utilizados para moldar e influenciar as decisões individuais.

Nesse processo, os dados pessoais são continuamente captados, processados, analisados e manipulados por sistemas computacionais, muitas vezes

sem transparência quanto às suas finalidades, esses algoritmos, sejam eles mais simples ou dotados de inteligência, operam com dados de múltiplos formatos de maneira constante, e a distinção principal entre eles está no nível de autonomia: enquanto os algoritmos convencionais são desenvolvidos para resolver tarefas específicas por meio de instruções fixas, os algoritmos inteligentes — como aqueles baseados em inteligência artificial ou aprendizado de máquina — são projetados para aprender com os dados e adaptar suas soluções, tornando-se capazes de enfrentar problemas complexos de maneira mais dinâmica (Meireles, 2021).

A construção e o aprimoramento de um software de inteligência artificial dependem fortemente da utilização de bancos de dados. Após o desenvolvimento do modelo, os dados são inseridos no sistema, possibilitando o processo conhecido como aprendizado de máquina (*machine learning*). Nessa etapa, a inteligência artificial analisa as informações recebidas conforme as diretrizes estabelecidas pelo algoritmo, com o objetivo de reconhecer padrões e, a partir deles, gerar previsões. A eficácia desses sistemas está diretamente relacionada à qualidade dos dados utilizados. Como esses dados refletem aspectos da sociedade — a qual é marcada por desigualdades, exclusões e preconceitos —, tais características podem influenciar significativamente os resultados gerados pelos sistemas de IA (Mendonça Junior; Nunes, 2023).

O aprendizado de máquina pode reproduzir padrões discriminatórios caso esses estejam presentes nos dados utilizados para treiná-lo. Isso significa que, ao identificar tais padrões, o sistema tende a replicá-los com precisão. Um exemplo marcante dessa problemática é o caso do Tay, um robô de inteligência artificial criado pela Microsoft com a finalidade de interagir com usuários no Twitter. O funcionamento do sistema dependia do conteúdo enviado pelos próprios usuários da rede. Contudo, em poucas horas, o robô passou a publicar mensagens de cunho xenofóbico, racista e antissemita. (Bahia; Nunes; Pedron, 2021)

Porém, o caso mais contundente de viés algorítmico no Poder Judiciário foi o caso que ocorreu nos Estados Unidos, especificamente no estado de Wisconsin, em que um software privado chamado (Correctional Offender Management Profiling for Alternative Sanctions) foi utilizado para determinar a pena de um condenado em 2013. O algoritmo utilizado não é de conhecimento público, levando o réu a recorrer à Suprema Corte de Wisconsin a fim de obter acesso às diretrizes utilizadas pelo COMPAS, que, assumidamente, foram usadas pelo juiz do

caso e resultaram em sua condenação, com a conclusão de que ele tinha um alto risco de cometer violência, fugir e reincidir. Vale ressaltar que nem mesmo os juízes que utilizam o sistema têm acesso ao código-fonte do algoritmo. Neste caso, houve também questionamentos sobre se o sistema estava discriminando afro-americanos. A Suprema Corte de Wisconsin rejeitou o recurso, afirmando que a sentença foi baseada na análise do juiz em relação ao crime cometido e aos antecedentes do acusado. A Suprema Corte dos Estados Unidos também não aceitou o pedido de revisão (writ of certiorari) apresentado pelo condenado, que permanece detido.(Ferrari,2018)

Diante desse exemplo emblemático, é essencial reconhecer que os dados alimentam e moldam os sistemas de IA, sendo necessário estabelecer critérios mais rígidos de seleção, curadoria e validação dessas informações, desde modo, poderá ser possível reduzir os vieses e promover o desenvolvimento de tecnologias mais justas, éticas e representativas da diversidade social.

4.3 Impactos do viés na equidade, imparcialidade e confiança do Poder Judiciário

Os impactos do viés nos sistemas judiciais, especialmente quando incorporados por tecnologias como a inteligência artificial, representam uma séria ameaça à equidade, imparcialidade e confiança no Judiciário, deste modo, quando decisões são influenciadas por dados enviesados ou algoritmos mal construídos, há o risco de perpetuar desigualdades históricas e reforçar discriminações estruturais. A imparcialidade, princípio fundamental da Justiça, é comprometida, e a confiança social nas instituições judiciais enfraquece, assim, é essencial refletir sobre a ética no desenvolvimento e uso dessas tecnologias, garantindo transparência, revisão crítica e responsabilidade humana nas decisões judiciais.

Nesse diálogo, segundo Suleiman e Silva (2023), a IA tem o potencial de transformar profundamente o setor jurídico, especialmente em um país como o Brasil, onde há uma sobrecarga nos tribunais e demora significativa na tramitação de processos, por meio da análise de grandes volumes de dados jurídicos, como leis,

jurisprudência e precedentes, os sistemas de IA podem auxiliar juízes na fundamentação de decisões, tornando o sistema mais ágil e eficiente, todavia, esses avanços não estão livres de desafios, sendo necessário equilibrar o uso da tecnologia com a preservação dos direitos humanos e das garantias constitucionais.

Um dos principais desafios identificados no estudo de Suleiman e Silva (2023) é a questão do viés algorítmico, caso os dados utilizados para treinar os sistemas de IA sejam tendenciosos ou não representem adequadamente a diversidade da população brasileira, a tecnologia pode replicar ou até mesmo amplificar preconceitos e desigualdades já existentes, esse problema pode resultar em decisões discriminatórias ou injustas, exacerbando as desigualdades sociais, e a solução para esse desafio passa pela implementação de mecanismos de transparência e prestação de contas, além da divulgação dos algoritmos e dos dados de treinamento utilizados, permitindo auditorias e supervisão contínua.

Outro ponto destacado pelos autores é a ausência de empatia nas decisões baseadas em IA, enquanto os sistemas podem processar informações de maneira objetiva e eficiente, eles não possuem a capacidade de compreender a complexidade das experiências humanas e as circunstâncias individuais de cada caso, neste sentido, conforme destaca os autores, a empatia é uma qualidade inerente aos operadores do direito e desempenha um papel fundamental na análise de casos sensíveis, onde é necessário considerar aspectos subjetivos e emocionais. Isso reforça a importância da atuação humana no processo de tomada de decisões judiciais (Suleiman; Silva, 2023).

Além disso, cumpre delimitar que o papel dos magistrados é essencial, mas, indubitavelmente, enfrentam uma pressão crescente devido ao aumento do volume de processos, o que impacta a eficácia e a agilidade na entrega da justiça. A busca pela duração razoável do processo tornou-se uma preocupação tanto da sociedade quanto dos profissionais do direito, já que a atual estrutura do judiciário reflete políticas passadas que, por permitir um acesso excessivo ao sistema de justiça, não atendem mais adequadamente às necessidades da sociedade moderna (Aquino, 2023).

Logo, “Com vários avanços significativos em algoritmos de aprendizagem de máquinas e processamento de linguagem natural, a Inteligência Artificial (IA) demonstra um grande potencial capaz de aprimorar a prestação da jurisdição estatal” (Simões; Morais, 2024, p. 6).

As instituições jurídicas enfrentam um aumento gradual do congestionamento, o que dificulta a prestação adequada de serviços e, para melhorar a eficiência do sistema judiciário, é necessário adotar medidas para reduzir o volume de processos, combatendo o excesso de litigiosidade sem prejudicar os direitos das partes envolvidas, cuja "curva das ações processuais" precisa ser tratada para garantir a produtividade e a justiça no sistema (Aquino, 2023).

a utilização da Inteligência Artificial (IA) no Poder Judiciário é uma inovação tecnológica que visa desafogar o grande volume de processos em trâmite, e assim aprimorar a prestação jurisdicional e garantir a celeridade processual, contribuindo para a realização de uma justiça mais efetiva, equitativa e transparente não só para as partes, mas para a toda a sociedade (Franco Junior; Cardoso, 2024, p. 38).

Nesse passo, considerando ser inegável a ineficácia do sistema jurídico brasileiro, especialmente pelo alto custo de pequenas demandas e a morosidade gerada por múltiplos recursos, a inteligência artificial surge como solução para maior eficiência e acessibilidade na justiça. No entanto, sua implementação exige supervisão humana rigorosa para evitar decisões automatizadas que possam contrariar princípios éticos e morais, garantindo a integridade do sistema jurídico (Silva; Bernardes, 2023).

Considerando este contexto, é inegável que “a estrutura jurídica dos institutos não consegue cumprir com aquilo que a lei preceitua, pois tem que lidar com casos simples que poderiam ser tratados por outros caminhos, e também com demandas mais complexas que não conseguem ser analisadas com maior rigor” (Aquino, 2023, p. 41).

Tem-se que, apesar das controvérsias, a Inteligência Artificial (IA) visa objetivos simples, como facilitar e otimizar processos para melhorar a eficiência e a rentabilidade, tornando os serviços jurídicos mais acessíveis, com ferramentas como *chatbots* e assistentes virtuais, que oferecem informações legais básicas a qualquer momento, sem necessidade de agendamento (Simões; Morais, 2024).

Um aspecto crítico que emerge reside na confiança dos usuários finais no sistema judicial automatizado, que parece ser diretamente impactada pela falta de transparência nos processos algorítmicos, mostrando-se necessária a adoção de mecanismos que tornem essas decisões mais transparentes e auditáveis,

fortalecendo a confiança através de estratégias de IA explicável (XAI) (Aquino, 2023).

É indiscutível que a IA é capaz de transformar os processos judiciais, aumentando a acessibilidade e a transparência, mas, destaca-se, de igual forma, o perigo de os algoritmos perpetuarem vieses existentes, comprometendo a imparcialidade das decisões legais. Essa pesquisa lançou as bases para debates subsequentes acerca dos riscos associados à aplicação de IA no contexto legal (Aquino, 2023).

Sob esse prisma, tem-se que o uso da Inteligência Artificial (IA) pode impulsionar o acesso à justiça no Brasil, tornando as decisões judiciais mais previsíveis e controladas, o que igualmente contribui para o aumento da eficiência na administração de fóruns e tribunais, podendo acelerar processos judiciais ao automatizar tarefas repetitivas, como análise de evidências, revisão de documentos e triagem de casos, além de auxiliar juízes na análise de dados complexos e precedentes legais, fornecendo informações valiosas para fundamentar decisões (Simões; Moraes, 2024).

A aplicação da IA no meio jurídico busca aumentar a produtividade e otimizar o fluxo de trabalho dos tribunais, mas, resta evidente que o seu uso deve respeitar princípios éticos, garantindo equidade e imparcialidade. A decisão humana permanece essencial para preservar a reflexão e interpretação nos julgamentos, sendo certo que uma abordagem prudente é iniciar a implementação da IA em casos de menor complexidade, minimizando riscos e assegurando uma transição responsável e eficiente no sistema judicial, conforme Silva e Bernardes (2023, p. 31231):

O emprego da IA deve ser pautado por princípios que preservem os direitos individuais e coletivos, garantindo a equidade e a imparcialidade nas decisões. Ademais, é crucial reconhecer o papel crucial da decisão humana como instância final, preservando o caráter reflexivo e interpretativo inerente ao exercício do julgamento. Em um estágio inicial, priorizar casos de menor complexidade e valor, atenuando o potencial impacto, surge como uma abordagem prudente para introduzir a IA de forma responsável e eficiente no âmbito judicial.

A delimitação de tecnologias no Judiciário tem promovido melhorias na eficiência dos processos, destacando-se pela automação e otimização do trabalho,

resultando em maior celeridade nos processos judiciais, redução do tempo de tramitação de casos e aumento da produtividade dos tribunais, além de aliviar a sobrecarga de trabalho dos magistrados e servidores, evidenciado em tribunais que adotaram sistemas automatizados (Schmidt et al , 2025).

“A busca por maior eficiência e celeridade processual tem levado à adoção de algoritmos e sistemas baseados em Inteligência Artificial – IA, para auxiliar juízes na administração do acervo e julgamento dos processos” (Franco Junior; Cardoso, 2024, p. 40).

Em que pese a implementação da Inteligência Artificial (IA) no sistema judiciário garantir maior eficiência e celeridade nas decisões judiciais, traz desafios como equidade, ética e a necessidade de supervisão adequada, de forma que a coleta e o armazenamento de grandes quantidades de dados pessoais exigem medidas rigorosas de segurança e privacidade para proteger informações sigilosas e reduzir riscos (Simões; Morais, 2024).

Os impactos dos vieses algorítmicos no âmbito do Poder Judiciário, em particular, os algoritmos examinados demonstram uma tendência a refletir desigualdades socioeconômicas já existentes, afetando a percepção de equidade e imparcialidade das decisões judiciais (Barocas; Selbst, 2016).

Percebe-se que, embora a tecnologia tenha avançado, o julgamento humano é insubstituível, pois envolve análise contextual, empatia e avaliação ética e, diante disso, a Inteligência Artificial (IA) deve ser uma ferramenta de apoio, não uma substituição, não podendo se olvidar que os algoritmos de IA podem reproduzir preconceitos presentes nos dados de treinamento, o que pode resultar em decisões discriminatórias. Para garantir justiça e igualdade, é essencial monitorar continuamente os vieses algorítmicos (Simões; Morais, 2024).

A implementação da IA no Poder Judiciário levanta preocupações éticas importantes. A transparência dos algoritmos, a equidade no tratamento de diferentes casos e a possibilidade de viés algorítmico são temas debatidos. A falta de diversidade nos conjuntos de dados utilizados para treinar os modelos pode resultar em discriminação, reproduzindo preconceitos existentes na sociedade (Franco Junior; Cardoso, 2024, p. 48).

A adoção da Inteligência Artificial (IA) no judiciário é capaz de ofertar maior eficiência, mas também levanta questões sobre privacidade, ética, discriminação e transparência e, apesar da tecnologia esteja em desenvolvimento, é

indispensável estabelecer regulamentações rigorosas para proteger os direitos individuais, não apenas no judiciário, mas em todas as áreas de aplicação da IA, garantindo-se, assim, segurança e evitando abusos (Simões; Morais, 2024).

Este entendimento permanece sendo partilhado, argumentando-se que embora a adoção de tecnologias no judiciário possa ensejar o aumento da eficiência jurisdicional, apresenta desafios relacionados à segurança de dados e à privacidade, uma vez que a digitalização crescente pode resultar em falhas, como vazamentos de dados. Além a isso, a automação dos processos gera resistência entre profissionais do direito, que temem comprometer o caráter humano da justiça e a personalização das decisões, além de enfrentarem dificuldades na adaptação aos novos sistemas (Schmidt et al, 2025).

“Há necessidade de destacar a importância de um bom gerenciamento de dados, baseados nas premissas da fidedignidade e da integralidade, a fim de que os algoritmos possam refletir o uso adequado, ético e benéfico da Inteligência Artificial” (Franco Junior; Cardoso, 2024, p. 49).

Apesar da inteligência artificial (IA) trazer benefícios na otimização de processos, ela não deve ser usada de forma autônoma em julgamentos judiciais, pois pode violar princípios constitucionais como a ampla defesa e o contraditório, razão pela qual a decisão final deve sempre ser responsabilidade de um magistrado, já que a IA pode apresentar vieses que comprometem a imparcialidade do julgamento. Aliado a isso, não se pode deixar de mencionar que a automação excessiva pode prejudicar os direitos fundamentais dos cidadãos, pois cada caso requer uma análise detalhada, considerando as particularidades e os princípios democráticos (Aquino, 2023).

Nessa perspectiva, “há uma preocupação de que a justiça ‘dita dos homens’ não se torne uma ‘justiça artificial’ a ponto de distorcer a principal função do Poder Judiciário, que é garantir os direitos individuais, coletivos e sociais e resolver conflitos entre cidadãos, entidades e Estado” (Franco Junior; Cardoso, 2024, p. 52).

O estabelecimento da IA no sistema judiciário deve respeitar princípios democráticos e diretrizes éticas, como supervisão humana, transparência e responsabilização e, sendo assim, o julgamento deve considerar as particularidades de cada caso, não sendo uma resposta padronizada. Embora a IA possa melhorar a eficiência, ela deve ser uma ferramenta de apoio, não substituindo o discernimento

humano, comportando como prioridade garantir a proteção dos direitos dos cidadãos e a integridade do processo judicial, consoante Aquino (2023, p. 43):

Assim, enquanto a IA tem potencial para melhorar a eficiência dos tribunais, seu uso não deve objetivar apenas a economia de recursos. A principal prioridade deve ser garantir que os direitos dos cidadãos e a integridade do processo judicial sejam mantidos. Em resumo, a inteligência artificial deve servir como uma ferramenta de apoio, não como um substituto para o discernimento humano no processo judicial.

A inteligência artificial deve ser entendida como uma ferramenta de apoio, especialmente útil para lidar com demandas repetitivas, previsíveis e baseadas em padrões já consolidados. Em atividades como análise de súmulas vinculantes, precedentes repetitivos e teses já firmadas, a IA pode oferecer agilidade, consistência e eficiência. Nesses contextos, a máquina consegue identificar rapidamente decisões anteriores semelhantes, cruzar dados e apontar caminhos com base no que já foi decidido, otimizando o tempo dos operadores do Direito e permitindo que eles foquem nas questões que realmente exigem interpretação humana.

No entanto, é fundamental lembrar que a IA só aprende a partir de dados passados — ou seja, de informações que já aconteceram. Isso significa que, quando nos deparamos com situações novas, complexas ou que exigem criatividade e sensibilidade, a máquina não tem como oferecer respostas inovadoras. Problemas inéditos, nuances específicas e decisões que exigem empatia ou consideração de contextos sociais e históricos fogem totalmente da capacidade da IA. Nessas horas, só o ser humano consegue romper com o padrão e propor soluções que dialoguem com os princípios de justiça e com as transformações da sociedade.

O melhor caminho não é pensar em substituir o humano pela máquina, nem o contrário. A solução está na construção de decisões híbridas, nas quais a IA atua como suporte técnico para tarefas repetitivas, enquanto o ser humano permanece no centro das decisões que exigem interpretação, ética e criatividade. Em um mundo cada vez mais dinâmico, com desafios jurídicos em constante mutação, a atuação humana continua sendo insubstituível — e é justamente essa parceria equilibrada que pode tornar o sistema de justiça mais eficiente, justo e inovador.

Percebe-se que a adoção crescente da inteligência artificial (IA) no processo brasileiro pode trazer melhorias para a eficiência e a celeridade processual, ajudando a reduzir a sobrecarga dos tribunais. Contudo, é fundamental que a IA respeite os direitos fundamentais, como o devido processo legal e a ampla defesa, devendo ser guiado por limites éticos e regulamentares, assegurando supervisão humana para evitar discriminação e garantir transparência nas decisões, sem substituir a análise crítica humana (Schmidt et al, 2025).

Em que pese os entendimentos mencionados, observa-se o quão complexo é estabelecer *accountability* para os respectivos algoritmos decisórios, levando-se em consideração o fato de que a opacidade dessas tecnologias pode impedir a identificação de erros ou vieses e dificultar a realização de auditorias eficazes, levantando, conseqüentemente, questões fundamentais sobre como as práticas regulatórias devem evoluir para acompanhar o desenvolvimento tecnológico e garantir que os sistemas de IA não comprometam a justiça (Aquino, 2023).

Isso porque a manutenção de injustiças estruturais mediadas por algoritmos é uma preocupação legítima que se materializa nos observáveis desafios de equidade e imparcialidade, corroborando o fato de que as decisões automatizadas podem frequentemente refletir preconceitos subjacentes ao invés de mitigá-los (Barocas; Selbst, 2016).

É indispensável garantir que as inovações tecnológicas no judiciário respeitem princípios como transparência e imparcialidade, além de assegurar a capacitação dos profissionais para o uso ético das novas ferramentas, de maneira que a modernização sirva de fortalecimento para a justiça, não apenas como meio de acelerar processos, mas, de igual forma, para manter a equidade e a segurança jurídica. Assim, o uso responsável das tecnologias pode tornar o sistema mais acessível, confiável e eficiente, sem comprometer os direitos individuais (Schmidt et al, 2025).

Levando-se em consideração o exposto, tem-se que estas vieses são capazes de comprometer sobremaneira princípios elencados, minando a confiança pública nas decisões judiciais e, sendo assim, a utilização dos vieses algoritmos, sem as devidas salvaguardas, tendem a reproduzir preconceitos preexistentes, sugerindo a necessidade de intervenções regulatórias e técnicas para mitigar esses riscos (Barocas; Selbst, 2016).

4.4 Mecanismos de Mitigação e Regulação

No contexto dos vieses algorítmicos, torna-se cada vez mais urgente e necessário desenvolver mecanismos eficazes de mitigação e regulação que garantam a equidade, a transparência e o respeito aos direitos fundamentais, assim, à medida que os sistemas de inteligência artificial ganham espaço em decisões que afetam diretamente a vida das pessoas – seja na área da justiça, saúde, segurança, crédito, ou consumo –, cresce também a responsabilidade ética, técnica e legal sobre a forma como esses algoritmos são desenvolvidos, treinados, utilizados e supervisionados.

Um dos principais elementos para a mitigação de riscos é a criação de políticas de governança que integrem aspectos éticos e técnicos no desenvolvimento e na implementação de sistemas de IA, a governança deve incluir a definição de padrões para a explicabilidade dos algoritmos, permitindo que os cidadãos compreendam como as decisões que os afetam são tomadas. Isso implica em desenvolver mecanismos que garantam a auditabilidade dos sistemas de IA, permitindo a verificação de suas operações e a identificação de possíveis falhas ou vieses, a promoção de uma cultura de transparência é essencial para que os sistemas de IA sejam aceitos pela sociedade, uma vez que a confiança no uso dessas tecnologias é fundamental para sua eficácia.

A regulação deve ser adaptativa e dinâmica, acompanhando o avanço rápido das tecnologias de IA, envolvendo assim, a necessidade de legislação que não apenas estabeleça normas claras, mas que também permita flexibilidade para se ajustar a novas realidades tecnológicas.

A implementação de auditorias regulares e independentes é outra estratégia fundamental para garantir a conformidade dos sistemas de IA com as normas estabelecidas, essas auditorias não apenas identificam problemas, mas também promovem a melhoria contínua dos processos, neste sentido, a criação de mecanismos de ética que atuem como supervisores das práticas de IA nas instituições públicas pode ser uma forma eficaz de assegurar que os princípios de justiça e equidade sejam respeitados na utilização dessas tecnologias, esses

mecanismos devem incluir representantes diversos, garantindo que diferentes perspectivas e experiências sejam consideradas nas decisões relacionadas à implementação de IA.

A educação e a capacitação dos profissionais envolvidos na implementação e supervisão de sistemas de IA são essenciais para o sucesso da regulação e mitigação de riscos. Investir em formação continuada que aborde tanto os aspectos técnicos quanto os éticos do uso de IA permitirá que os profissionais estejam melhor preparados para lidar com os desafios e lacunas que surgem.

Destaca-se que ascensão dos algoritmos como componentes centrais das sociedades contemporâneas levanta questões cruciais sobre o poder que exercem nas relações sociais e políticas, Mendonça, Filgueiras e Almeida (2025) argumentam que os algoritmos não são meras ferramentas tecnológicas, mas instituições que definem e organizam a vida cotidiana, impactando decisões em diversas áreas, desde a seleção de candidatos em processos de recrutamento até a avaliação de risco em sistemas de justiça, como artefatos sociotécnicos, os algoritmos são moldados por contextos socio-históricos e pelas intenções de quem os projeta, mas suas consequências vão além das intenções originais, gerando impactos profundos nas interações humanas.

Neste cenário, um dos aspectos mais preocupantes do uso de algoritmos é a forma como eles categorizam e definem indivíduos, reforçando estereótipos e preconceitos existentes, e o uso de dados históricos para treinar modelos de IA pode resultar em discriminação sistemática, como demonstrado por Obermeyer et al. (2019), que evidenciaram como algoritmos de saúde podem subestimar as necessidades de tratamento de populações racializadas, essa reproduzida desigualdade revela a necessidade urgente de uma governança mais rigorosa e de políticas de curadoria de dados que assegurem a representatividade e a equidade, e a falta de atenção a essas práticas pode resultar em algoritmos que não apenas falham em fornecer resultados justos, mas que também perpetuam injustiças sociais.

As formas de resistência ao poder algorítmico são variadas e frequentemente sutis, Mendonça, Filgueiras e Almeida (2025) introduzem o conceito de "desidentificação", um mecanismo que permite aos indivíduos se afastarem das definições rígidas impostas pelos algoritmos, essa desidentificação pode ser vista como uma expressão da infrapolítica, conceito desenvolvido por James Scott, que se refere a formas de resistência menos visíveis e organizadas, tais ações incluem a

manipulação de dados e comportamentos de forma a confundir os algoritmos, desafiando assim a categorização e a vigilância. Isso pode ser observado em práticas cotidianas em que os indivíduos jogam com os sistemas algorítmicos, utilizando táticas criativas para contornar as determinações algorítmicas.

Entre as estratégias de resistência, a anonimização e o engano se destacam como táticas centrais, e a anonimização envolve ocultar a identidade em contextos digitais, permitindo que indivíduos ajam fora do alcance dos sistemas de vigilância. Isso pode ser alcançado através de navegação em modo anônimo ou pelo uso da dark web, onde a identificação é dificultada, e o engano, por sua vez, refere-se a ações deliberadas que visam confundir os algoritmos sobre interesses e preferências pessoais, como demonstrado por práticas onde os usuários ajustam suas interações para manipular os resultados que os algoritmos produzem, essas táticas representam formas de resistência cotidianas que, embora possam parecer pequenas, têm o potencial de influenciar as dinâmicas de poder estabelecidas (Mendonça; Filgueiras; Almeida, 2025).

Nesta perspectiva, a análise das interações entre humanos e algoritmos revela a necessidade de um entendimento mais profundo sobre como esses sistemas moldam identidades e comportamentos, e a resistência ao poder algorítmico, conforme discutido por Mendonça, Filgueiras e Almeida (2025), não é apenas uma luta contra a vigilância, mas uma busca por formas alternativas de ser e agir em uma sociedade cada vez mais mediada por tecnologia, as práticas de desidentificação e as táticas de resistência são fundamentais para desafiar as normas e regras impostas pelos algoritmos, permitindo que os indivíduos reivindiquem sua agência em um ambiente que frequentemente tenta reduzi-los a meros pontos de dados.

Um dos primeiros passos nesse processo é a implementação de políticas de governança e curadoria de dados, pois os vieses geralmente têm origem no próprio conjunto de dados utilizado para treinar os modelos de IA, esses dados muitas vezes refletem desigualdades sociais, discriminações históricas e padrões comportamentais enviesados que, ao serem reproduzidos por máquinas, ganham aparência de neutralidade e objetividade.

O avanço da inteligência artificial tem transformado diversos setores, desde a saúde até o sistema judiciário, contudo, um dos maiores desafios enfrentados pelas organizações que adotam soluções baseadas em IA é assegurar

que os modelos sejam éticos, justos e imparciais, assim, grande parte dos vieses encontrados nesses sistemas tem origem no conjunto de dados utilizado para treiná-los, segundo Barocas, Hardt e Narayanan (2019), os dados, muitas vezes, refletem padrões históricos de discriminação e desigualdade presentes na sociedade, que acabam sendo replicados ou amplificados pelos algoritmos, e para abordar esse problema, a implementação de políticas robustas de governança e curadoria de dados torna-se essencial, com o objetivo de garantir a qualidade, representatividade e transparência dos conjuntos de dados utilizados no treinamento de modelos de IA.

A governança de dados consiste em um conjunto de práticas e diretrizes que regulam a coleta, armazenamento, uso e gerenciamento de dados dentro de uma organização, em um contexto de IA, uma governança eficaz deve garantir que os dados utilizados sejam confiáveis e representem adequadamente a diversidade da população. Mitchell et al. (2021) argumentam que a falta de governança pode levar à utilização de dados enviesados, resultando em decisões automatizadas que prejudicam grupos já marginalizados, e a ausência de políticas claras pode permitir que dados incompletos ou distorcidos sejam usados para treinar os modelos, comprometendo a imparcialidade dos resultados produzidos pelos sistemas de IA, dessa forma, a governança deve incluir mecanismos para identificar e corrigir vieses, bem como para auditar continuamente os conjuntos de dados utilizados.

Outro ponto fundamental é a curadoria de dados, que envolve a seleção, limpeza e organização dos dados utilizados para o treinamento de modelos de IA, de acordo com Obermeyer et al. (2019), práticas inadequadas de curadoria podem introduzir distorções que exacerbam desigualdades sociais, esses autores destacam que estudos na área da saúde revelaram que sistemas de IA treinados com dados baseados em custos médicos tendem a subestimar as necessidades de pacientes de baixa renda, pois esses indivíduos têm menor acesso a serviços de saúde e, conseqüentemente, apresentam custos menores nos registros hospitalares, esse caso ilustra como a curadoria de dados deve ser conduzida com atenção aos contextos sociais e históricos, a fim de evitar que os modelos reforcem desigualdades estruturais.

É importante que as políticas de governança e curadoria sejam combinadas com a implementação de práticas de transparência e prestação de contas, Barocas, Hardt e Narayanan (2019) destacam que a transparência é um dos pilares para a construção de confiança em sistemas de IA, os usuários e as partes

interessadas devem ter acesso a informações sobre como os dados foram coletados, tratados e utilizados no treinamento dos modelos, dessa forma, ferramentas como explicabilidade dos algoritmos e relatórios detalhados sobre o impacto dos dados nos resultados finais podem ajudar a identificar possíveis problemas e a promover a responsabilização das organizações que desenvolvem ou utilizam soluções de IA, dessa forma, a transparência não apenas fortalece a confiança nos sistemas, mas também permite que os vieses sejam identificados e corrigidos com maior rapidez.

Nessa senda, a colaboração interdisciplinar entre cientistas de dados, especialistas em ética e representantes de grupos impactados também é essencial para a implementação de políticas eficazes de governança e curadoria de dados, segundo Mitchell et al. (2021), equipes diversas são mais capazes de identificar e abordar potenciais fontes de viés nos dados, uma vez que trazem diferentes perspectivas e experiências para o processo de desenvolvimento dos modelos, e a criação de diretrizes éticas e a realização de auditorias regulares devem ser conduzidas de forma participativa, garantindo que os sistemas de IA sejam projetados para beneficiar todas as populações de maneira justa e inclusiva.

A implementação de políticas de governança e curadoria de dados deve ser acompanhada de iniciativas de capacitação e conscientização. Obermeyer et al. (2019) enfatizam que os tomadores de decisão e os usuários finais precisam estar cientes dos riscos associados ao uso de dados enviesados e das medidas necessárias para mitigá-los, assim incluindo treinamentos sobre como avaliar a qualidade dos dados, identificar possíveis fontes de viés e interpretar os resultados dos modelos de IA de maneira crítica, e a combinação de governança, curadoria, transparência e capacitação é essencial para garantir que as soluções baseadas em inteligência artificial sejam éticas, eficazes e alinhadas aos princípios de justiça e igualdade.

A implementação de políticas de governança e curadoria de dados é indispensável para mitigar os vieses nos sistemas de IA, essas políticas devem garantir a qualidade e a representatividade dos dados, promover a transparência e facilitar a prestação de contas, além de envolver equipes diversas e capacitadas no processo de desenvolvimento dos modelos, destarte, ao adotar essas práticas, as organizações podem reduzir os riscos associados aos vieses algorítmicos e assegurar que a inteligência artificial seja utilizada de forma ética e benéfica

A curadoria responsável dos dados implica em selecionar, classificar e tratar os dados com atenção crítica, garantindo diversidade, representatividade e qualidade da informação, já a governança de dados diz respeito ao estabelecimento de normas, responsabilidades e fluxos organizacionais que assegurem uma gestão ética e eficiente dos dados ao longo de todo o ciclo de vida da tecnologia, desde sua concepção até seu uso prático.

A auditoria e o monitoramento contínuo dos algoritmos são estratégias fundamentais para identificar e corrigir vieses durante a aplicação dos sistemas de IA, a auditoria algorítmica deve ser realizada tanto por especialistas internos quanto por agentes independentes, garantindo a análise crítica de como os algoritmos operam, com quais dados trabalham, quais critérios utilizam para tomar decisões e quais impactos sociais estão gerando.

Desse modo, as auditorias devem observar se o uso da inteligência artificial no Judiciário deve estar alinhado à Constituição e a tratados internacionais, em estrita proteção aos direitos fundamentais, além de proibir a utilização de dados sigilosos ou protegidos por segredo de justiça sem anonimização prévia, determinando que as empresas fornecedoras cumpram a legislação brasileira, especialmente a LGPD, adotando medidas rigorosas de segurança da informação (Travain, 2025).

O monitoramento regular permite detectar desvios de comportamento nos sistemas, verificar se estão mantendo o desempenho esperado e evitar que reproduzam ou agravem desigualdades sociais, essa prática também envolve o uso de métricas de justiça algorítmica, como paridade demográfica ou igualdade de oportunidades, para verificar se diferentes grupos populacionais estão sendo tratados de maneira justa e equitativa.

Além disso, de acordo com um estudo proposto por Abeba Birhane, observa-se que o aumento do volume de dados não garante a melhoria dos sistemas de IA, podendo até ampliar conteúdos odiosos contra minorias e, em decorrência disso, são necessários mecanismos rigorosos de avaliação, curadoria e auditoria, tendo em vista que mesmo os sistemas considerados de baixo risco, como os usados para sumarização e análise de jurisprudência, exigem cautela, pois podem cristalizar injustiças, razão pela qual é essencial revisar periodicamente sua classificação para verificar se seus impactos não os tornam, na prática, sistemas de alto risco, conforme Rodrigues e Lopes (2024, p. 1):

Mesmo no uso de tecnologias de IA classificadas como de baixo risco, é preciso atenção. A adoção de sistemas para sumarização e análise de jurisprudência, como a identificação de padrões decisórios, não pode resultar no engessamento da justiça – ou pior, no engessamento de injustiças, de modo que a utilização desse tipo de ferramenta precisa ser acompanhada de uma atuação que se preocupe em promover também as atualizações necessárias ao Direito com o passar do tempo. Futuramente, é relevante a revisão da classificação desses sistemas, a fim de identificar se os seus efeitos não justificam a sua classificação, na verdade, como um sistema de alto risco.

Dentro deste panorama, a aplicação de uma análise estatística combinada com análise de conteúdo é capaz de possibilitar que os dados sejam transformados em insights acionáveis, fornecendo uma base sólida para recomendações tanto para políticas de governança e curadoria de dados quanto para estratégias regulamentares, o que assegura que as conclusões do estudo sejam não apenas teoricamente fundadas, mas também relevantes em um contexto prático (Bardin, 2011).

A presença de vieses algorítmicos no judiciário não é apenas uma falha técnica, mas uma manifestação de problemas mais amplos relacionados à desigualdade estrutural e à falta de transparência. Dentro deste enfoque, acaba se tornando indispensável que as futuras intervenções tecnológicas sejam acompanhadas por estratégias de auditoria regular e monitoramento para assegurar que os sistemas permaneçam responsivos e justos (Cowgill et al., 2020).

A transparência no uso da inteligência artificial (IA) no Judiciário exige debate público, definição de parâmetros claros e auditorias contínuas, visando elidir a presença de modelos de IA opacos, conhecidos como "caixas-pretas", dificultando o acesso a informações fundamentais. No contexto judicial, não se sabe quais dados são fornecidos à IA, nem o funcionamento de seus códigos ou a organização dos dados, o que acaba comprometendo o direito à ampla defesa e ao contraditório, cuja falta de transparência fere o princípio da publicidade, essencial ao Estado Democrático de Direito, conforme Toledo e Pessoa (2023, p. 1):

Outros aspectos da transparência são a necessidade de amplo debate público sobre a utilização da IA para a decisão judicial, seus parâmetros e a

realização contínua e permanente de auditorias sobre a IA. Entretanto, no mercado das IA, prevalecem os modelos opacos - “caixas pretas inescrutáveis” -, de forma que os modelos transparentes são raras exceções. No âmbito do Poder Judiciário, não se conhecem (i) os dados fornecidos à IA, (ii) o código da IA utilizada na decisão judicial, e (iii) como são organizados os dados destinados à máquina, então se desconhecem as informações envolvidas no processo judicial, o que afeta diretamente tanto a ampla defesa quanto o contraditório, pois as partes não sabem do que se defender, nem o que contraditar. A falta de transparência no uso de IA pelo Judiciário configura evidente afronta ao princípio da publicidade, orientador do Estado Democrático de Direito.

Nesse cenário, ganha destaque a necessidade de desenvolver sistemas de inteligência artificial explicável, conhecidos pela sigla XAI (*eXplainable Artificial Intelligence*), cuja proposta é tornar as decisões dos algoritmos compreensíveis para os usuários, desenvolvedores, reguladores e demais interessados.

A crescente adoção de sistemas de inteligência artificial em diversos contextos, como saúde, justiça e segurança, trouxe à tona a necessidade de tornar as decisões algorítmicas mais transparentes e compreensíveis para os usuários, e desse modo, essa demanda está diretamente relacionada ao conceito de *eXplainable Artificial Intelligence* (XAI), ou inteligência artificial explicável, que visa permitir que humanos compreendam e confiem nas decisões tomadas por sistemas de IA, e a explicabilidade é fundamental não apenas para promover a confiança nos sistemas automatizados, mas também para assegurar que suas decisões estejam alinhadas a valores éticos e aos direitos fundamentais.

Em apertada síntese, a inteligência artificial explicável (XAI) compreende métodos que tornam os resultados dos algoritmos de aprendizado de máquina compreensíveis e confiáveis para os usuários, capaz de descrever o funcionamento do modelo, seus impactos e possíveis vieses, promovendo transparência, justiça e precisão, sendo indispensável para a construção da confiança no uso da IA, adotando uma abordagem responsável no seu desenvolvimento e aplicação, de acordo com o IBM (2020, p. 1):

A inteligência artificial explicável (XAI) é um conjunto de processos e métodos que permite aos usuários humanos compreender e confiar nos resultados e na saída criados por algoritmos de **aprendizado de máquina**. A IA explicável é usada para descrever um modelo de IA, seu impacto esperado e possíveis vieses. Isso ajuda a caracterizar a precisão do

modelo, a justiça, a transparência e os resultados na tomada de decisões com tecnologia de IA. A IA explicável é crucial para uma organização na construção de confiança ao colocar modelos de IA em produção. A explicabilidade da IA também ajuda a uma organização a adotar uma abordagem responsável para o desenvolvimento da IA.

A explicabilidade da IA (XAI) refere-se ao uso de métodos e ferramentas que tornam os modelos de machine learning compreensíveis aos humanos, permitindo entender como as decisões são tomadas e, mais precisamente no contexto jurídico brasileiro, essa explicabilidade é essencial, pois as decisões judiciais devem ser fundamentadas de maneira inteligível, conforme exigências constitucionais e legais, mesmo quando auxiliadas por sistemas de IA (Pecego; Teixeira, 2024).

Observe que o artigo 93, IX, da Constituição Federal, exige que todas as decisões do Judiciário sejam devidamente fundamentadas, sob pena de nulidade, ao passo que o Código de Processo Civil reforça essa exigência nos artigos 10, 11 e 489, estabelecendo que o juiz deve permitir manifestação das partes sobre os fundamentos utilizados, inclusive os gerados por inteligência artificial. Assim, a simples adoção das conclusões da IA não é suficiente, eis que o magistrado deve justificar, de forma clara e compreensível, por que tais recomendações são pertinentes ao caso concreto, de acordo com Pecego e Teixeira (2024, p. 10):

O artigo 93 da Constituição Federal Brasileira (CRFB), dispõe expressamente no inciso IX que todas as decisões dos órgãos do Poder Judiciário devem ser fundamentadas, sob pena de nulidade. Desta forma, a simples adoção das conclusões fornecidas pela IA sem uma explicação detalhada não é suficiente, devendo o juiz fundamentar o porquê de as recomendações da IA serem pertinentes e aplicáveis ao caso concreto submetido à sua análise.

Isso porque decisões judiciais que apenas reproduzem normas usam conceitos genéricos (como “razoabilidade” ou “boa-fé”), ou citam precedentes sem contextualização, não são consideradas devidamente fundamentadas, sendo certa a necessidade de ser explicada de forma clara e específica a pertinência dos argumentos ao caso concreto, inclusive quando envolver o uso de IA. Além disso, o juiz deve enfrentar todos os argumentos relevantes trazidos pelas partes e justificar eventuais divergências de jurisprudência, garantindo uma decisão íntegra e compreensível (Pecego; Teixeira, 2024).

Veja-se que a explicabilidade da IA (XAI) está diretamente relacionada ao

princípio do devido processo legal, sendo essencial para garantir a contestação e revisão das decisões judiciais, cujo princípio, previsto no artigo 5º, inciso LV da Constituição, assegura o contraditório e a ampla defesa, os quais só podem ser plenamente exercidos se as partes compreendem claramente as razões que fundamentam a decisão do magistrado (Pecego; Teixeira, 2024).

O surgimento da IA explicável (XAI) como solução viável para aumentar a transparência dos processos algorítmicos consiste em uma inovação relevante para o fim de construir confiança entre usuários do sistema judicial e o público em geral, permitindo que os mecanismos de decisão sejam mais compreensíveis e, portanto, mais auditáveis por humanos, o que também pode contribuir na identificação de vieses persistentes que requerem intervenção (Zarsky, 2016).

A viabilidade da IA explicável (XAI) como uma solução de curto e médio prazo para aumentar a transparência dos processos algorítmicos é um achado importante, uma vez que esta abordagem pode auxiliar na construção da confiança das tecnologias emergentes, promovendo uma adesão mais ampla às práticas digitais na justiça, o que da mesma forma permite a identificação precoce de vieses, assegurando uma adaptabilidade contínua ao evoluir do panorama tecnológico e social (Cowgill et al., 2020).

A dificuldade em compreender como os algoritmos tomam decisões, especialmente em modelos mais complexos como deep learning, é uma das principais barreiras para a aceitação e confiança do público nessas tecnologias, e de acordo com Miller (2019), a explicabilidade não é apenas um requisito técnico, mas também um compromisso ético e social, o autor argumenta que sistemas opacos, frequentemente chamados de "caixas-pretas", podem tomar decisões que afetam diretamente a vida das pessoas sem fornecer justificativas claras ou compreensíveis, isso é particularmente problemático em áreas sensíveis, como decisões judiciais ou diagnósticos médicos, onde as consequências de uma decisão errada podem ser graves, deste modo, a XAI, nesse contexto, aparece como uma solução para mitigar os riscos associados à opacidade, ao permitir que especialistas e usuários finais entendam as razões por trás das escolhas algorítmicas.

Outro aspecto fundamental da XAI é a sua capacidade de promover a accountability (prestação de contas) e a supervisão humana em sistemas automatizados, conforme Ribeiro, Singh e Guestrin (2016), a explicabilidade é essencial para identificar e corrigir vieses nos algoritmos, além de facilitar auditorias

e garantir que as decisões estejam em conformidade com padrões éticos e legais, os autores desenvolveram o método LIME (Local Interpretable Model-Agnostic Explanations), que busca explicar as previsões de qualquer modelo de IA, mesmo os mais complexos, ferramentas como essa são indispensáveis para garantir que os sistemas de IA não apenas sejam eficientes, mas também justos e imparciais, de modo que, a explicabilidade permite que os usuários contestem decisões automatizadas, protegendo-os contra possíveis injustiças ou discriminações.

Neste caminho, a integração da XAI em sistemas de IA também desempenha um papel estratégico para aumentar a confiança do público e das instituições nessas tecnologias, conforme Gunning e Aha (2019), a falta de transparência é uma das razões pelas quais muitas organizações hesitam em adotar soluções baseadas em IA, especialmente em contextos regulamentados, os autores destacam que a XAI é fundamental para garantir que as decisões automatizadas sejam compreensíveis para uma ampla gama de stakeholders, incluindo usuários finais, reguladores e desenvolvedores, outrossim, a implementação de sistemas explicáveis contribui para o fortalecimento da colaboração entre humanos e máquinas, ao permitir que os usuários entendam e confiem nos processos automatizados, em vez de apenas aceitá-los passivamente.

A importância da XAI transcende o aspecto técnico, envolvendo debates éticos e regulatórios, e a explicabilidade é um dos pilares para o desenvolvimento de uma IA responsável, que respeite os direitos fundamentais e promova a inclusão, Miller (2019) ressalta que, ao tornar os sistemas mais transparentes, a XAI ajuda a mitigar riscos de discriminação algorítmica e garante que as tecnologias sejam projetadas para beneficiar a sociedade como um todo, todavia, Gunning e Aha (2019) alertam que o desenvolvimento de sistemas explicáveis não deve comprometer a precisão ou a eficiência dos modelos de IA, sendo necessário encontrar um equilíbrio entre a complexidade técnica e a clareza para os usuários.

A criação de sistemas de inteligência artificial explicável (XAI) é uma necessidade urgente em um cenário de crescente dependência de tecnologias algorítmicas, e a explicabilidade não apenas aumenta a confiança e a aceitação pública, mas também desempenha um papel crucial na mitigação de riscos, na promoção da justiça e na proteção de direitos fundamentais, dessarte, ferramentas como LIME e avanços em pesquisa mostram que a XAI é viável e pode ser integrada a soluções de IA sem comprometer sua eficácia, entretanto, para que isso

se torne uma prática padrão, é indispensável que desenvolvedores, reguladores e a sociedade civil trabalhem juntos para criar diretrizes e políticas que incentivem a adoção de sistemas mais transparentes e explicáveis.

Neste diálogo, o uso da Inteligência Artificial para apoiar decisões na esfera pública, como na administração pública e no judiciário, levanta questões importantes sobre a necessidade de explicabilidade dessas tecnologias, de acordo com Pádua e Lorenzetto (2024), a explicabilidade se torna um direito fundamental no contexto das decisões estatais, visando garantir que os cidadãos possam compreender e questionar as decisões automatizadas que os afetam, essa preocupação é especialmente relevante em um cenário onde a transparência e a accountability são essenciais para a construção da confiança pública nas instituições governamentais.

Um dos aspectos centrais discutidos pelos autores é a interseção entre a teoria da justificação das decisões estatais e as técnicas de IA explicável (xAI), a justificativa para essa interseção reside na necessidade de assegurar que as decisões tomadas com o auxílio de algoritmos estejam em conformidade com princípios éticos e normativos, assim, a pesquisa desses autores revela que a explicabilidade não deve ser vista como um obstáculo ao desenvolvimento da IA, mas como uma necessidade para fortalecer a integridade das decisões públicas, dessarte, a construção do direito à explicabilidade deve se fundamentar em diretrizes que considerem as especificidades da tecnologia e as exigências legais e éticas (Pádua; Lorenzetto, 2024).

A explicabilidade pode ser dividida em dois conceitos: a interpretabilidade, que se refere à compreensão dos processos computacionais, e a explicabilidade em sentido estrito, que busca traduzir esses processos em termos compreensíveis para o público em geral, essa distinção é elementar para a formulação de políticas públicas que integrem a IA de forma responsável e ética, assegurando que as decisões automatizadas sejam auditáveis e passíveis de contestação, a necessidade de uma abordagem clara e acessível à explicabilidade é reforçada por diretrizes internacionais, como as recomendações da OCDE e da UNESCO, que enfatizam a importância de sistemas de IA que sejam transparentes e responsivos às necessidades dos cidadãos (Pádua; Lorenzetto, 2024).

Os autores também abordam a importância de estabelecer marcos regulatórios que garantam a proteção do direito à explicabilidade, a legislação

existente, como a Lei Geral de Proteção de Dados (LGPD) no Brasil, já estabelece princípios que favorecem a transparência no tratamento de dados pessoais, mas é necessário que esses princípios sejam aplicados de maneira eficaz às decisões baseadas em IA, a accountability deve ser um componente central na governança dos sistemas de IA, permitindo que os cidadãos possam questionar e entender as decisões que impactam suas vidas, nesse sentido, os princípios de responsabilidade e transparência devem ser incorporados no design e na implementação da IA, assegurando que as instituições cumpram seu papel de prestação de contas (Pádua; Lorenzetto, 2024).

Deste modo, na visão desses autores, o direito à explicabilidade da IA nas decisões estatais é um reflexo da necessidade de proteger a dignidade humana e garantir a justiça e a igualdade no acesso à informação, e a pesquisa deles conclui que a explicabilidade deve ser vista como um direito fundamental que acompanha a evolução das tecnologias e das práticas de governança, desta maneira é imperativo que as instituições desenvolvam mecanismos que assegurem a transparência das decisões automatizadas, promovendo um ambiente institucional onde a responsabilidade e a confiança nas decisões públicas sejam constantemente fortalecidas, outrossim, a construção de um arcabouço normativo que reconheça e proteja o direito à explicabilidade é essencial para a consolidação da democracia e a promoção dos direitos humanos na era digital (Pádua; Lorenzetto, 2024).

Por sua vez, como é de conhecimento, o CNJ tem demonstrado atenção à inteligência artificial no Judiciário e, em 2020, publicou a Resolução nº 332/2020, cuja norma instituiu a Política de Governança e Gestão de Dados Judiciários, além de criar o Comitê Gestor de Governança de Dados e o Observatório Nacional de Inteligência Artificial e Justiça, comportando o escopo de definir diretrizes para a coleta, armazenamento, gestão e compartilhamento de dados judiciários, inclusive os usados em soluções de IA, de acordo com Carini (2024, p. 767):

O CNJ tem se mostrado atento à questão da inteligência artificial no âmbito do Judiciário e já editou outras normas sobre o tema. Em 2020, foi publicada a Resolução CNJ nº 332/2020, que instituiu a Política de Governança e de Gestão de Dados Judiciários no âmbito do Poder Judiciário e criou o Comitê Gestor de Governança de Dados Judiciários e do Observatório Nacional de Inteligência Artificial e Justiça. A norma tem como objetivo estabelecer diretrizes para a coleta, armazenamento, gestão e compartilhamento de dados judiciários, incluindo aqueles utilizados em

soluções de inteligência artificial.

Aliado a isso, cabe mencionar que a Resolução do CNJ nº 363/2021 instituiu a Estratégia Nacional de Inteligência Artificial no Poder Judiciário e criou seu Comitê Gestor, tendo como objetivo a promoção do uso e desenvolvimento de soluções de IA no Judiciário com base em princípios éticos, transparentes e responsáveis, além de incentivar parcerias e cooperação com instituições públicas e privadas (Carini, 2024).

A Resolução nº 363 determina que os sistemas de inteligência artificial no Judiciário devem ser transparentes, justificáveis e explicáveis, assegurando a proteção dos direitos fundamentais, sendo que de igual forma exige avaliações periódicas desses sistemas para garantir sua eficácia e conformidade com normas éticas e legais (Carini, 2024).

Neste cenário, no início desse ano, foi publicada a Resolução CNJ Nº 615/2025 que estabelece normas para o uso ético, transparente e seguro de soluções baseadas em inteligência artificial no Poder Judiciário, com o intuito de promover a inovação tecnológica e a eficiência na prestação jurisdicional, seu conteúdo aborda definições, fundamentos, princípios, categorização de riscos, governança, supervisão e outras diretrizes específicas para o desenvolvimento e uso de tecnologias de IA no âmbito judicial.

A Resolução estabelece que o uso de soluções de IA deve respeitar direitos fundamentais e valores democráticos, entre os fundamentos, destacam-se a centralidade da pessoa humana, a supervisão humana em todas as etapas do ciclo de vida das soluções de IA, a promoção da igualdade, a proteção de dados pessoais e a transparência (Art. 2º), e os princípios incluem justiça, equidade, inclusão, transparência, explicabilidade, contestabilidade, auditabilidade e eficiência. A supervisão humana é obrigatória para garantir que as decisões automatizadas sejam apenas complementares, e não substitutas da atuação humana (Art. 3º) (CNJ, 2025).

A governança das soluções de IA é um aspecto central da Resolução, exigindo que os tribunais implementem processos internos para garantir a segurança, transparência e eficiência das ferramentas utilizadas, deste modo, devem ser adotados mecanismos de auditoria contínua e monitoramento, especialmente em soluções classificadas como de alto risco, para evitar vieses

discriminatórios e garantir conformidade com os direitos fundamentais (Art. 12), e os relatórios de auditoria e avaliações de impacto algorítmico devem ser públicos e acessíveis, promovendo a confiança e a fiscalização pela sociedade e pelos órgãos de controle (Art. 14 e 18) (CNJ, 2025).

A Resolução introduz um sistema de categorização de riscos das soluções de IA em dois níveis: alto risco e baixo risco, soluções de alto risco são aquelas que podem impactar diretamente direitos fundamentais, como as que envolvem a identificação de padrões comportamentais, decisões judiciais ou reconhecimento facial para monitoramento de pessoas, já as soluções de baixo risco são aquelas que auxiliam na gestão processual, como a transcrição de áudios, anonimização de documentos e produção de textos de apoio, sempre com supervisão humana (Art. 9º e Anexo de Classificação de Riscos), as soluções de alto risco exigem auditorias mais rigorosas e monitoramento contínuo, enquanto as de baixo risco têm um acompanhamento mais simplificado.

Ademais, a Resolução enfatiza a importância da qualidade e representatividade dos dados utilizados para treinar soluções de IA, os dados devem ser coletados de fontes confiáveis, preferencialmente públicas ou governamentais, e devem ser anonimizados sempre que possível, em conformidade com a Lei Geral de Proteção de Dados (LGPD) a curadoria dos dados é imprescindível para evitar vieses discriminatórios e garantir a equidade nas decisões judiciais, além disso, é vedado o uso de dados sigilosos ou protegidos por segredo de justiça para treinamento de modelos de IA, salvo nos casos em que os dados sejam anonimizados na origem (Art. 7º, 26 e 30)) (CNJ, 2025).

A supervisão humana é mandatória em todas as etapas do ciclo de vida das soluções de IA, assim, o uso de IA deve ser complementar e auxiliar, nunca substituindo a autoridade final dos magistrados ou servidores, decisões judiciais devem ser revisadas por humanos, e os usuários externos devem ser informados de forma clara e acessível sobre o uso de IA em processos judiciais, destacando o caráter consultivo e não vinculante dessas ferramentas (Art. 33 e 34), de modo que, a Resolução exige a capacitação contínua de magistrados, servidores e outros usuários internos sobre os riscos, limitações e usos éticos das tecnologias de IA (Art. 16 e 19) (CNJ, 2025).

Conforme CNJ (2025) a Resolução proíbe o uso de soluções de IA que não permitam revisão humana ou que apresentem riscos excessivos à segurança da

informação e aos direitos fundamentais, deste modo, exemplos incluem sistemas que criem perfis para prever crimes ou recidiva criminal, ranqueiem pessoas com base em atributos pessoais ou utilizem reconhecimento facial para monitoramento comportamental sem justificativa adequada (Art. 10), e essas medidas visam prevenir o uso abusivo ou discriminatório de tecnologias no sistema de justiça.

Neste contexto, a transparência é um dos pilares da Resolução. Todas as soluções de IA desenvolvidas ou utilizadas no Poder Judiciário devem ser registradas na plataforma Sinapses, que mantém um catálogo público das tecnologias empregadas, e a plataforma também disponibiliza informações sobre classificações de risco, avaliações de impacto algorítmico e relatórios de auditoria, garantindo o controle social e a prestação de contas (Art. 24 e 25) (CNJ, 2025).

Destaca-se com isso que a formação de equipes multidisciplinares e representativas é incentivada para o desenvolvimento de soluções de IA, a diversidade de gênero, etnia e áreas de conhecimento é essencial para evitar vieses durante o planejamento, coleta de dados, desenvolvimento e implementação dos modelos (Art. 35), assim, além disso, os tribunais devem promover treinamentos para magistrados e servidores sobre o uso responsável e eficiente das tecnologias, garantindo que estejam cientes de seus riscos e limitações (Art. 19 e 20) (CNJ, 2025).

Ressalta-se que a resolução encoraja a colaboração entre tribunais, órgãos públicos e a sociedade civil para o desenvolvimento de soluções inovadoras. Modelos de IA devem ser interoperáveis e compartilhados entre diferentes instituições judiciais, evitando duplicidade de esforços e promovendo a eficiência do sistema (Art. 12 e 43), e a cooperação com parceiros externos deve seguir estritas diretrizes de segurança e proteção de dados (CNJ, 2025).

Os tribunais têm um prazo de 12 meses para adequar suas soluções de IA às novas disposições da Resolução, que entra em vigor após 120 dias da sua publicação, e durante esse período, os projetos em andamento devem ser ajustados para cumprir com as normas estabelecidas, garantindo a conformidade com os princípios éticos e jurídicos (Art. 45 a 47) (CNJ, 2025).

Mediante uma análise mais acurada da Resolução CNJ nº 615/2025, foi revelado que, embora existam diretrizes que promovem articulações para o uso seguro da tecnologia, há um caminho a percorrer na definição de mecanismos de monitoramento contínuos e reativos, uma vez que a aludida diretriz muitas vezes

menciona a necessidade de uma IA explicável, mas ainda não estabelecem metodologias claras sobre sua implementação prática (Kenney; Zysman, 2018).

Embora a Resolução CNJ nº 615/2025 represente um progresso substancial na regulamentação do uso de tecnologia de IA no judiciário, é relevante destacar a necessidade de serem implementadas políticas mais robustas que incluam metodologias práticas de auditoria e monitoramento, de forma que a instituição de uma IA explicável é indispensável não apenas para aumentar a confiança do público, mas também para alinhar práticas tecnológicas com princípios regulatórios (Kenney; Zysman, 2018).

Assim, a Resolução CNJ Nº 615/2025 é um marco regulatório fundamental para o uso de inteligência artificial no Poder Judiciário brasileiro, ao estabelecer diretrizes para governança, transparência, auditoria e supervisão, a Resolução busca equilibrar a inovação tecnológica com a proteção dos direitos fundamentais, promovendo um uso ético, seguro e eficiente das tecnologias no sistema de justiça.

Portanto, os mecanismos de mitigação e regulação dos vieses algorítmicos envolvem um conjunto de práticas e princípios interligados, que vão desde a etapa inicial de coleta e curadoria dos dados, passando pela auditoria contínua dos modelos, até a transparência e inteligibilidade das decisões tomadas pelas máquinas, esses instrumentos, quando aplicados de forma articulada e ética, são essenciais para garantir que os avanços tecnológicos não reproduzam as injustiças do passado, mas sim contribuam para a construção de uma sociedade mais justa, inclusiva e democrática.

Diante do impacto crescente da inteligência artificial nas decisões que afetam diretamente a vida das pessoas, torna-se indispensável identificar e mitigar os riscos relacionados ao viés algorítmico e à discriminação automatizada, dessarte, pode se dizer que o desafio está em assegurar que essas tecnologias não apenas otimizem processos, mas também respeitem os princípios de equidade, justiça e direitos fundamentais, sendo que a construção de sistemas mais transparentes, auditáveis e livres de preconceitos é um passo fundamental rumo a uma sociedade digital mais ligada aos direitos fundamentais.

5 CONCLUSÃO

Este trabalho analisou a complexidade que envolve o uso de algoritmos em decisões judiciais, com especial atenção ao problema do viés algorítmico. Embora as tecnologias digitais possam trazer ganhos significativos em termos de eficiência, celeridade processual e padronização de entendimentos, a pesquisa demonstrou que a implementação da Inteligência Artificial (IA) no Judiciário, se realizada sem critérios rigorosos e mecanismos de controle, pode comprometer valores essenciais da justiça. Nesse sentido, o estudo reforça a importância de se buscar um equilíbrio entre inovação tecnológica e os princípios constitucionais, garantindo que os avanços digitais não resultem em discriminação ou em perdas de direitos fundamentais.

A análise dos vieses algorítmicos revelou que uma das principais causas de distorções nos resultados produzidos por sistemas automatizados encontra-se nos dados utilizados para seu treinamento. Tais dados, muitas vezes reflexo de desigualdades históricas e estruturais, tendem a reproduzir e perpetuar preconceitos sociais que atingem, sobretudo, grupos em situação de vulnerabilidade. Dessa forma, torna-se evidente que a governança de dados e a transparência algorítmica devem ser compreendidas como pilares fundamentais para a construção de um sistema de IA que respeite os valores democráticos. A seleção criteriosa, a limpeza adequada e a auditoria contínua dos dados são etapas indispensáveis para evitar a replicação de padrões discriminatórios em decisões judiciais.

Além disso, a pesquisa destacou a necessidade de uma abordagem interdisciplinar para o desenvolvimento e a supervisão de sistemas de IA. A colaboração entre especialistas das áreas do Direito, da ciência de dados, da ética e da tecnologia é essencial para mitigar riscos e assegurar que as decisões automatizadas sejam legítimas, imparciais e justas. Nesse contexto, a formação continuada de profissionais do Judiciário e de desenvolvedores tecnológicos, com foco na ética, nos direitos fundamentais e na explicabilidade, é apontada como elemento-chave para a construção de soluções que estejam em consonância com os valores do Estado Democrático de Direito. Ademais, ressaltou-se que a supervisão humana deve permanecer como componente indispensável durante todo o ciclo de vida dos sistemas de IA, garantindo que a tecnologia complemente — e não substitua — o papel crítico dos operadores da justiça.

A pesquisa conclui que a inteligência artificial deve ser entendida como uma ferramenta de apoio, e não como substituta da atuação humana no Judiciário. Apesar de sua sofisticação, a IA não é capaz de interpretar normas com sensibilidade, considerar aspectos éticos ou lidar com situações singulares. Seu aprendizado se baseia exclusivamente em padrões do passado, o que a torna limitada diante de casos que exigem julgamento crítico, criatividade e soluções inéditas — competências que continuam sendo exclusivas dos seres humanos.

Nesse sentido, a IA é especialmente útil em demandas repetitivas, previsíveis e baseadas em padrões já consolidados, nas quais pode agilizar o trabalho ao identificar soluções já aplicadas. No entanto, sua natureza estatística e retroativa impede que ela ofereça respostas inovadoras ou trate adequadamente casos que envolvem complexidade, nuances jurídicas e valores humanos. Assim, embora seja uma aliada importante na automação de tarefas, a IA jamais poderá substituir o papel insubstituível do ser humano nas decisões judiciais que envolvem ética, interpretação e inovação.

Alem disso, a utilização da Inteligência Artificial no Poder Judiciário brasileiro apresenta-se como um campo repleto de potencialidades, mas também de desafios que não podem ser ignorados. O principal objetivo dessas tecnologias é modernizar o sistema judicial, oferecendo maior eficiência, celeridade processual e uniformidade nas decisões, ao mesmo tempo em que busca ampliar o acesso à justiça. No entanto, o problema do viés algorítmico demonstra ser um entrave significativo à concretização desses objetivos, pois ameaça comprometer pilares fundamentais do sistema jurídico, como a equidade, a imparcialidade e a segurança jurídica. A introdução da IA nos tribunais exige, portanto, uma abordagem responsável e sensível às complexidades éticas, técnicas e sociais envolvidas, considerando os múltiplos impactos dessas tecnologias nos processos decisórios.

Entre os principais problemas identificados ao longo da pesquisa, o viés algorítmico mostrou-se especialmente preocupante, uma vez que pode comprometer a imparcialidade, a equidade e a transparência das decisões judiciais. Tais vieses, que muitas vezes decorrem de dados históricos enviesados ou de escolhas inadequadas no desenvolvimento dos algoritmos, têm o potencial de perpetuar desigualdades sociais e afetar diretamente a legitimidade das decisões automatizadas. Assim, restou demonstrado que, embora a IA possa oferecer avanços consideráveis para o Judiciário, sua adoção deve ser acompanhada de

critérios técnicos e jurídicos rigorosos, bem como de mecanismos de mitigação eficazes que garantam a justiça, a equidade e o respeito aos direitos fundamentais.

Nesse contexto, a governança de dados emerge como um dos elementos mais relevantes para enfrentar os desafios levantados. A curadoria de dados — que compreende a seleção, organização e limpeza das informações utilizadas no treinamento de algoritmos — é fundamental para assegurar que os conjuntos de dados sejam representativos, diversificados e livres de preconceitos estruturais. Paralelamente, é imprescindível que os sistemas de IA sejam concebidos com mecanismos de explicabilidade e auditabilidade, permitindo que as decisões automatizadas sejam compreensíveis, rastreáveis e passíveis de contestação. Essas medidas, além de promoverem a transparência, reforçam a confiança pública no uso de tecnologias no âmbito do Judiciário.

Constatou-se que a ausência de explicabilidade nas decisões automatizadas compromete diretamente os princípios do contraditório e da ampla defesa, dificultando a fiscalização e a responsabilização por eventuais falhas ou injustiças. Os vieses algorítmicos, muitas vezes invisíveis, refletem desigualdades históricas e estruturais presentes nos dados que alimentam os sistemas de IA. Esses vieses podem surgir em diferentes etapas do desenvolvimento e da aplicação das tecnologias, desde a coleta e a curadoria dos dados até a modelagem, treinamento e implementação prática dos algoritmos. Quando não mitigados, esses problemas têm o potencial de reproduzir ou intensificar desigualdades, afetando diretamente a legitimidade das decisões judiciais e a confiança da sociedade nas instituições.

Diante disso, esta dissertação traz uma contribuição significativa ao enfatizar a necessidade de práticas de governança e curadoria de dados que assegurem a qualidade, a representatividade e a diversidade dos conjuntos informacionais utilizados nos sistemas de IA aplicados ao Judiciário. A seleção criteriosa, a higienização e a auditoria periódica dos dados são etapas indispensáveis para evitar que padrões discriminatórios sejam replicados em decisões judiciais. Além disso, a transparência na construção e na utilização de algoritmos é essencial para fomentar a confiança social, possibilitando que os cidadãos compreendam os fundamentos das decisões e disponham de instrumentos eficazes para contestá-las, quando necessário.

Outro ponto de destaque refere-se à necessidade de uma regulamentação robusta, dinâmica e adaptável que acompanhe os avanços tecnológicos e ofereça diretrizes claras para sua aplicação no Poder Judiciário. A criação de marcos regulatórios como a Resolução CNJ nº 615/2025 representa um avanço importante, ao estabelecer princípios de governança, transparência e auditabilidade voltados ao uso da IA nas decisões judiciais. Tais regulamentações contribuem para prevenir riscos relacionados à opacidade e à discriminação, promovendo maior segurança jurídica e fortalecimento da confiança pública. Contudo, ainda persistem lacunas a serem enfrentadas, especialmente no que se refere à inclusão de mecanismos mais eficazes de responsabilização, explicabilidade e supervisão contínua.

A regulamentação e a supervisão das tecnologias de IA devem ser compreendidas como elementos centrais de uma política pública voltada à promoção da justiça algorítmica. A implementação de auditorias regulares, avaliações de impacto, categorização de risco e supervisão humana permanente, conforme previsto na Resolução CNJ nº 615/2025, assegura que as tecnologias utilizadas não violem direitos fundamentais. Ademais, a transparência e a prestação de contas devem ser incorporadas como princípios fundamentais do desenvolvimento e da gestão de sistemas de IA no Judiciário, garantindo a compatibilidade das tecnologias com os valores democráticos e constitucionais.

A interdisciplinaridade também se revela como aspecto crucial na mitigação dos riscos associados ao viés algorítmico. A colaboração entre juristas, cientistas de dados, engenheiros da computação e especialistas em ética proporciona a construção de soluções mais robustas, justas e ajustadas aos princípios do Estado de Direito. Essa perspectiva é particularmente relevante no contexto brasileiro, marcado por desigualdades estruturais e desafios no acesso à justiça. A capacitação contínua de operadores do Direito e profissionais da tecnologia é, portanto, indispensável para assegurar uma implementação responsável e efetiva da IA no sistema de justiça.

Assim, a criação de marcos normativos claros e abrangentes, aliados à supervisão humana constante e à capacitação interdisciplinar, representa a base para uma incorporação ética da Inteligência Artificial no Judiciário. A IA deve ser compreendida como um instrumento de apoio à atividade jurisdicional, jamais como substituta do discernimento humano, de modo a preservar a essência do julgamento

como prática racional, empática e humanizada. O futuro da justiça dependerá, sobretudo, da forma como a tecnologia será integrada aos valores humanos, de forma equilibrada, justa e responsável.

Portanto, a formulação de políticas públicas específicas, a implementação de mecanismos eficientes de supervisão e a adoção de práticas consistentes de governança e curadoria de dados são passos imprescindíveis para assegurar que a tecnologia atue como aliada na promoção de um sistema judiciário mais acessível, transparente e eficiente. O futuro da inteligência artificial no Poder Judiciário dependerá diretamente da capacidade institucional de equilibrar inovação tecnológica e proteção dos direitos fundamentais, consolidando um modelo de justiça que seja, ao mesmo tempo, inclusivo, ético e transparente.

Por fim, a presente pesquisa reforça a importância de conciliar a busca por inovação com a preservação dos direitos fundamentais. A eficiência e a celeridade processual não podem ser alcançadas em detrimento de princípios como a equidade, a imparcialidade e a justiça. A consolidação de políticas públicas, regulamentos e práticas de governança robustas, aliada a uma abordagem verdadeiramente interdisciplinar, constitui a melhor estratégia para mitigar os riscos associados ao viés algorítmico e garantir que as decisões automatizadas sejam legítimas, transparentes e compatíveis com os valores democráticos. Dessa forma, a integração equilibrada entre tecnologia e justiça poderá fortalecer a segurança jurídica e contribuir para um Judiciário mais eficiente, acessível e justo.

REFERÊNCIAS

ABBOTT, D. **What is Digital Curation?** Digital Curation Centre, [s.d.]. Disponível em: <https://www.dcc.ac.uk/guidance/briefing-papers/introduction-curation/what-digital-curation>. Acesso em: 2 abr. 2025.

ALE, Humberto. Juiz do TRF1 que usou o ChatGPT para elaborar decisão será investigado pelo CNJ. 2023. Disponível em: <https://www.jota.info/justica/juiz-do-trf1-que-usou-o-chatgpt-para-elaborar-decisao-sera-investigado-pelo-cnj-13112023>. Acesso em: 3 jun. 2024.

ALKAISSI, Hussam; MCFARLANE, Samy I. Artificial hallucinations in ChatGPT: implications in scientific writing. **Cureus**, [S.l.], v. 15, n. 2, 2023.

ALCOLEA, E. E. G. Vida e inteligencia artificial. **ACIMED**, Ciudad de La Habana, v. 19, n. 1, jan. 2009. Disponível em: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S102494352009000100006&lng=es&nrm=iso&tlng=es. Acesso em: 18 fev. 2024.

ALEXOPOULOS, Charalampos et al. ChatGPT application vis-a-vis Open Government Data (OGD): capabilities, public values, issues and a research agenda. 2023. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=437877. Acesso em: 2 abr. 2024.

ALMEIDA, Virgilio et al. ChatGPT: tecnologia, limitações e impactos. **Ciência Hoje**, [S.l.], mar. 2023. Edição 396. Disponível em: <https://cienciahoje.org.br/artigo/chatgpttecnologia-limitacoes-e-impactos/>. Acesso em: 2 abr. 2024.

ANDRADE, Mariana Dionísio de; ROSA, Beatriz de Castro; PINTO, Eduardo Régis Girão de Castro. Legal tech: analytics, inteligência artificial e as novas perspectivas para a prática da advocacia privada. **Revista Direito GV**, São Paulo, v. 16, n. 1, p. [não informado], jan./abr. 2020.

ANDRIGHI, Fátima Nancy; BIANCHI, José Flávio. Reflexão sobre os riscos do uso da inteligência artificial ao processo de tomada de decisões no Poder Judiciário. In: PINTO, Henrique Alves; GUEDES, Jefferson Carús; CÉSAR, Joaquim Portes de Cerqueira (coord.). **Inteligência artificial aplicada ao processo de tomada de decisões**. Belo Horizonte; São Paulo: D'Plácido, 2020. p. 173–190.

AQUINO, Pedro Gabriel dos Santos. **Inteligência artificial: os impactos causados no judiciário brasileiro**. Gama: UNICEPLAC, 2023.

AVELAR, Dayanne. Construindo um Judiciário sustentável: a incorporação da inteligência artificial como catalisador de mudança. **Migalhas**, 2024. Disponível em: <https://www.migalhas.com.br/depeso/406804/construindo-um-judiciario-sustentavel>. Acesso em: 18 jun. 2024.

ÁVILA, Gustavo Noronha de; CORAZZA, Thaís Aline Mazetto. **Revista da AJURIS**,

Porto Alegre, v. 49, n. 152, jun. 2022.

BARDIN, L. **Análise de conteúdo**. Lisboa: Edições 70, 2011.

BAROCAS, S.; HARDT, M.; NARAYANAN, A. **Fairness and Machine Learning: Limitations and Opportunities**. Cambridge: MIT Press, 2019. Disponível em: <https://fairmlbook.org>. Acesso em: 4 abr. 2025.

BAROCAS, S.; SELBST, A. D. **Big Data's Disparate Impact**. *California Law Review*, 2016.

BARROSO, Luís Roberto. **Neoconstitucionalismo e constitucionalização do Direito**. *Revista Quaestio Iuris*, v. 2, n. 1, p. 1-48, 2006.

BARROZO, V. L. S. **Curadoria e preservação digital: diretrizes para a gestão de dados científicos**. 2022. Dissertação (Mestrado em Gestão & Organização do Conhecimento) – Universidade Federal de Minas Gerais. Disponível em: <http://hdl.handle.net/1843/34550>. Acesso em: 4 abr. 2025.

BAHIA, A.; NUNES, D.; PEDRON, F. **Teoria geral do processo**. Salvador: JusPodvim, 2021.

BENNETT MOSES, L. **Allocating Accountability in AI Decision Making: A New Perspective on AI Liability**. *University of New South Wales Law Journal*, 2017.
BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: https://www.planalto.gov.br/ccivil_03/ato2015-2018/2018/lei/113709.htm. Acesso em: 16 mar. 2025.

BRAVE. CNJ vai investigar juiz que usou tese inventada pelo ChatGPT para escrever decisão. *Conjur*, 2023. Disponível em: <https://www.conjur.com.br/2023-nov-12/cnj-vai-investigar-165juiz-que-usou-tese-inventada-pelo-chatgpt-para-escrever-decisao/>. Acesso em: 15 nov. 2023.

BROCHADO, M. **Inteligência Artificial: um diálogo com Lima Vaz**. *Kriterion: Revista de Filosofia*, v. 64, n. 154, p. 75-98, 2023. DOI: <https://doi.org/10.1590/0100-512X2023n15404mb>. Disponível em: <https://www.scielo.br/j/kr/a/4NKGBGSPn3J8KDBb44VBTBf/?format=pdf&lang=pt>. Acesso em: 23 fev. 2024.

BURRELL, J. **How the machine “thinks”: understanding opacity in machine learning algorithms**. *Big Data & Society*, jan.–jun. 2016, p. 1–12. Disponível em: <http://journals.sagepub.com/doi/pdf10.1177/2053951715622512>.

CAMBI, E. A. S.; AMARAI, M. E. T. P. T. **Inteligência artificial no Poder Judiciário, discriminação algorítmica e direitos humanos fundamentais**. *SUPREMA – Revista de Estudos Constitucionais*, Brasília, v. 3, n. 2, p. 189-218, jul./dez. 2023. Disponível em: <https://suprema.stf.jus.br/index.php/suprema/article/download/250/122/527>. Acesso em: 14 mar. 2025.

CARINI, Lucas. **Inteligência artificial e Poder Judiciário: análise das políticas regulatórias no Brasil e na Itália.** *Lumen et Virtus*, São José dos Pinhais, v. XV, n. XXXVIII, 2024.

CARLE, Eben. **Ask a Techspert: What is generative AI?** *Google The Keyword*, 2023. Disponível em: <https://blog.google/inside-google/googlers/ask-a-techspert/what-is-generative-ai/>. Acesso em: 02 abr. 2024.

CARVALHO, A. C. P. L. F. **Inteligência Artificial: riscos, benefícios e uso responsável.** *Estudos Avançados*, v. 35, n. 101, p. 21-36, 2021. DOI: <https://doi.org/10.1590/s0103-4014.2021.35101.003>. Disponível em: <https://www.scielo.br/j/ea/a/ZnKyr-crLVqzhZbXGgXTwDtn/?format=pdf&lang=pt>. Acesso em: 12 mar. 2025.

CARVALHO, Allan Pereira de. **Viés Algorítmico e Discriminação: possíveis soluções regulatórias para o Brasil.** Porto Alegre: UFRGS, 2020.

CASTRO, Eduardo Viveiros de. **As três inteligências – notas para um artigo.** *ResearchGate*, 2021. Disponível em: <https://www.researchgate.net/publication/351273509> As tres inteligencias - notas para um artigo. Acesso em: 11 set. 2023.

CAVALCANTE, J. S. C. **Os desafios éticos no desenvolvimento de sistemas inteligentes no poder judiciário brasileiro: Uma perspectiva acerca dos riscos aos Direitos Fundamentais.** 2023. Dissertação (Mestrado em Direito) – Universidade Estácio de Sá, 2023. Disponível em: <https://dissertacao.estacio.br/direito/2023/JAMILE%20SABBAD%20CARECHO%20CAVALCANTE.pdf>. Acesso em: 12 mar. 2025.

CHURCHLAND, P. M.; CHURCHLAND, P. S. **Uma máquina poderia pensar? É improvável que a IA clássica produza máquinas conscientes, mas sistemas que imitam o cérebro talvez consigam.** *Revista Eletrônica de Filosofia*, v. 12, n. 1, p. 157-169, jan./jun. 2015. Disponível em: <https://revistas.pucsp.br/index.php/cognitio/article/view/22125/16870>. Acesso em: 28 fev. 2024.

CNJ. Conselho Nacional de Justiça. **Curso Introdução à Inteligência Artificial para o Poder Judiciário.** 2023. Disponível em: <https://www.cnj.jus.br/curso-online-introduz-conceitos-de-inteligencia-artificial-para-o-judiciario/>. Acesso em: 18 jun. 2024.

CNJ. Conselho Nacional de Justiça. **O uso da inteligência artificial generativa no Poder Judiciário Brasileiro.** Brasília: CNJ, 2024.

CNJ. Conselho Nacional de Justiça. **Resolução nº 332/2020.** Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3429>. Acesso em: 14 mar. 2025.

CNJ. Conselho Nacional de Justiça. **Resolução nº 615, de 11 de março de 2025.** Disponível em: <https://atos.cnj.jus.br/files/original1555302025031467d4517244566.pdf>. Acesso em:

02 abr. 2025.

COELHO, J. V. A. B. R. **Aplicações e implicações da inteligência artificial no direito**. 2017. Monografia (Bacharelado em Direito) – Universidade de Brasília, Brasília, 2017.

CORETRUSTSEAL STANDARDS AND CERTIFICATION BOARD. **CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022**. CoreTrustSeal [online]. [viewed 20 October 2021]. Available from: <https://doi.org/10.5281/zenodo.3638211>. Acesso em: 30 mar. 2025.

CORTIZ, Diogo. **Inteligência Artificial: equidade, justiça e consequências**. *Panorama Setorial da Internet*, n. 1, maio 2020, pp. 1-5, p. 3. Disponível em: https://cetic.br/media/docs/publicacoes/6/20200626161010/panorama_setorial_anoxii_n_1_inteligencia_artificial_equidade_justica.pdf. Acesso em: 26 fev. 2024.

COSTA, Roselene Henrique Pereira; LIMA, Bruno Rodrigues; FERNANDES, Antonio de Pádua Lima; GONÇALVES, Carla Maria Montanari; Paula, Edmar Mendes de. **[Título da obra faltando]**.

COWGILL, B.; DELL'ACQUA, F.; DENG, S. **The Impact of Algorithms on Judicial Decision-Making**, 2020.

COZMAN, F. G. **No canal da Inteligência Artificial – Nova temporada de desgrenhados e empertigados**. *Estudos Avançados*, v. 35, n. 101, p. 7-20, 2021. DOI: <https://doi.org/10.1590/s0103-4014.2021.35101.002>. Disponível em: <https://www.scielo.br/j/ea/a/q3MZJVGqtrrhYwZy4vt54w/?format=pdf&lang=pt>. Acesso em: 22 fev. 2024.

CRAWFORD, K. **Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence**. New Haven: Yale University Press, 2021.

Curty, R. L.; Silva, F. C. **Gestão de Dados Científicos: Boas Práticas e Diretrizes**. *Revista Brasileira de Ciência da Informação*, v. 15, n. 2, p. 177-193, 2021. Disponível em: https://www.researchgate.net/publication/343234567_Gestao_de_Dados_Cientificos_Boas_Pratica_e_Diretrizes. Acesso em: 30 mar. 2025.

DABASS, Jyoti; DABASS, Bhupender Singh. **Scope of Artificial Intelligence in Law**. The Northcap University, 2018. DOI: 10.20944/preprints201806.0474.v1.

DAMILANO, C. T. **O uso da inteligência artificial como ferramenta para garantia de uma maior eficácia pela administração pública no Brasil**. Disponível em: <https://www.academia.edu/download/91800816/cap. livro Padova Vol. 1 Direitos Fundamentais Italobrasileiro 1 .pdf#page=470>. Acesso em: 01 out. 2023.

DE PLÁCIDO, E. Silva; SLAIBI FILHO, Nagib; CARVALHO, Gláucia. **Vocabulário jurídico**. Forense, 2002.

DIGIAMPIETRI, Luciano Antonio. **O que é ChatGPT?** EACH-USP. Disponível em: http://www.each.usp.br/digiampietri/SIN5028/06_ChatGPT.pdf. Acesso em: 02 abr.

2024.

DIGITAL CURATION CENTER (DCC). **A graphical, high-level overview of the stages required for successful management, curation and preservation of data.** Reino Unido, julho de 2022. Disponível em: <https://www.dcc.ac.uk/guidance/curation-lifecycle-model>. Acesso em: 03 abr. 2025.

DIAS, Z. **Algoritmos de Busca: Algoritmos e Programação de Computadores.** Instituto de Computação. Universidade Estadual de Campinas. 2020. Disponível em: <https://ic.unicamp.br/~mc102/aulas/aula11.pdf>. Acesso em: 23 fev. 2024.

DIAS, Z. **Algoritmos e Programação de Computadores.** Instituto de Computação. Universidade Estadual de Campinas. 2023. Disponível em: <https://ic.unicamp.br/~mc102/aulas/aula12.pdf>. Acesso em: 23 fev. 2024.

DINAMARCO, Cândido Rangel. **Instituições de direito processual civil**, v. I e II. São Paulo: Malheiros, 2001. p. 489.

DUQUE-PEREIRA, Ives da Silva; MOURA, Sergio Arruda de. **Understanding Generative Artificial Intelligence From The Language Perspective.** SciELO Preprints, 2023, p. 11. DOI: 10.1590/SciELOPreprints.7077.

ENAP. **Análise de Dados em Linguagem R.** Brasília: Enap, 2020.

ESCOVEDO, Tatiana; KOSHIYAMA, Adriano. **Introdução a Data Science.** Casa do Código, 2020.

EUROPARL. Parlamento Europeu. **Regulamento Inteligência Artificial: Parlamento aprova legislação histórica.** 2024. Disponível em: <https://www.europarl.europa.eu/news/pt/pressroom/20240308IPR19015/regulamento-inteligencia-artificial-parlamento-aprova-legislacao-historica>. Acesso em: 10 mar. 2025.

EUBANKS, V. **Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor.** New York: St. Martin's Press, 2018.

FARACO, F. M.; TODESCO, J. L. **Inteligência artificial aplicada ao direito – uma revisão sistemática da literatura.** VIII Congresso Internacional de Conocimiento e Innovación, 2018.

FERRARI, Fabricio; CECHINEL, Cristian. **Introdução a algoritmos e programação.** Disponível em: <https://www.ferrari.pro.br/home/documents/FFerrari-CCechinel-Introducao-a-algoritmos.pdf>. Acesso em: 24 mar. 2025.

FERRARI, O. **Accountability de Algoritmos: a falácia do acesso ao código e caminhos para uma explicabilidade efetiva.** Inteligência Artificial: 3º Grupo de Pesquisa ITS, 2018. Disponível em: <https://itsrio.org/pt/publicacoes/inteligencia-artificial-gp3/>. Acesso em: 24 fev. 2024.

FERREIRA, H. et al. **Introduzindo aprendizado de máquina em cursos de física:**

o caso do rolamento no plano inclinado. *Revista Brasileira de Ensino de Física*, v. 44, 2022. DOI: <https://doi.org/10.1590/1806-9126-RBEF-2022-0214>. Disponível em: <https://www.scielo.br/j/rbef/a/JS7GjGRH5HmCngFZSpvFZHK/?format=pdf&lang=pt>. Acesso em: 27 fev. 2024.

FERREIRA, R. A. M. **Direito e democracia: tecnologia, inteligência artificial e fake news.** 2020. Disponível em: <https://books.google.com/books?hl=en&lr=&id=BcgOEAAAQBAJ&oi=fnd&pg=PT3&dq=Intelig%C3%Aancia+Artificial+e+Processo+Civil:+Riscos+do+positivismo+tecnol%C3%B3gico+no+decisionismo+Judicial+na+Direito&ots=Erz6eGOiMo&sig=gE0ID6-tfrK0UkCvuCFY40cZO8w>. Acesso em: 01 out. 2023.

FERRAZ, T. P. et al. **Inteligência Artificial explicável para atenuar a falta de transparência e a legitimidade na moderação da Internet.** *Estudos Avançados*, v. 38, n. 111, 2024. Disponível em: <https://www.scielo.br/j/ea/a/KPMcWYkkqHy5ZK3zTFCBpFj/?format=pdf&lang=pt>. Acesso em: 15 mar. 2025.

FLACH, P. **Machine Learning: The Art and Science of Algorithms that Make Sense of Data.** Cambridge University Press, 2012.

FONTANA, E. **Introdução aos Algoritmos de Aprendizagem Supervisionada.** Universidade Federal do Paraná - UFPR, Departamento de Engenharia Química, 2020. Disponível em: https://fontana.paginas.ufsc.br/files/2018/03/apostila_ML_pt2.pdf. Acesso em: 27 fev. 2024.

FROOMKIN, A. M.; KERR, I. R.; PINEAU, J. **When AIs Outperform Doctors: Confronting the Challenges of a Tort-Induced Over-Reliance on Machine Learning.** *Ariz. L. Rev.*, v. 61, p. 33, 2019. University of Miami Legal Studies Research Paper No. 18-3. DOI: <http://dx.doi.org/10.2139/ssrn.3114347>. Disponível em: <https://ssrn.com/abstract=3114347>. Acesso em: 15 fev. 2024.

FRANCO JUNIOR, Abelardo; CARDOSO, Aline Cristine Machado. **A inteligência artificial utilizada no Poder Judiciário para garantia da celeridade processual: uma análise de sua aplicação, suas implicações éticas, desafios e os limites de sua utilização.** VII Encontro Virtual do CONPEDI, Direito, Governança e Novas Tecnologias II, 2024.

GABRIEL, Markus. **O sentido do pensar: A filosofia desafia a inteligência artificial.** Tradução de Lucas Machado. Petrópolis, RJ: Vozes, 2021.

GANGULI, Deep et al. **Predictability and Surprise in Large Generative Models.** In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery, New York, NY, USA. p. 5. Disponível em: <https://doi.org/10.1145/3531146.3533229>. Acesso em: 02 abr. 2024.

GOMES, E. B.; VAZ, A. A.; DIAS, S. M. O. **Limites éticos para o uso da inteligência artificial no sistema de justiça brasileiro, de acordo com a Lei.**

13.709 de 2018 (LGPD) e Resoluções 331 e 332 do Conselho Nacional de Justiça. *CONSINTER*, v. 13, 2021. Disponível em: https://scielo.pt/scielo.php?pid=S2183-95222021000200107&script=sci_arttext. Acesso em: 25 mar. 2025.

GONÇALVES, Giselle Aline dos Santos et al. **Algoritmos e Técnicas de Programação**. São Paulo: Científica, 2023.

GONÇALVES, Rafael. **Eficiência como viés algorítmico nas técnicas de aprendizado de máquina: caracterização baseada na produção tecnocientífica da Google**. XXX Congresso de Iniciação Científica da UNICAMP, 2022.

GROSSMAN, Maura R. et al. **The GPTJudge: Justice in a Generative AI World**. *Duke Law & Technology Review*, v. 23, n. 1, 2023. Disponível em: <https://scholarship.law.duke.edu/dltr/vol23/iss1/1>. Acesso em: 04 abr. 2024.

GUNNING, D.; AHA, D. **DARPA's Explainable Artificial Intelligence (XAI) Program**. *AI Magazine*, v. 40, n. 2, p. 44-58, 2019. Disponível em: <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/2850>. Acesso em: 29 mar. 2025.

HACKER, Philipp et al. **Regulating ChatGPT and other Large Generative AI Models**. Disponível em: <https://arxiv.org/abs/2302.02337>. Acesso em: 03 abr. 2024.

HAN, Byung-Chul. **Não-coisas: Reviravolta no mundo das coisas**. Tradução de Rafael Rodrigues Garcia. Petrópolis, RJ: Vozes, 2022.

HARTMANN, F.; ZUMBLICK, R. **Inteligência Artificial e Direito**. Curitiba: Alteridade Editora, 2019. p. 100.

HASSELMANN, Gustavo. **Os limites da inteligência artificial no Direito**. 2019. *Migalhas*. Disponível em: <https://www.migalhas.com.br/depeso/310527/os-limites-da-inteligencia-artificial-no-direito>. Acesso em: 18 jun. 2024.

HIGASI, P. K. B. **O direito aplicável às decisões produzidas por software e machine learning**. 2020. 74 f. Dissertação (Mestrado em Tecnologias da Inteligência e Designe Digital) – Pontifícia Universidade Católica de São Paulo, PUC-SP, São Paulo. Disponível em: <https://sapientia.pucsp.br/bitstream/handle/23659/2/PI%C3%ADnio%20Kentar%C3%ADe%20Britto%20Costa%20Higasi.pdf>. Acesso em: 28 fev. 2024.

HOGEMANN, E. R. **O futuro do Direito e do ensino jurídico diante das novas tecnologias**. *Revista Interdisciplinar de Direito*, v. 16, n. 1, 2018.

IBM. **Structured, semi-structured, and unstructured data: what's the difference?** *IBM Cloud Education*, 2023.

IBM. **O que é IA explicável?** Disponível em: <https://www.ibm.com/think/topics/explainable-ai>. Acesso em: 01 abr. 2025.

ILUMEO. **Vieses sociais da inteligência artificial: como eles acontecem na prática?** Disponível em: <https://ilumeo.com.br/categorias/2021-07-20-vieses-sociais-da-inteligencia-artificial-como-eles-acontecem-na-pratica/#:~:text=Dados%2C%20algoritmos%20e%20pessoas%2C%20as,selecionam%20os%20dados%20de%20treinamento.> Acesso em: 24 mar. 2025.

JONKER, Alexandra; ROGERS, Julie. **O que é viés algorítmico?** Disponível em: <https://www.ibm.com/br-pt/think/topics/algorithmic-bias#:~:text=O%20vi%C3%A9s%20algor%C3%ADtmico%20ocorre%20quando,e%20de%20q%C3%AAnero%20j%C3%A1%20existentes.> Acesso em: 24 mar. 2025.

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

KABIR, Samia et al. Is stack overflow obsolete? an empirical study of the characteristics of chatgpt answers to stack overflow questions. In: **Proceedings of the CHI Conference on Human Factors in Computing Systems.** 2024. p. 1-17.

KENNEY, M.; ZYSMAN, J. **The Next Phase in the Digital Revolution: Intelligent Tools, Platforms, Growth, and Employment.** Journal of Industry, Competition and Trade, 2018.

LAGE, Fernanda de Carvalho. **Manual de Inteligência Artificial no Direito Brasileiro.** Editora JusPodivm, 2021.

LEE, Kai-Fu. **Inteligência artificial: como os robôs estão mudando o mundo, a forma como amamos, nos relacionamos, trabalhamos e vivemos.** Rio de Janeiro: Globo Livros, 2019.

LEITE, T. M. **Deep learning em dois estágios para detecção e classificação de doenças em folhas de plantas com aplicação em dispositivos móveis.** 2022. 70 f. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) Universidade de São Paulo – São Carlos. 2022. Disponível em: https://www.teses.usp.br/teses/disponiveis/55/55134/tde-02022022-121135/publico/TiagodeMirandaLeite_revisada.pdf. Acesso em: 28 fev. 2024.

LIMA, Patrícia Cunha de; SIQUEIRA, Fernando Soares de. Hadoop: uma plataforma de dados não convencional para apoiar a implantação da cultura de orientação a dados nas empresas. v. 9, n. 1, 2019: **Anais do XV Encontro de Iniciação Científica da UNI7.**

LIMA, T. M. M.; SÁ, M. F. F. Inteligência Artificial e Lei Geral de Proteção de Dados Pessoais: O direito à explicação nas decisões automatizadas. **Revista Brasileira de Direito Civil. RBD Civil**, v. 26, p. 227-246, out./dez. 2020. DOI: 10.33242/rbdc.2020.04.011. Disponível em: <https://rbdcivil.ibdcivil.org.br/rbdc/article/view/584/425>. Acesso em: 28 fev. 2024.

LUDERMIR, T. B. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. **Revista Estudos Avançados**, v. 35, n. 101, p. 85-94, 2021. DOI: <https://doi.org/10.1590/s0103-4014.2021.35101.007>. Disponível em:

<https://www.scielo.br/j/ea/a/wXBdv8yHBV9xHz8qG5RCgZd/?format=pdf&lang=pt>. Acesso em: 20 fev. 2024.

MAIA FILHO, M. S.; JUNQUILHO, T. A. Projeto Victor: perspectivas de aplicação da inteligência artificial ao direito. **R. Dir. Gar. Fund.**, Vitória, v. 19, n. 3, p. 219-238, set./dez. 2018.

MARANHÃO, Juliano Souza de Albuquerque; JUNQUILHO, Tainá Aguiar; TASSO, Fernando Antonio. Transparência sobre o emprego de Inteligência Artificial no Judiciário: um modelo de governança. **Suprema-Revista de Estudos Constitucionais**, v. 3, n. 2, p. 145-187, 2023. Disponível em: <https://suprema.stf.jus.br/index.php/suprema/article/view/231>. Acesso em: 26 maio. 2024.

MARTINS, T. G. S.; SCHOR, P. Desembalando a caixa preta. **Revista Einstein**, v. 19. 2021. DOI: https://doi.org/10.31744/einstein_journal/2021ED6037. Disponível em: <https://www.scielo.br/j/eins/a/Y3XQ5tqdTTV66HsXJFgXt9v/?format=pdf&lang=pt>. Acesso em: 25 fev. 2024.

MATIAS, E. A.; MOUTA ARAÚJO, J. H. Inteligência Artificial e o Direito: uma reflexão sobre as novas tendências, perspectivas e desafios à prática jurídica no Brasil. **Revista De Direito E Atualidades**, v. 2, n. 5, 2023.

MCGINNIS, John O.; PEARCE, Russell G. The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services, **Northwestern Public Law Research Paper**, New York, v. 17, n. 14, p.1-26, 2014.

MEIRELES, A. V. Algoritmos e autonomia: relações de poder e resistência no capitalismo de vigilância. **Opinião Pública**, v. 27, n. 1, 2021. Disponível em: <https://www.scielo.br/j/op/a/vryT7RHQCQ8q8RvYXF3zKvZS/?format=pdf&lang=pt>. Acesso em: 26 mar. 2025.

MENDES, Laura Schertel; MATTIUZZO, Marcela. Discriminação Algorítmica: Conceito, Fundamento Legal e Tipologia. **RDU**, Porto Alegre, Volume 16, n. 90, 2019, 39-64, nov-dez 2019.

MENDONÇA JUNIOR, C.D.N.; NUNES, D. Desafios e Oportunidades para a Regulação da Inteligência Artificial :A necessidade de Compreensão e Mitigação dos Riscos da IA. **Revista Contemporânea**, 2023. Disponível em: <<https://ojs.revistacontemporanea.com/ojs/index.php/home/article/view/1146>> Acesso em: 01. Out.2023.

MENDONÇA, R. F.; FILGUEIRAS, F.; ALMEIDA, V. Algoritmos, desidentificação e infrapolítica da resistência. **Revista Brasileira de Ciência Política**, v. 44, p. 1-31, e280252, 2025. Disponível em: <https://www.scielo.br/j/rbcpol/a/9dHmHNSzsGWt4CTb3nthVTd/>. Acesso em: 31 mar. 2025

MILLER, Tim. Explanation in Artificial Intelligence: Insights from the Social

Sciences. **Artificial Intelligence**, v. 267, p. 1-38, 2019. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0004370218305988>. Acesso em: 30 mar. 2025.

MITCHELL, M. *et al.* **Diversity and inclusion metrics in subset selection**. In: **Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society**. ACM, New York, pp 117–123. 2021.

MOREIRA, A. O. S. Sobre o direito e a inteligência artificial (e robótica). **Oscorp**, 2017.

NASCIMENTO, S. M. *et al.* Inteligência artificial e suas implicações éticas e legais: revisão integrativa. **Revista Bioética**, Brasília, v. 32, p. e3729, 2024. Disponível em: <https://www.scielo.br/j/bioet/a/NjRmBYTfwTy9HFQPf7dm8Ny/>. Acesso em: 04 abr. 2025.

NOBLE, S. U. **Algorithms of Oppression: How Search Engines Reinforce Racism**. New York: NYU Press, 2018.

NOVAIS, P.; FREITAS, P. M. **Inteligência Artificial e Regulação de algoritmos**. Diálogos União Europeia Brasil. Universidade do Minho. 2018. Disponível em: https://eubrdialogues.com/documentos/noticias/adjuntos/ef9c1b_Intelig%C3%A2ncia%20Artificial%20e%20Regula%C3%A7%C3%A3o%20de%20Algoritmos.pdf. Acesso em: 19 fev. 2024.

OBERMEYER, Z. *et al.* Dissecting racial bias in an algorithm used to manage the health of populations. **Science**, v. 366, n. 6464, p. 447-453, 2019. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/31649194/>. Acesso em: 30 mar. 2025.

ORTEGA, Pepita. **Como o Judiciário brasileiro está usando IA para reduzir a burocracia e agilizar processos**. 2023. ESTADÃO. Disponível em: <https://www.estadao.com.br/politica/blog-do-fausto-macedo/ia-ja-ajuda-a-reduzirtempo-e-agilizar-processos-no-judiciario/>. Acesso em: 18 jun. 2024.

PADHY, A. K.; PADHY, A. K. Criminal liability of the artificial intelligence entities. Nirma University **Law Journal**: Volume-8, Issue-2, July-2019. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3443002. Acesso em: 18 fev. 2024.

PÁDUA, Sérgio Rodrigo de; LORENZETTO, Bruno Meneses. O direito fundamental à explicabilidade da inteligência artificial utilizada em decisões estatais. **Revista da AGU**, Brasília-DF, v. 23, n. 02, p. 350-371, jun. 2024. Disponível em: <https://doi.org/10.25109/2525-328X.v.23.n.02.2024>. Acesso em: 04 abr. 2025.

PALETTA, Francisco Carlos; SILVA, Armando Manuel Barreiro Malheiro da. Dado, informação e inteligência artificial: contribuições para a ciência da informação. I **Simpósio Informação, Ciência, Tecnologia e Inovação**, 2020.

PECEGO, Daniel Nunes; TEIXEIRA, Raphael Lobato Collet Janny. Inteligência

artificial no judiciário: da opacidade à explicabilidade das decisões judiciais. **RFD – Revista da Faculdade de Direito da UERJ**, n. 43, 2024.

PEIXOTO, Fabiano Hartmann. Inteligência Artificial e Direito: Convergência Ética e Estratégica. Coleção Direito, Racionalidade e Inteligência Artificial – volume 5. Curitiba: Alteridade, edição Kindle, 2020

PEREIRA, Alexandre Libório Dias. Ius ex machina? Da informática jurídica ao computador-juiz. **Revista Jurídica Luso-Brasileira**, Lisboa, v. 3, n. 1, p. 43-126, 2017.

POSSA, Alisson; AZEVEDO, Rebeca. **A utilização de IA generativa pelo Judiciário brasileiro: a crônica de uma tragédia anunciada do julgador que objetivou a celeridade e feriu uma garantia fundamental**. 2023. Migalhas. Disponível em: <https://www.migalhas.com.br/coluna/dados-publicos/397564/a-utilizacao-de-iagenerativa-pelo-judiciario-brasileiro>. Acesso em: 18 jun. 2024.

REQUIÃO, Maurício; COSTA, Diego Carneiro. Discriminação algorítmica: ações afirmativas como estratégia de combate. **Civilistica.com.**, a. 11, n. 3, 2022.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In: **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**, 2016. Disponível em: <https://dl.acm.org/doi/10.1145/2939672.2939778>. Acesso em: 01 abr. 2025.

ROCHA, I. F.; KISSIMOTO, K. O. Barreiras e benefícios na adoção de inteligência artificial e IoT na gestão da operação. **RAM**, São Paulo, v. 23, n. 4, 2022. Disponível em: <https://www.scielo.br/j/ram/a/mGpm3mhb5vZ5VLPbmmfYBwt/?format=pdf&lang=pt>. Acesso em: 10 mar. 2025.

RODRIGUES, Fernanda; LOPES, Lucas. **IA no Judiciário**: da curadoria de dados e pesquisas independentes à regulação. Disponível em: <<https://www.jota.info/opiniao-e-analise/colunas/ia-regulacao-democracia/ia-no-judiciario-da-curadoria-de-dados-e-pesquisas-independentes-a-regulacao>>. Acesso em: 01 abr. 2025.

ROSSETTI, R.; ANGELUCI, A. Ética Algorítmica: questões e desafios éticos do avanço tecnológico da sociedade da informação. **Galáxia**, v. 46, 2021. Disponível em: <https://www.scielo.br/j/gal/a/R9F45HyqFZMpQp9BGTfZnyr#>. Acesso em: 14 mar. 2025.

ROVER, A. J. A tecnologia como fator de democratização do Direito. **Seqüência Estudos Jurídicos e Políticos**, Florianópolis, v. 18, n. 35, p. 50–55, 1997.

RUSSELL, S. J.; NORVIG, P. **Inteligência artificial**. tradução Regina Célia Simille. Rio de Janeiro: Elsevier, 2013.

SAINZ, N.; GABARDO, E.; ONGARATTO, N. Discriminação algorítmica no Brasil: uma análise da pesquisa jurídica e suas perspectivas para a compreensão do fenômeno. **Revista Direito Público**, v. 21, n. 110, 2024. Disponível em: [SALOMAO, Luis Felipe. **Inteligência artificial: tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário brasileiro**. 2. ed. Rio de Janeiro: FGV Conhecimento Centro de Inovação, Administração e Pesquisa do Judiciário, 2022. Disponível em: \[https://ciapj.fgv.br/sites/ciapj.fgv.br/files/relatorio_ia_2fase.pdf\]\(https://ciapj.fgv.br/sites/ciapj.fgv.br/files/relatorio_ia_2fase.pdf\). Acesso em: 23 abr. 2024.](https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/download/7295/3391/27759#:~:text=As%20decis%C3%B5es%20humanas%20sempre%20foram,de%20Fortes%20(2020%2C%20p. Acesso em: 14 mar. 2025.</p></div><div data-bbox=)

SANTOS, A. C. Variações conceituais entre a ética e a moral. **Revista Filosofia Unisinos**, v. 22, n. 2, p. 1-10. 2021. DOI: <https://doi.org/10.4013/fsu.2021.222.07>. Disponível em: <https://www.scielo.br/j/fun/a/L373KJHwbW4TcWk6Lcjyfvj/?format=pdf&lang=pt>. Acesso em 21 fev. 2024.

SANTOS, Luiza Carolina dos. Inteligência Artificial conversacional e o paradigma simulativo: pistas antropomórficas nas assistentes digitais. In: **Anais do 31º Encontro Anual da Compós**, Imperatriz (MA), 2022. Disponível em: <https://proceedings.science/compos/compos-2022/trabalhos/inteligencia-artificialconversacional-e-o-paradigma-simulativo-pistas-antropomo?lang=pt-br>. Acesso em: 02 abr. 2024.

SAP (System Applications and Products in Data Processing). **O que é IA generativa?** 2023. Disponível em: <https://www.sap.com/brazil/products/artificial-intelligence/what-is-generative-ai.html>. Acesso em: 02 abr. 2024.

SARTOR, Giovanni; BRANTING, L. Karl. Introduction: judicial applications of artificial intelligence. In: **Judicial applications of artificial intelligence**. Dordrecht: Springer Netherlands, 1998. p. 1-6. Disponível em: <https://bit.ly/2G6zyQz>. Acesso em: 24 abr. 24.

SAYÃO, L. F.; SALES, L. F. Curadoria digital e dados de pesquisa. **AtoZ: Novas práticas Em informação E Conhecimento**, v. 5, n. 2, 2016. Disponível em: <https://revistas.ufpr.br/atoz/article/view/49708>. Acesso em: 03 abr. 2025.

SCHIMIDT, Eduardo Moreira; AGUIAR, Lucas Lima; RODRIGUES, Paulo Henrique Andrade; NASCIMENTO, Marcio de Jesus Lima do. Os impactos do uso da inteligência artificial pelo Poder Judiciário: análise das publicações dos últimos anos. **Revista Ibero-Americana de Humanidades, Ciências e Educação**. São Paulo, v. 11, n.3, mar. 2025. ISSN: 2675-3375.

SCHWAB, Klaus, **A quarta revolução industrial**, trad. Daniel Moreira Miranda, São Paulo: Edipro, 2016.

SCHWARTZ, Hamilton Rafael Marins. Inteligência Artificial na judicialização da saúde: marcos legais e ferramentas tecnológicas. **Gralha Azul: Periódico Científico da EJUD/PR**, Curitiba, v. 1, n. 20, p. 20-31, nov. 2023. Disponível em: <https://ejud.tjpr.jus.br/documents/d/ejud/2-hamilton-rafael-marins-schwartz-2>. Acesso em: 02 maio 2024.

SENADO. **Projeto de Lei nº 2338, de 2023**. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>. Acesso em: 14 mar. 2025.

SICHMAN, J. S. Inteligência Artificial e sociedade: avanços e riscos. **Revista Estudos Avançados**, v. 35, n. 101, p. 37-49. 2021. DOI: <https://doi.org/10.1590/s0103-4014.2021.35101.004>. Disponível em: <https://www.scielo.br/j/ea/a/c4sqqrthGMS3ngdBhGWtKhh/?format=pdf&lang=pt>. Acesso em 26 fev. 2024.

SILVA, Alessandro Azevedo de Sousa; BERNARDES, Rochele Juliane Lima Firmeza. Desafios e perspectivas da inteligência artificial no sistema judiciário brasileiro. **Revista Contemporânea**, v. 3, n. 12, 2023. ISSN 2447-0961.

SILVA, A. D. F. **Processo judicial eletrônico e a informática jurídica: um olhar para o uso da inteligência artificial como ferramenta de eficiência na prestação jurisdicional**. 2017. Dissertação (Mestrado em Direito) Universidade Nove de Julho – UNINOVE. São Paulo, 2017.

SILVA, G. L.; SOUZA, L. S.; SOUZA, I. S. M. Desafios e riscos de utilização (in)adequada da inteligência artificial como ferramenta de resolução de demandas no sistema judiciário. **Revista FT**, v. 28, 2024. Disponível em: <https://revistaft.com.br/desafios-e-riscos-de-utilizacao-inadequada-da-inteligencia-artificial-como-ferramenta-de-resolucao-de-demandas-no-sistema-judiciario/>. Acesso em: 15 mar. 2025.

SILVA, Ricardo Augusto Ferreira e; FILHO, Antônio Isidro da Silva. **Inteligência Artificial nos tribunais brasileiros: retórica ou realidade?**. Administration of Justice Meeting. CWB, Jun, 2020. Disponível em: <http://www.enajus.org.br/anais/2020/inteligencia-artificial-em-tribunais-brasileiros-retorica-ou-realidade>. Acesso em 01 mai. 2024

SIMÕES, Nathália Chagas; MORAIS, Laislla Ferreira. As reflexões da inteligência artificial no Poder Judiciário e a sua efetividade. Race Interdisciplinar. **Revista Científica Eletrônica**, 2024.

SIMÕES-GOMES, Letícia; ROBERTO, Enrico; MENDONÇA, Jônatas. Viés Algorítmico – Um Balanço Provisório. **Estud. sociol.** Araraquara, v. 25, n. 48, p. 139-166, jan.-jun., 2020.

SLEEMAN, D.; BERNSEN, O. **Artificial Intelligence**. London and New York: Routledge. 2020.

SOUSA, Guilherme Castro. **SARA, a inteligência artificial generativa feita para uso em tribunais**. 2024. Jornal da USP. Disponível em: <https://jornal.usp.br/atualidades/sara-a-inteligenciaartificial-generativa-feita-para-uso-em-tribunais/>. Acesso em: 02 maio 2024.

SOUSA, S. A.; FIDALGO, S.; COSTA, M. J. A.; PEREIRA, A. L. D.; RODRIGUES, A. M. M. P. M. **Inteligência Artificial. Desafios Societais e a Investigação em Direito – 04**. 2022. Coimbra: Instituto Jurídico da Faculdade de Direito da Universidade de Coimbra. Disponível em: <https://estudogeral.uc.pt/handle/10316/103984>. Acesso em: 15 fev. 2024.

SOUZA, R. M.; OLIVEIRA, F. D.; PINTO, P.E. D. Uma Nova Proposta para a Obtenção da Complexidade de Pior Caso do ShellSort. **Revista TEMA**, v. 20, n. 3, p. 457-476, 2019. <https://doi.org/10.5540/tema.2019.020.03.0457>. Acesso em: 24 fev. 2024.

SPAULDING, Norma W. Is human judgment necessary? In: DUBBER, Markus D.; PASQUALE, Frank; DAS, Sunit (org.). **The Oxford handbook of ethics of AI**. New York: Oxford University Press, 2020. p. 375-402.

SPERANDIO, H. R. C. **Desafios da inteligência artificial para a profissão jurídica**. 2018. Dissertação (Mestrado em Direito) Fundação Getúlio Vargas, São Paulo, 2018. Disponível em: <https://repositorio.fgv.br/items/97c7f566-43b0-45ad-86e2-a161df1b2de3>. Acesso em: 14 mar. 2025.

SPIEGELEIRE, S.; et al. **“What is Artificial Intelligence?” Artificial Intelligence and the future of defense: strategic implications for small- and medium-sized force providers**, Hague Centre for Strategic Studies, 2017, p. 25–42. JSTOR, Disponível em: <http://www.jstor.org/stable/resrep12564.7>. Acesso em 20 fev. 2024.

STAATS, Sabrina; MORAIS, Fausto Santos de. A Utilização de Inteligência Artificial no Poder Judiciário Brasileiro como exercício de sustentabilidade. **Revista Ilustração**, v. 1, n. 2, p. 7-18, 2020.

SULEIMAN, N.; SILVA, N. F. A utilização da inteligência artificial no sistema judiciário brasileiro: análise dos benefícios e desafios. **Revista Interciência – IMES**, Catanduva, v. 1, n. 12, p. 12-19, dez. 2023. Disponível em: <https://fafica.br/revista/index.php/interciencia/article/view/502>. Acesso em: 29 mar. 2025.

TEIXEIRA, Raoni Florentino da Silva. **Introdução a algoritmos**. Mato Grosso: UFMT, 2019.

TOLEDO, C.; PESSOA, D. O uso de inteligência artificial na tomada de decisão judicial. **Rev. Investig. Const.**, Curitiba, vol. 10, n. 1, e237, jan./abr. 2023. Disponível em: <https://www.scielo.br/j/rinc/a/qRC4TmVXVDJ8Wkv7Ns49jxH/?format=pdf&lang=pt>. Acesso em: 13 mar. 2025.

TOMASEVICIUS FILHO, E. Inteligência artificial e direitos da personalidade: uma

contradição em termos? **R. Fac. Dir. Univ.** São Paulo, v. 113, 2018.

TRAVAIN, Luiz Antonio Loureiro. Inteligência artificial e jurimetria no Poder Judiciário. **Revista Acadêmica Online**, Brazil, v.11, n.55, p. 01-29, 2025.

UFSM. **Viés Algorítmico**. Disponível em: <<https://www.ufsm.br/pet/sistemas-de-informacao/2023/11/28/vies-algoritmico>>. Acesso em: 24 mar. 2025.

TRIBUNAL DE JUSTIÇA DO ESTADO DO PARANÁ (TJPR). *TJPR lança ferramenta de IA para pesquisa jurisprudencial*. Curitiba, 2024. Disponível em: <https://www.tjpr.jus.br>. Acesso em: 26 mar. 2025.

TRIBUNAL DE JUSTIÇA DE MINAS GERAIS. TJMG apresenta SAVIA, nova ferramenta de inteligência artificial baseada no chatGPT. **TJMG**, 2023. Disponível em:< TJMG apresenta SAVIA, nova ferramenta de inteligência artificial baseada no ChatGPT | Portal TJMG>. Acesso em 20 de out. de 2023.

TRIBUNAL REGIONAL FEDERAL (3ª Região). Laboratório de Inteligência Artificial aplicada. **TRF3**, 2020. Disponível em: <<https://www.trf3.jus.br/adeq/governanca-e-estrategia/laboratorio-de-inovacao-do-trf3a-regiao/laboratorio-de-inteligencia-artificial-aplicada-liaa-3r>>. Acesso em 11 de set. de 2023.

VALENTINI, R. S. **Julgamento por computadores? as novas possibilidades da juscibernética no século XXI e suas implicações para o futuro do direito e do trabalho dos juristas**. 2017. Tese (Doutorado em Direito) Universidade Federal de Minas Gerais, 2017.

VALLE, V. C. L.; GASÓ, J. R. F.; AJUS, A. M. Decisão judicial assistida por inteligência artificial e o Sistema Víctor do Supremo Tribunal Federal. **Rev. Investig. Const.**, Curitiba, vol. 10, n. 2, e252, maio/ago. 2023. Disponível em: <https://www.scielo.br/j/rinc/a/YKZfQPLJqT7F3P445KkmwnC/?lang=pt>. Acesso em: 19 mar. 2025.

VILAÇA, M. L. C.; ARAÚJO, E. V. F. **Tecnologia, Sociedade e Educação na Era Digital**. Duque de Caxias, RJ: UNIGRANRIO, 2016.

ZARSKY, T. **The Trouble with Algorithmic Decisions**: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. Science, Technology, & Human Values, 2016.

ZHANG, L. **DeepSeek-R1: A New Era in Cost-Effective AI**. White Paper DeepSeek, 2025.

WILKINSON, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, London, v. 3, n. 1, p. 1-9, mar. 2016. Disponível em: <https://www.nature.com/articles/sdata201618#citeas>. Acesso em: 02 abr. 2025.