

UNIVERSIDADE ESTADUAL DE PONTA GROSSA
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS SOCIAIS APLICADAS

MELINA CARLA DE SOUZA BRITTO

INTELIGÊNCIA ARTIFICIAL E GÊNERO NO PODER JUDICIÁRIO
BRASILEIRO: DESAFIOS ÉTICOS E DIRETRIZES PARA UMA REGULAÇÃO
FEMINISTA E INCLUSIVA

PONTA GROSSA

2025

MELINA CARLA DE SOUZA BRITTO

**INTELIGÊNCIA ARTIFICIAL E GÊNERO NO PODER JUDICIÁRIO
BRASILEIRO: DESAFIOS ÉTICOS E DIRETRIZES PARA UMA REGULAÇÃO
FEMINISTA E INCLUSIVA**

Tese apresentada como requisito para obtenção do título de Doutora em Ciências Sociais Aplicadas, pelo Programa de Pós-Graduação em Ciências Sociais Aplicadas da Universidade Estadual de Ponta Grossa (UEPG). Área de concentração: Cidadania e Políticas Públicas. Linha de Pesquisa: História, cultura e cidadania.

Orientador: Prof. Dr. Fabrício Bittencourt da Cruz.
Coorientador: Prof. Dr. Flávio Bortolozzi.

PONTA GROSSA

2025

B862 Britto, Melina Carla de Souza
Inteligência artificial e gênero no Poder Judiciário brasileiro: desafios éticos e diretrizes para uma regulação feminista e inclusiva / Melina Carla de Souza Britto. Ponta Grossa, 2025.
302 f.

Tese (Doutorado em Ciências Sociais Aplicadas - Área de Concentração: Cidadania e Políticas Públicas), Universidade Estadual de Ponta Grossa.

Orientador: Prof. Dr. Fabrício Bittencourt da Cruz.
Coorientador: Prof. Dr. Flávio Bortolozzi.

1. Inteligência artificial. 2. Poder Judiciário. 3. Viés algorítmico. 4. Viés de gênero. 5. Technofeminismo. I. Cruz, Fabrício Bittencourt da. II. Bortolozzi, Flávio. III. Universidade Estadual de Ponta Grossa. Cidadania e Políticas Públicas. IV.T.

CDD: 340

TERMO DE APROVAÇÃO

MELINA CARLA DE SOUZA BRITTO

"Inteligência Artificial e gênero no Poder Judiciário brasileiro: desafios éticos e diretrizes para uma regulação feminista e inclusiva".

Tese aprovada como requisito parcial para obtenção do grau de Doutor(a) no Programa de Pós-Graduação em Ciências Sociais Aplicadas, Setor de Ciências Sociais Aplicadas da Universidade Estadual de Ponta Grossa, pela seguinte banca examinadora:

Documento assinado digitalmente
 **FABRICIO BITTENCOURT DA CRUZ**
Data: 08/12/2025 18:06:16-0300
Verifique em <https://validar.itl.gov.br>

Prof. Dr. Fabrício Bittencourt da Cruz - UEPG-PR - Presidente

Documento assinado digitalmente
 **FLAVIO BORTOLOZZI**
Data: 09/12/2025 13:10:16-0300
Verifique em <https://validar.itl.gov.br>

Prof. Dr. Flavio Bortolozzi - Coorientador

CINTHIA OBLADEN DE
ALMENDRA
FREITAS:59256931915

Assinado de forma digital por CINTHIA
OBLADEN DE ALMENDRA
FREITAS:59256931915
Dados: 2025.12.10 12:02:05 -03'00'

Profa. Dra. Cinthia Obladen de Almendra Freitas - PUC-PR - Membro Externo

Documento assinado digitalmente
 **TAIS SCHILLING FERRAZ**
Data: 09/12/2025 13:34:56-0300
Verifique em <https://validar.itl.gov.br>

Profa. Dra. Taís Schilling Ferraz - Escola Nac. de Form. e Aperf. de Magistrados -
Membro Externo

Documento assinado digitalmente
 **ANTONIO CESAR BOCHENEK**
Data: 09/12/2025 17:34:43-0300
Verifique em <https://validar.itl.gov.br>

Prof. Dr. Antônio César Bochenek - UEPG-PR - Membro Interno

Documento assinado digitalmente
 **RESHAD TAWFEIQ**
Data: 09/12/2025 17:54:51-0300
Verifique em <https://validar.itl.gov.br>

Prof. Reshad Tawfeiq - UEPG-PR - Membro Interno

Ponta Grossa, 08 de dezembro de 2025.

Break down
Every door they built
To keep you out
And bring all your people with you

- *storm*

(rupi kaur – home body)

Dedico esta conquista à minha família.

AGRADECIMENTOS

A Deus, pela dádiva da vida, pela força e sabedoria concedidas em cada etapa desta jornada. Agradeço por iluminar meus caminhos e me sustentar nos momentos de incerteza.

À minha família, meu alicerce e maior fonte de amor e inspiração. Em especial: aos meus pais e irmãs, obrigada pela paciência, incentivo e compreensão nos momentos em que precisei me dedicar intensamente a este trabalho; e ao meu namorado, companheiro de todos os dias, que está sempre ao meu lado, compartilhando sonhos, desafios e conquistas. A presença de vocês, apoio e amor foram fundamentais para que eu chegasse até aqui.

Ao meu orientador, professor Dr. Fabrício Bittencourt da Cruz, pela orientação competente, pelas valiosas contribuições e pela dedicação durante todo o processo. Sua experiência e disponibilidade foram essenciais para o desenvolvimento deste trabalho.

Ao meu coorientador, professor Dr. Flávio Bortolozzi, pela atenção, pelas sugestões precisas e pelo incentivo constante, que enriqueceram significativamente esta pesquisa.

A todos que, de alguma forma, contribuíram para esta conquista, meu sincero agradecimento.

RESUMO

BRITTO, Melina Carla de Souza. **Inteligência artificial e gênero no Poder Judiciário brasileiro**: desafios éticos e diretrizes para uma regulação feminista e inclusiva. Orientador: Fabrício Bittencourt da Cruz. Coorientador: Flávio Bortolozzi. Ponta Grossa, 2025, 302f. Tese (Doutorado em Ciências Sociais Aplicadas) – Universidade Estadual de Ponta Grossa, Ponta Grossa, 2025.

A inteligência artificial (IA) consolida-se como uma das tecnologias mais transformadoras do século XXI, impactando profundamente as formas de produção do conhecimento, as relações sociais e as estruturas institucionais. No campo das ciências sociais aplicadas, sua relevância é ainda mais expressiva, sobretudo quando se considera sua incorporação em sistemas de poder e decisão, como o Poder Judiciário. A crescente utilização de algoritmos e sistemas automatizados para triagem, classificação e apoio à decisão judicial tem sido justificada pelo discurso da eficiência e da racionalidade técnica. Contudo, a literatura contemporânea e experiências empíricas recentes revelam que tais tecnologias não são neutras: elas podem reproduzir e ampliar desigualdades pré-existentes, especialmente as de gênero, raça e classe. Diante desse contexto, esta tese investiga de que maneira a IA aplicada ao Poder Judiciário brasileiro pode reproduzir ou intensificar discriminações de gênero e em que medida o Direito pode atuar de forma eficaz na prevenção desses vieses e na reorientação da tecnologia para promover a ética e a justiça social. Parte-se da hipótese de que o Direito pode funcionar como instrumento regulatório eficaz, desde que articule simultaneamente garantias técnicas (p.e., *fairness*, transparência, explicabilidade e auditabilidade), garantias regulatórias (p.e., governança de dados, direitos humanos e princípios constitucionais relacionados ao devido processo legal) e garantias políticas (p.e., conscientização, formação crítica e letramento tecnológico), assegurando que a tecnologia seja orientada para promover justiça social e inclusão. O objetivo geral é propor protocolos e diretrizes que assegurem o uso justo, transparente e inclusivo da IA no Judiciário brasileiro, a partir da identificação dos vieses de gênero que esses sistemas podem reproduzir ou intensificar e da análise do papel do Direito como instrumento regulatório capaz de preveni-los. A pesquisa adota abordagem qualitativa e interpretativa, de natureza exploratória e aplicada, fundamentada em estudo de caso sobre o Poder Judiciário brasileiro. Foram utilizadas três estratégias metodológicas principais: (a) levantamento bibliográfico em bases acadêmicas nacionais e internacionais; (b) análise documental de atos normativos, resoluções e relatórios do Conselho Nacional de Justiça (CNJ) e de tribunais; e (c) análise de conteúdo, segundo Bardin (2016), aplicada a documentos institucionais, jurisprudenciais e técnicos. O recorte temporal compreende o período de 2018 a 2025, correspondente à consolidação das políticas de transformação digital e dos primeiros projetos de IA no sistema judicial. Os resultados evidenciam que a IA, ao ser implementada em um Judiciário ainda marcado por assimetrias de gênero e por uma cultura jurídica predominantemente masculina, tende a reproduzir desigualdades estruturais. Contudo, a pesquisa demonstra que o Direito pode atuar como instrumento de mitigação desses vieses, especialmente quando orientado por parâmetros feministas e interseccionais, promovendo a criação de instrumentos técnicos, regulatórios, pedagógicos e políticos voltados à justiça algorítmica. Nesse sentido, apresentam-se dois protocolos éticos que materializam e operacionalizam as três dimensões de garantias propostas na tese: um protocolo para desenvolvedores e equipes técnicas, que traduz principalmente as garantias técnicas de planejamento ético, governança de dados, testes de equidade, explicabilidade e auditoria contínua; e um protocolo para magistrados, assessorias e usuários, que concretiza as garantias políticas e regulatórias por meio do uso responsável e transparente da IA, com julgamento com

perspectiva de gênero e preservação da autonomia decisória humana. Como contribuição teórica e prática, a tese propõe um modelo de governança algorítmica feminista e inclusiva da IA no sistema de justiça, estruturado nos três ramos de garantias éticas interdependentes – técnicas, regulatórias e políticas – e exemplificado por meio dos protocolos apresentados. Esses instrumentos constituem a principal contribuição aplicada do trabalho, pois oferecem referências operacionais e verificáveis para orientar o desenvolvimento, o uso e a fiscalização de sistemas inteligentes no Poder Judiciário. A pesquisa, assim, fornece uma base conceitual e normativa capaz de sustentar políticas públicas e práticas institucionais comprometidas com a equidade, a ética e a justiça social no uso de tecnologias judiciais.

PALAVRAS-CHAVE: Inteligência artificial; Poder Judiciário; Viés algorítmico; Viés de gênero; Technofeminismo; Garantias éticas.

ABSTRACT

BRITTO, Melina Carla de Souza. **Artificial Intelligence and Gender in the Brazilian Judiciary: Ethical Challenges and Guidelines for a Feminist and Inclusive Regulation.** Advisor: Fabrício Bittencourt da Cruz; Co-advisor: Flávio Bortolozzi. Ponta Grossa, 2025, 302p. Tesis (Doctorate in Applied Social Sciences) – Universidade Estadual de Ponta Grossa, Ponta Grossa, 2025.

Artificial Intelligence (AI) has consolidated itself as one of the most transformative technologies of the twenty-first century, profoundly reshaping the production of knowledge, social relations, and institutional structures. Within the field of applied social sciences, its relevance is particularly pronounced when considering its incorporation into systems of power and decision-making, such as the Judiciary. The increasing use of algorithms and automated systems for case screening, classification, and judicial decision support has often been justified through the discourse of efficiency and technical rationality. However, contemporary academic studies and recent empirical evidence reveal that such technologies are not neutral; rather, they may reproduce and exacerbate pre-existing inequalities, especially those related to gender, race, and class. In this context, the present doctoral dissertation investigates how artificial intelligence, when applied to the Brazilian Judiciary, may reproduce or intensify gender-based discrimination, and to what extent the law can operate effectively in preventing such biases and reorienting technology toward the promotion of ethics and social justice. The study is grounded in the hypothesis that the law can serve as an effective regulatory instrument, provided that it simultaneously integrates technical safeguards (e.g., fairness, transparency, explainability, and auditability), regulatory safeguards (e.g., data governance, human rights, and constitutional principles of due process), and political safeguards (e.g., awareness, critical education, and gender-sensitive technological literacy). These dimensions are understood as interdependent and essential for ensuring that technology is aligned with principles of social justice and inclusion. The main objective is to propose protocols and guidelines that guarantee the fair, transparent, and inclusive use of AI in the Brazilian Judiciary, based on the identification of gender biases potentially reproduced or amplified by such systems and on the analysis of the law's regulatory capacity to mitigate them. Methodologically, the research adopts a qualitative and interpretative approach, exploratory and applied in nature, structured as a case study of the Brazilian Judiciary. Three principal methodological strategies were employed: (a) a comprehensive literature review in national and international academic databases; (b) documentary analysis of normative acts, resolutions, and reports issued by the Conselho Nacional de Justiça (CNJ) and judicial institutions; and (c) content analysis, as proposed by Bardin (2016), applied to institutional, jurisprudential, and technical documents. The temporal scope spans from 2018 to 2025, a period corresponding to the consolidation of digital transformation policies and the initial implementation of AI projects within the judicial system. The findings indicate that AI, when integrated into a judiciary still characterized by gender asymmetries and a predominantly male legal culture, tends to replicate structural inequalities. Nonetheless, the research demonstrates that the law can function as a mitigating mechanism, particularly when informed by feminist and intersectional perspectives, promoting the development of technical, regulatory, pedagogical, and political instruments aimed at algorithmic justice. Two ethical protocols are proposed to operationalize the three dimensions of safeguards: (1) a protocol for developers and technical teams, translating technical safeguards into practices of ethical planning, data governance, fairness testing, explainability, and continuous auditing; and (2) a protocol for judges, advisors, and users, implementing regulatory and political safeguards through responsible and transparent AI use, gender-sensitive adjudication, and the preservation of human decision-making autonomy. As both a theoretical

and practical contribution, the dissertation proposes a feminist and inclusive model of algorithmic governance for AI within the justice system, structured around interdependent technical, regulatory, and political safeguards and exemplified through the proposed protocols. These instruments constitute the main applied contribution of the study, offering operational and verifiable references for the ethical development, deployment, and oversight of intelligent systems in the Judiciary. Ultimately, this research provides a conceptual and normative framework capable of informing public policies and institutional practices committed to equity, ethics, and social justice in the use of judicial technologies.

KEYWORDS: Artificial intelligence; Judiciary; Algorithmic bias; Gender bias; Technofeminism; Ethical safeguards.

LISTA DE FIGURAS

Figura 1. Mapa mental sobre viés nos dados de treinamento.	70
Figura 2. Mapa mental sobre viés de modelagem ou viés algorítmico.	71
Figura 3. Mapa mental sobre viés na implementação e nos resultados.	72
Figura 4. Mapa mental sobre viés por interpretação de contextos culturais, políticos ou sociais.	73
Figura 5. Sexo e posição na carreira da magistratura.	94
Figura 6. Percentual de magistradas no Poder Judiciário brasileiro.	95
Figura 7. Identidade de gênero no Poder Judiciário - 2023.	96
Figura 8. Percentuais de juízes (homens e mulheres) no 1º grau de jurisdição no Brasil em 2022 – Parte 1.	98
Figura 9. Percentuais de juízes (homens e mulheres) no 1º grau de jurisdição no Brasil em 2022 – Parte 2.	99
Figura 10. Percentuais de desembargadores e ministros (homens e mulheres) no Poder Judiciário brasileiro em 2022 – Parte 1.	100
Figura 11. Percentuais de desembargadores e ministros (homens e mulheres) no Poder Judiciário brasileiro em 2022 – Parte 2.	101
Figura 12. Comparação de gênero na composição geral do Poder Judiciário no ano-base 2024.	107
Figura 13. Comparação de gênero de ministros nos Tribunais Superiores e de desembargadores nos demais ramos de justiça.	107
Figura 14. Comparação de gênero no primeiro grau, por ramo de justiça.	107
Figura 15. Distribuição de magistrados por gênero em outubro de 2025.	108
Figura 16. Distribuição de magistrados por gênero e por ramo de atuação em outubro de 2025.	109
Figura 17. Série histórica de magistradas no Poder Judiciário brasileiro, entre dezembro de 2023 e junho de 2025.	110
Figura 18. Números de iniciativas de uso de IA no setor jurídico.	137
Figura 19. Quantidade de iniciativas de IA por tribunal e situação.	138
Figura 20. Abordagens de AI nos tribunais - I.	138
Figura 21. Abordagens de AI nos tribunais - II.	139
Figura 22. Distribuição de projetos de IA no Poder Judiciário.	140
Figura 23. Tipos de algoritmos usados.	142

Figura 24. Tipos de aprendizado de máquina usados.....	142
Figura 25. IAGs utilizadas por magistrados e servidores do Poder Judiciário.....	145
Figura 26. Tipos de uso de ferramentas de IAG por magistrados e servidores.....	146
Figura 27. Principais limitações apontadas pelos respondentes quanto ao uso de IAGs.	146
Figura 28. Distribuição das ferramentas de IA até março de 2025.	148
Figura 29. Tipos de tarefas aplicadas nos projetos de IA no Judiciário em 2024.	150
Figura 30. Projetos de IA por ramo de Justiça.	151
Figura 31. Congestionamento e índice de atendimento à demanda no Poder Judiciário.	158
Figura 32. Classificação dos riscos das IAs no Sinapse em 2023.	168
Figura 33. Princípios da Declaração de Montreal para o Desenvolvimento Responsável da Inteligência Artificial.....	223
Figura 34. Panorama dos princípios sobre IA da OCDE.....	227
Figura 35. Eixos temáticos da EBIA.	229

LISTA DE QUADROS

Quadro 1. Definições de inteligência artificial.....	38
Quadro 2. Taxonomia evolutiva do viés cognitivo.....	52
Quadro 3. A relação entre vieses e heurísticas.	54
Quadro 4. Exemplos de incorporação de viés em sistemas de IA.....	63
Quadro 5. Categorias de viés em sistemas computacionais.	66
Quadro 6. Garantias técnicas para uma IA ética, confiável, segura e alinhada aos direitos humanos e à justiça social.	215
Quadro 7. Princípios-guia para o desenvolvimento e uso ético de IA, segundo García.	224
Quadro 8. Pontos fortes e limites da Resolução CNJ n. 615/2025.....	239
Quadro 9. Razões para a incorporação do Protocolo para Julgamento com Perspectiva de Gênero na IA aplicada dentro do Poder Judiciário.....	244
Quadro 10. Garantias regulatórias para uma IA ética, responsável e alinhada à proteção dos direitos humanos e princípios jurídicos fundamentais.	246
Quadro 11. Garantias políticas para uma IA ética, inclusiva e alinhada à sustentabilidade socioambiental, protegendo direitos humanos fundamentais.....	255

LISTA DE ABREVIATURAS E SIGLAS

ApoIA	Assistente Pessoal Operada por Inteligência Artificial
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
CEDAW	Convenção sobre a Eliminação de Todas as Formas de Discriminação contra as Mulheres
CNJ	Conselho Nacional de Justiça
COMPAS	Correctional Offender Management Profiling for Alternative Sanctions
DPJ	Departamento de Pesquisas Judiciárias
ECOSOC	Conselho Econômico e Social das Nações Unidas
ENFAM	Escola Nacional de Formação e Aperfeiçoamento de Magistrados
FONAVIM	Fórum Nacional de Enfrentamento à Violência Contra a Mulher
FPR	Taxa de falsos positivos
IA	Inteligência artificial
IAG	Inteligência artificial generativa
IRCAI	International Research Centre on Artificial Intelligence under the auspices of UNESCO
KNN	K-nearest neighbors
LGPD	Lei Geral de Proteção de Dados
LIME	Local Interpretable Model-agnostic Explanations
LLM	Large Language Models
OCR	Reconhecimento óptico de caracteres
ODR	Online Dispute Resolution Systems
ODS	Objetivo de desenvolvimento sustentável
ONU	Organização das Nações Unidas
PDPJ-Br	Plataforma Digital do Poder Judiciário Brasileiro
PNUD	Programa das Nações Unidas para o Desenvolvimento
PPGCSA	Programa de Pós-Graduação em Ciências Sociais Aplicadas
PUCPR	Pontifícia Universidade Católica do Paraná
RNN	Redes Neurais Recorrentes (Recurrent Neural Network)
SHAP	SHapley Additive exPlanations
STF	Supremo Tribunal Federal
STJ	Superior Tribunal de Justiça

SVM – Máquina de vetores de suporte

TJAL – Tribunal de Justiça de Alagoas

TJAM – Tribunal de Justiça do Amapá

TJBA – Tribunal de Justiça da Bahia

TJCE – Tribunal de Justiça do Ceará

TJMMG – Tribunal de Justiça Militar de Minas Gerais

TJMRS – Tribunal de Justiça Militar do Rio Grande do Sul

TJMSP – Tribunal de Justiça Militar de São Paulo

TJPA – Tribunal de Justiça do Pará

TJPE – Tribunal de Justiça de Pernambuco

TJPI – Tribunal de Justiça do Piauí

TJRJ – Tribunal de Justiça do Rio de Janeiro

TJRN – Tribunal de Justiça do Rio Grande do Norte

TJRS – Tribunal de Justiça do Rio Grande do Sul

TJTO – Tribunal de Justiça do Tocantins

TPR – Taxa de verdadeiros positivos

TRE-CE – Tribunal Regional Eleitoral do Ceará

TRE-DF – Tribunal Regional Eleitoral do Distrito Federal

TRE-MG – Tribunal Regional Eleitoral de Minas Gerais

TRE-MS – Tribunal Regional Eleitoral do Mato Grosso do Sul

TRE-PE – Tribunal Regional Eleitoral de Pernambuco

TRE-RN – Tribunal Regional Eleitoral do Rio Grande do Norte

TRE-RO – Tribunal Regional Eleitoral de Rondônia

TRE-RR – Tribunal Regional Eleitoral de Roraima

TRE-SC – Tribunal Regional Eleitoral de Santa Catarina

TRE-SE – Tribunal Regional Eleitoral de Sergipe

TRE-SP – Tribunal Regional Eleitoral de São Paulo

TRF1 – Tribunal Regional Federal da 1ª Região

TRF2 – Tribunal Regional Federal da 2ª Região

TRF4 – Tribunal Regional Federal da 4ª Região

TRF5 – Tribunal Regional Federal da 5ª Região

TRT11 – Tribunal Regional do Trabalho da 11ª Região

TRT13 – Tribunal Regional do Trabalho da 13ª Região

- TRT17 – Tribunal Regional do Trabalho da 17ª Região
- TRT19 – Tribunal Regional do Trabalho da 19ª Região
- TRT22 – Tribunal Regional do Trabalho da 22ª Região
- TRT24 – Tribunal Regional do Trabalho da 24ª Região
- TRT5 – Tribunal Regional do Trabalho da 5ª Região
- TST – Tribunal Superior do Trabalho
- UEPG – Universidade Estadual de Ponta Grossa
- UNESCO – Organização das Nações Unidas para a Educação, a Ciência e a Cultura
(United Nations Educational, Scientific and Cultural Organization)

SUMÁRIO

INTRODUÇÃO	20
CAPÍTULO 1 – DA METODOLOGIA DA TESE.....	26
1.1 Abordagem da pesquisa	26
1.2 Tipos de pesquisa	27
<i>1.2.1 Pesquisa exploratória.....</i>	<i>27</i>
<i>1.2.2 Pesquisa descritiva.....</i>	<i>28</i>
1.3 Método de investigação	28
1.4 Universo, amostra e sujeitos da pesquisa	29
1.5 Técnicas de coleta de dados	30
<i>1.5.1 Levantamento bibliográfico.....</i>	<i>31</i>
<i>1.5.2 Análise documental direta e indireta</i>	<i>32</i>
1.6 Procedimento para análise dos dados.....	32
1.7 Aspectos éticos	34
1.8 Relevância do método adotado.....	34
1.9 Limitações metodológicas	35
CAPÍTULO 2 – A INTELIGÊNCIA ARTIFICIAL E O DESAFIO DO VIÉS DE GÊNERO.....	36
2.1 Definindo inteligência artificial e aprendizagem de máquina	38
2.2 Principais técnicas de aprendizagem de máquina e algumas de suas aplicações no campo do Direito.....	43
2.3 O viés humano e da inteligência artificial: a neutralidade existe?	47
<i>2.3.1 O que são e como ocorrem os vieses humanos</i>	<i>49</i>
<i>2.3.2 A inteligência artificial e o mito da neutralidade: vieses nos dados e nos algoritmos... </i>	<i>59</i>
2.4 A inteligência artificial e papel do feminismo no enfrentamento do viés de gênero ..	78
<i>2.4.1. O feminismo, a construção de identidades e os espaços de poder.....</i>	<i>80</i>
<i>2.4.2 A emergência do (tecno)feminismo em contraposição ao viés de gênero na inteligência artificial</i>	<i>82</i>
CAPÍTULO 3 – O PODER JUDICIÁRIO BRASILEIRO E O DESAFIO DO VIÉS DE GÊNERO.....	92
3.1 O perfil sociodemográfico do Poder Judiciário brasileiro.....	93
3.2 O mito da neutralidade do Direito e o impacto da desigualdade de gênero no Judiciário brasileiro: o viés humano-jurídico.....	113

3.3 O Protocolo para Julgamento com Perspectiva de Gênero e o resgate feminista da magistratura no Brasil	128
CAPÍTULO 4 – A IMPLEMENTAÇÃO DA INTELIGÊNCIA ARTIFICIAL NO PODER JUDICIÁRIO BRASILEIRO	135
4.1 O estado da arte da implementação de sistemas inteligentes no Poder Judiciário brasileiro.....	135
4.2 O que podemos esperar no futuro?	152
<i>4.2.1 Os benefícios da implementação da inteligência artificial no Poder Judiciário brasileiro</i>	<i>157</i>
<i>4.2.2 Os riscos e desafios da implementação da inteligência artificial no Poder Judiciário brasileiro</i>	<i>161</i>
CAPÍTULO 5 – GARANTIAS ÉTICAS PARA O ENFRENTAMENTO DOS VIESES DE GÊNERO EM SISTEMAS INTELIGENTES APLICADOS NO JUDICIÁRIO....	170
5.1 Garantias técnicas para uma inteligência artificial ética e feminista no Poder Judiciário brasileiro	173
<i>5.1.1 Métricas de fairness (justiça ou equidade algorítmica).....</i>	<i>174</i>
<i>5.1.2 Transparência dos sistemas inteligentes</i>	<i>182</i>
<i>5.1.3 Explicabilidade dos sistemas inteligentes</i>	<i>187</i>
<i>5.1.4 Diversidade e não discriminação na construção e no emprego dos sistemas inteligentes</i>	<i>193</i>
<i>5.1.5 Design acessível dos sistemas inteligentes.....</i>	<i>196</i>
<i>5.1.6 Acurácia e robusteza dos sistemas inteligentes.....</i>	<i>199</i>
<i>5.1.7 Segurança cibernética e segurança da informação</i>	<i>201</i>
<i>5.1.8 Compreensão dos limites da linguagem natural</i>	<i>206</i>
<i>5.1.9 Personalização e atualização constante dos sistemas inteligentes.....</i>	<i>207</i>
<i>5.1.10 Interpretabilidade e auditabilidade dos sistemas inteligentes e prestação de contas (accountability) quanto ao seu desenvolvimento e uso</i>	<i>210</i>
<i>5.1.11 Quadro resumo das garantias técnicas para uma inteligência artificial ética e feminista no Poder Judiciário brasileiro</i>	<i>215</i>
5.2 Garantias regulatórias para uma inteligência artificial ética e feminista no Poder Judiciário brasileiro	217
<i>5.2.1 O papel dos direitos humanos</i>	<i>218</i>
<i>5.2.2 O papel da regulamentação geral sobre ética e inteligência artificial.....</i>	<i>226</i>

5.2.3 Governança dos sistemas inteligentes.....	231
5.2.4 A Resolução CNJ n. 615/2025.....	237
5.2.5 Regras próprias envolvendo o Poder Judiciário: princípios do devido processo legal, do juiz natural e da fundamentação das decisões judiciais	240
5.2.6 O Protocolo para Julgamento com Perspectiva de Gênero.....	242
5.2.7 Quadro resumo das garantias regulatórias para uma inteligência artificial ética e feminista no Poder Judiciário	246
5.3 Garantias políticas para uma inteligência artificial ética e feminista no Poder Judiciário brasileiro	249
5.3.1 Conscientização e letramento tecnológico.....	249
5.3.2 Bem-estar socioambiental quanto ao desenvolvimento e uso de sistemas inteligentes pelo Poder Judiciário.....	252
5.2.3 Manutenção da conscientização a respeito da disparidade de gênero e a luta constante pela igualdade	254
5.2.4 Quadro resumo das garantias políticas para uma inteligência artificial ética e feminista no Poder Judiciário.....	255
CAPÍTULO 6 – PROTOCOLOS PARA O CUMPRIMENTO DE GARANTIAS ÉTICAS PARA O DESENVOLVIMENTO E O USO DA INTELIGÊNCIA ARTIFICIAL NO PODER JUDICIÁRIO	258
6.1 Protocolo Ético para Desenvolvedores e Equipes Técnicas dos Sistemas Inteligentes Judiciais	258
6.1.1 Planejamento ético e análise de impacto	259
6.1.2 Curadoria e governança de dados	259
6.1.3 Desenvolvimento, testes e validação ética	260
6.1.4 Implementação e monitoramento contínuo	260
6.1.5 Responsabilidade e prestação de contas	261
6.2 Protocolo Ético para Magistrados, Assessorias e Demais Usuários dos Sistemas Inteligentes	261
6.2.1 Compreensão e consentimento informado	261
6.2.2 Uso responsável e transparente	262
6.2.3 Avaliação de equidade e impacto.....	262
6.2.4 Julgamento com perspectiva de gênero.....	262
6.2.5 Supervisão, capacitação e educação ética	263

CONSIDERAÇÕES FINAIS.....	264
REFERÊNCIAS	268
APÊNDICE A – PROTOCOLO ÉTICO PARA DESENVOLVEDORES E EQUIPES TÉCNICAS DOS SISTEMAS INTELIGENTES JUDICIAIS	298
APÊNDICE B – PROTOCOLO ÉTICO PARA MAGISTRADOS, ASSESSORIAS E DEMAIS USUÁRIOS DOS SISTEMAS INTELIGENTES.....	301

INTRODUÇÃO

A inteligência artificial (IA) consolida-se como uma das tecnologias mais transformadoras do século XXI, atravessando diversas áreas do conhecimento e impactando setores estratégicos da sociedade contemporânea. No campo das ciências sociais aplicadas, sua relevância é expressiva, sobretudo quando se analisa a interseção entre tecnologia, direito e questões de gênero.

Sua aplicação em campos decisórios – especialmente no Poder Judiciário – tem sido defendida com base em discursos de eficiência, racionalidade técnica e modernização. Entretanto, a promessa de neutralidade e objetividade dos algoritmos se revela ilusória quando se observam os impactos sociais e éticos decorrentes de sua incorporação em estruturas historicamente marcadas por desigualdades. Ao reproduzirem padrões de comportamento e decisões extraídos de dados e práticas humanas, sistemas de IA podem perpetuar ou amplificar vieses preexistentes, entre eles os de gênero, raça e classe.

No Brasil, o processo de digitalização do Poder Judiciário intensificou-se a partir de 2018, com a adoção de projetos de automação e de IA em diversos tribunais, sob a coordenação do Conselho Nacional de Justiça (CNJ). Diversos sistemas inteligentes foram e estão sendo desenvolvidos e aplicados no contexto judicial, representando um novo paradigma de gestão e decisão judicial orientado por dados. Contudo, no caso brasileiro, a incorporação dessas tecnologias ocorre em um ambiente institucional que ainda reflete profunda assimetria de gênero: as mulheres representam menos de 40% da magistratura brasileira (Conselho Nacional de Justiça, 2025b; Carta de Brasília pela Igualdade de Gênero no Poder Judiciário, 2022). Essa sub-representação, combinada a uma cultura jurídica predominantemente masculina, levanta questões sobre como a IA – alimentada por decisões históricas – pode reforçar estereótipos e desigualdades, em vez de contribuir para superá-los.

Nesse cenário, a tese se concentra na análise da viabilidade e da efetividade do Direito como instrumento para mitigar os vieses de gênero – tanto algorítmicos quanto de dados – presentes na utilização de técnicas de IA no Poder Judiciário brasileiro. Assim, parte-se da seguinte indagação central: De que maneira a IA aplicada ao Poder Judiciário brasileiro pode reproduzir ou intensificar discriminações de gênero e em que medida o Direito pode atuar de forma eficaz na prevenção desses vieses e reorientar a tecnologia para promover a ética e a justiça social?

Tem-se como hipótese que o Direito pode atuar como instrumento regulatório eficaz para prevenir e mitigar a reprodução de vieses de gênero pela IA aplicada ao Poder Judiciário brasileiro, desde que articule simultaneamente garantias técnicas (como *fairness*, transparência e auditabilidade), garantias regulatórias (governança de dados, direitos humanos e devido processo legal) e garantias políticas (conscientização, formação crítica e letramento tecnológico). Essa articulação permitiria reorientar o uso da tecnologia para promover a ética, a equidade e a justiça social, assegurando decisões judiciais mais transparentes e inclusivas.

O objetivo geral desta pesquisa é propor protocolos e diretrizes que assegurem a utilização justa, transparente e inclusiva da IA no Poder Judiciário brasileiro, a partir da identificação dos vieses de gênero que esses sistemas podem reproduzir ou intensificar e da análise do papel do Direito como instrumento regulatório capaz de preveni-los e reorientar a tecnologia para a promoção da justiça social. Para alcançar esse propósito, a pesquisa desdobra-se em seis objetivos específicos, que orientam as etapas teórica, empírica e propositiva do trabalho.

O primeiro consiste em compreender os fundamentos técnicos e conceituais da IA aplicada ao campo jurídico, examinando os princípios da *machine learning*, do *natural language processing*, da *fairness* e da explicabilidade. Essa etapa inclui um levantamento bibliográfico sistemático em bases acadêmicas nacionais e internacionais, bem como a revisão da literatura sobre Direito digital e *LawTech*, de modo a contextualizar a incorporação da IA às rotinas e decisões do sistema judicial.

O segundo objetivo busca identificar e classificar os principais tipos de vieses presentes em sistemas de IA, distinguindo viés humano, viés de dados e viés algorítmico, a partir de uma análise crítica informada pela epistemologia feminista e pelos estudos de ciência, tecnologia e sociedade. Essa etapa envolve a revisão de relatórios institucionais e a aplicação de técnicas de análise de conteúdo, conforme Bardin (2016), para a categorização dos vieses e compreensão de seus efeitos no contexto jurídico.

O terceiro objetivo consiste em analisar o contexto estrutural do Poder Judiciário brasileiro, com foco nas desigualdades de gênero e em suas implicações para a adoção e operação de tecnologias de IA. Serão utilizados dados estatísticos do CNJ, a fim de avaliar a sub-representação feminina em diferentes instâncias e compreender como essa composição socioprofissional pode influenciar a reprodução de assimetrias no uso de tecnologias inteligentes.

O quarto objetivo visa mapear e examinar as iniciativas de uso de IA no Judiciário brasileiro, identificando os principais projetos implementados, seus potenciais benefícios e riscos em relação à reprodução de discriminações de gênero. Essa etapa envolve a análise documental de experiências emblemáticas e de relatórios técnicos e normativos do CNJ, buscando identificar padrões discursivos e práticas institucionais associadas à transformação digital da justiça.

O quinto objetivo propõe avaliar o papel e os limites do Direito como instrumento regulatório no enfrentamento dos vieses de gênero, articulando três dimensões de garantia: as técnicas, que envolvem parâmetros de *fairness*, transparência e auditabilidade; as regulatórias, relacionadas à governança de dados, aos direitos humanos e ao devido processo legal; e as políticas, ligadas à conscientização e ao letramento tecnológico. Essa análise abrange a revisão de legislações e marcos normativos nacionais e internacionais a fim de compreender suas potencialidades e lacunas.

Por fim, o sexto objetivo consiste em formular protocolos e diretrizes normativas, técnicas e pedagógicas que incorporem a perspectiva de gênero no desenvolvimento e aplicação da IA no sistema de justiça. A partir da síntese dos resultados empíricos e teóricos, essa etapa visa propor ajustes normativos e mecanismos de formação e letramento tecnológico voltados à construção de uma governança ética, inclusiva e democrática da tecnologia no Poder Judiciário brasileiro.

O referencial teórico desta tese é estruturado a partir de um diálogo interdisciplinar entre os campos da IA, da epistemologia feminista e da crítica do Direito, articulando conceitos técnicos, filosóficos e normativos para compreender a reprodução de vieses de gênero no uso de sistemas inteligentes pelo Poder Judiciário. A base teórica da IA é sustentada pelas formulações clássicas de Turing (1950), McCarthy et al. (1995) e Bellman (1998), bem como pelas conceituações de Russell e Norvig (2010), que definem a IA como um campo voltado ao desenvolvimento de agentes racionais capazes de tomar decisões autônomas. Ao incorporar essa perspectiva técnica, a tese reconhece que os algoritmos, ainda que concebidos para otimizar processos, refletem as escolhas humanas que os configuram – e, portanto, também reproduzem seus limites cognitivos, morais e sociais.

No âmbito das ciências cognitivas, as contribuições de Kahneman e Tversky (1974; 1981; 1982; 2013) e de Herbert Simon (1955) são fundamentais para estabelecer a conexão entre heurísticas e vieses humanos e os vieses algorítmicos. Esses autores demonstram que a racionalidade humana é limitada e orientada por atalhos mentais que simplificam decisões, mas

também geram distorções sistemáticas. Essa compreensão sustenta o argumento de que os sistemas de IA, desenvolvidos e treinados a partir de dados produzidos por seres humanos, acabam herdando e reproduzindo as mesmas distorções cognitivas e culturais. Nesse sentido, o referencial de Bardin (2016) sobre análise de conteúdo também é mobilizado para compreender como estruturas interpretativas – humanas ou computacionais – influenciam a produção de sentido e podem conduzir a resultados enviesados.

A sustentação crítica e transformadora da pesquisa decorre do aporte da epistemologia feminista e do tecnofeminismo, com destaque para Haraway (1987; 1991) e Wajcman (1991; 2004; 2010; 2020; 2023), que questionam a pretensa neutralidade da ciência e da tecnologia e evidenciam como estas reproduzem desigualdades de gênero. Esses referenciais, articulados à crítica do Direito tal como em Streck (2010) e em Rezende (2018), fundamentam a análise sobre o papel do Judiciário como espaço de poder historicamente masculinizado e como possível instrumento regulatório de transformação social. Assim, o referencial teórico da tese combina a compreensão técnica da IA com uma abordagem filosófica e jurídica crítica, propondo uma leitura integrada que revela como as decisões automatizadas podem amplificar injustiças históricas, mas também como o Direito pode atuar para reorientar a tecnologia em direção à ética, à transparência e à equidade de gênero.

A relevância desta pesquisa manifesta-se em três dimensões complementares: social, acadêmica e prática. Do ponto de vista social, o Judiciário é espaço central na garantia de direitos e na consolidação da democracia. A introdução da IA promete eficiência e celeridade processual, mas também pode cristalizar desigualdades históricas, especialmente em um contexto de sub-representação feminina na magistratura (Conselho Nacional de Justiça, 2025b; Carta de Brasília pela Igualdade de Gênero no Poder Judiciário, 2022). Se decisões passadas carregam estereótipos de gênero, algoritmos treinados com esses dados correm o risco de reproduzir injustiças em escala ampliada, comprometendo o princípio da igualdade material e a legitimidade democrática da Justiça.

No campo acadêmico, há escassez de estudos no Brasil que articulem de modo integrado IA, Direito e perspectiva feminista. Esta pesquisa propõe um diálogo entre três campos críticos: a epistemologia feminista, que questiona a neutralidade do conhecimento; a crítica do Direito, que problematiza a universalidade aparente das normas jurídicas; e os estudos de ciência e tecnologia, que demonstram que as tecnologias incorporam valores sociais e políticos. Essa convergência teórica constitui um marco interdisciplinar inédito para pensar a regulação da IA no sistema de justiça.

Já do ponto de vista prático e político, o trabalho busca oferecer protocolos e diretrizes concretas que incorporem a perspectiva de gênero à governança da IA no Judiciário, contribuindo para um sistema mais inclusivo, transparente e democrático, além de subsidiar políticas públicas baseadas em evidências e preocupações éticas.

A originalidade desta tese reside na articulação entre Direito, tecnofeminismo e governança algorítmica, propondo uma abordagem que ultrapassa as fronteiras tradicionais da dogmática jurídica. Ao integrar análises técnicas sobre o funcionamento da IA com reflexões filosóficas e sociológicas sobre poder e desigualdade, a pesquisa formula uma estrutura analítica capaz de compreender como a tecnologia pode tanto reproduzir quanto transformar padrões de exclusão de gênero. A proposta de protocolos e diretrizes técnicas, regulatórias e pedagógicas que incorporam a perspectiva feminista constitui uma contribuição inovadora e aplicável à prática institucional.

A pesquisa se insere no campo das Ciências Sociais Aplicadas, dialogando com diferentes áreas do conhecimento. No eixo do Direito, examina questões de neutralidade, hermenêutica jurídica e viés decisório, bem como os fundamentos normativos e institucionais da regulação da IA; no campo da Computação, discute princípios técnicos como *fairness*, explicabilidade e auditabilidade de sistemas inteligentes aplicados no Judiciário; a partir da Filosofia e da Sociologia, mobiliza conceitos de ética, poder e justiça social; e nos Estudos de Gênero e na epistemologia feminista, encontra o suporte crítico para repensar a neutralidade tecnológica e os mecanismos de exclusão simbólica. Essa pluralidade metodológica e conceitual reflete o caráter interdisciplinar do Programa de Pós-Graduação em Ciências Sociais Aplicadas (PPGCSA) da Universidade Estadual de Ponta Grossa (UEPG), ao qual esta tese se vincula, reforçando sua vocação para investigar fenômenos complexos que atravessam simultaneamente o social, o político, o jurídico e o tecnológico.

A tese está estruturada em cinco capítulos interdependentes, que articulam dimensões metodológicas, teóricas, empíricas e normativas. O primeiro capítulo expõe a metodologia, de caráter qualitativo e ancorada no estudo de caso, permitindo uma análise aprofundada da relação entre gênero, IA e Direito no contexto judicial brasileiro. Essa abordagem envolve pesquisa bibliográfica, análise documental e análise de conteúdo, conduzidas em etapas de pré-análise, codificação e interpretação crítica, sempre com atenção aos limites da análise qualitativa e às exigências éticas da investigação. O segundo capítulo estabelece o alicerce teórico, discutindo os conceitos fundamentais de IA, aprendizado de máquina e suas principais aplicações no campo jurídico. Questiona-se a suposta neutralidade da tecnologia e evidencia-se como vieses

humanos podem ser incorporados a algoritmos e bases de dados. Nesse ponto, introduz-se a contribuição do feminismo e do tecnofeminismo como ferramentas críticas e transformadoras, capazes de desconstruir estereótipos, propor narrativas inclusivas e oferecer caminhos inovadores para enfrentar desigualdades nas tecnologias digitais e nos processos decisórios. No terceiro capítulo, o foco desloca-se para o Poder Judiciário brasileiro, analisando seu perfil sociodemográfico e as implicações da desigualdade de gênero para a hermenêutica jurídica. Examina-se criticamente o Protocolo para Julgamento com Perspectiva de Gênero (Conselho Nacional de Justiça, 2021a), além de iniciativas recentes que buscam ampliar a representatividade feminina e promover uma abordagem mais equitativa nas decisões judiciais. O quarto capítulo volta-se à implementação da IA no Poder Judiciário, descrevendo experiências já em curso e mapeando desafios futuros relacionados à expansão do uso de sistemas inteligentes na prática decisória. O quinto capítulo apresenta as garantias éticas, técnicas e regulatórias indispensáveis para assegurar julgamentos livres de vieses de gênero. São debatidos princípios como *fairness*, transparência, explicabilidade e diversidade, bem como questões relacionadas à governança de dados e à regulação da IA. O capítulo propõe diretrizes concretas que combinam medidas técnicas (transparência e auditabilidade), regulatórias (governança algorítmica e conformidade com direitos humanos) e políticas (conscientização e letramento tecnológico em gênero), com vistas à promoção de julgamentos mais éticos e inclusivos. Por fim, o sexto capítulo traduz essas garantias em instrumentos operacionais, apresentando dois protocolos éticos: (i) um protocolo para desenvolvedores e equipes técnicas, que organiza práticas de planejamento ético, governança de dados, testes de equidade, explicabilidade e auditoria contínua; e (ii) um protocolo para magistrados, assessorias e usuários, que orienta o uso responsável, transparente e com julgamento com perspectiva de gênero, preservando a autonomia decisória humana e a legitimidade democrática das decisões judiciais.

Ao longo do trabalho, sustenta-se que a IA, quando desenvolvida e aplicada a partir de uma perspectiva crítica e feminista, pode se constituir em poderosa ferramenta de combate às desigualdades de gênero e de promoção da justiça social. Contudo, isso exige esforços contínuos e multidisciplinares, envolvendo juristas, técnicos, pesquisadores, ativistas e sociedade civil organizada. A presente pesquisa pretende contribuir para esse debate, oferecendo reflexões teóricas e propostas práticas que orientem a adoção responsável, ética e inclusiva da IA no Poder Judiciário e em outros espaços de poder.

CAPÍTULO 1 – DA METODOLOGIA DA TESE

Este capítulo tem por objetivo apresentar, detalhadamente, o percurso metodológico adotado nesta tese, esclarecendo todas as etapas que sustentam cientificamente o desenvolvimento do trabalho. Considerando a complexidade interdisciplinar do tema que envolve gênero, IA e Direito no contexto específico do Poder Judiciário brasileiro, optou-se pela adoção de uma metodologia qualitativa, baseada em um estudo de caso e complementada por técnicas de coleta e análise de dados.

A metodologia qualitativa revelou-se a mais apropriada por permitir uma análise aprofundada e interpretativa dos fenômenos investigados, facilitando o alcance dos objetivos específicos traçados pela pesquisa. Ao longo deste capítulo serão apresentados, de forma detalhada, a abordagem qualitativa escolhida, o tipo de pesquisa adotado, o método de investigação empregado, as técnicas de coleta e análise de dados utilizadas, bem como as limitações metodológicas reconhecidas e os aspectos éticos fundamentais que guiaram esta investigação. Finalmente, será evidenciada a relevância do método adotado para a contribuição científica e prática desta tese na compreensão e proposição de soluções concretas para enfrentar vieses algorítmicos de gênero no Judiciário, especialmente com o fim de promover a implementação ética e responsável da IA no Poder Judiciário brasileiro.

1.1 Abordagem da pesquisa

A abordagem qualitativa foi adotada devido à necessidade de explorar e interpretar fenômenos sociais complexos relacionados à interação entre gênero e IA, particularmente no contexto do Poder Judiciário brasileiro. Esta escolha metodológica está diretamente relacionada com as propostas de:

- a) explorar como os vieses algorítmicos podem reproduzir e amplificar discriminações de gênero, destacando suas implicações éticas e sociais;
- b) investigar as limitações e potencialidades do Direito como ferramenta para enfrentar tais vieses; e
- c) propor diretrizes concretas para garantir uma implementação ética e inclusiva da IA no Judiciário brasileiro.

A abordagem qualitativa permite compreender profundamente as percepções, experiências e práticas relacionadas ao uso da IA no contexto jurídico, identificando aspectos subjetivos, simbólicos e contextuais fundamentais para responder ao objetivo geral da pesquisa, que visa analisar como o Direito pode enfrentar os vieses de gênero produzidos por sistemas de IA, a partir de uma perspectiva feminista crítica. O estudo de Godoy (1995, p. 63) demonstra que a pesquisa qualitativa é apropriada para captar a complexidade inerente às relações sociais e institucionais, proporcionando uma análise detalhada e reflexiva, tornando-se, assim, fundamental para propor soluções práticas e teóricas para os desafios identificados nesta tese.

Ademais, adotou-se também uma abordagem de pesquisa interdisciplinar no campo das Ciências Sociais Aplicadas. O trabalho propõe uma articulação entre: Direito (enquanto instrumento normativo e de justiça social); Ciência da Computação (responsável pela criação e aplicação dos algoritmos); Estudos de Gênero (que analisam a desigualdade estrutural e a invisibilidade das mulheres na ciência e na tecnologia); e Sociologia e Filosofia (que problematizam as implicações éticas e sociais da IA).

Essa abordagem interdisciplinar é fundamental para compreender como os vieses de gênero se manifestam nos sistemas de IA e quais ferramentas podem ser eficazes para mitigá-los. A interdisciplinaridade não é apenas metodológica, mas constitutiva da análise, já que nenhum campo isolado é capaz de abarcar a complexidade do problema.

1.2 Tipos de pesquisa

A presente pesquisa é classificada como exploratória e descritiva.

1.2.1 Pesquisa exploratória

Inicialmente, adotou-se a pesquisa exploratória visando obter maior familiaridade e compreensão inicial sobre a temática da IA e suas implicações sociais e jurídicas, especialmente relacionadas ao gênero. Além disso, esta etapa da pesquisa serviu para identificar e compreender a composição do Poder Judiciário brasileiro, bem como verificar a crescente implementação tecnológica nesta esfera de poder. A pesquisa exploratória auxiliou na identificação dos conceitos fundamentais relacionados aos temas da IA, da ética e do feminismo tecnológico. Como explicado por Gil (2008, p. 27), a pesquisa exploratória é recomendada

quando há necessidade de adquirir clareza sobre um fenômeno pouco conhecido, proporcionando uma compreensão inicial sólida para embasar fases posteriores da pesquisa.

1.2.2 Pesquisa descritiva

Na etapa seguinte, foi utilizada a pesquisa descritiva, visando detalhar as características específicas e as relações estabelecidas entre o uso de IA e a reprodução ou mitigação de vieses de gênero, abordando o tema no contexto do Judiciário brasileiro. Essa etapa incluiu a descrição detalhada da conjectura institucional, procedimentos, práticas, protocolos e regulamentos existentes no âmbito judicial. Tal como aponta Gil (2008, p. 28), a pesquisa descritiva é necessária para mapear e descrever detalhadamente fenômenos já identificados, permitindo compreender melhor suas dinâmicas internas e externas.

Essas duas fases foram escolhidas por proporcionarem uma compreensão abrangente e aprofundada do objeto de estudo, permitindo tanto a exploração inicial quanto uma análise detalhada e sistemática do fenômeno investigado. Esta combinação metodológica revelou-se essencial para atingir os objetivos específicos da pesquisa, contribuindo diretamente para responder à questão central proposta por esta tese.

1.3 Método de investigação

O método utilizado nesta pesquisa foi o estudo de caso, conforme proposto por Yin (2001). A escolha é justificada em função da proposta central da tese e dos objetivos específicos.

Inicialmente, o estudo de caso foi selecionado por permitir um exame detalhado e aprofundado de um fenômeno contemporâneo específico, que é a aplicação da IA no contexto do Poder Judiciário brasileiro, especialmente relacionado ao viés de gênero. Depois, justificou-se seu uso pelo fato deste método ser particularmente apropriado devido à complexidade e à novidade do fenômeno estudado, caracterizado por fronteiras pouco definidas entre tecnologia, aspectos jurídicos e questões sociais e éticas.

Ainda, vale destacar que o estudo de caso possibilita a análise contextualizada das práticas judiciais, políticas institucionais e estratégias tecnológicas adotadas pelo Judiciário brasileiro, permitindo uma compreensão rica e detalhada sobre como os vieses algorítmicos são

reproduzidos pela IA em um aspecto geral, bem como possíveis mecanismos para sua mitigação, buscando garantias éticas que promovam a implementação e o bom uso da IA no Judiciário. Finalmente, a escolha deste método é diretamente relacionada ao alcance dos objetivos específicos da pesquisa, permitindo explorar concretamente as implicações éticas e sociais dos vieses, investigar limitações e potencialidades do Direito nesse contexto e elaborar diretrizes específicas e aplicáveis para a implementação ética da IA no âmbito judicial.

Em resumo, o estudo de caso foi o método mais adequado e eficaz para atender integralmente aos objetivos específicos propostos pela tese, fornecendo uma base empírica sólida para responder ao problema central e contribuir significativamente para as discussões acadêmicas e práticas sobre o tema.

1.4 Universo, amostra e sujeitos da pesquisa

O universo da pesquisa compreende o Poder Judiciário brasileiro, considerado em sua dimensão institucional, normativa e tecnológica, na medida em que se insere em um processo mais amplo de digitalização e de adoção de sistemas de IA voltados à automação de rotinas e ao apoio à tomada de decisão. Esse universo abarca os órgãos de cúpula e de coordenação nacional – em especial o CNJ –, os tribunais superiores, federais e estaduais e demais instâncias do sistema de justiça que vêm incorporando ferramentas de IA em suas práticas administrativas e jurisdicionais.

A amostra é composta por três conjuntos de fontes documentais:

- a) documentos oficiais emitidos por órgãos do Judiciário, tais como atos normativos, resoluções, recomendações, relatórios técnicos e planos estratégicos produzidos pelo CNJ e pelos tribunais, com destaque para os materiais que abordam o uso da IA no contexto da gestão judicial e da formulação de políticas judiciárias;
- b) casos emblemáticos em que o uso de algoritmos ou sistemas automatizados influenciou decisões judiciais ou procedimentos processuais, permitindo problematizar a dimensão prática e ética da aplicação dessas tecnologias no processo decisório;
- c) produções acadêmicas e técnicas – artigos, dissertações, teses, pareceres e relatórios de pesquisa – que discutem o uso da IA no

Judiciário, seus potenciais benefícios e riscos, bem como os fundamentos teóricos e jurídicos que sustentam ou criticam sua adoção.

A seleção desse material tem caráter intencional e qualitativo, voltada à compreensão crítica do fenômeno, e não à representatividade estatística. Os documentos foram escolhidos conforme sua relevância temática, abrangência institucional e impacto no debate público sobre IA e justiça.

Não há, coleta de dados com sujeitos humanos, uma vez que o foco da investigação recai sobre fontes documentais institucionais, doutrinárias e jurisprudenciais, analisadas segundo procedimentos de pesquisa documental e análise de conteúdo.

Essa estratégia metodológica permite examinar de forma sistemática as narrativas, justificativas e discursos presentes nos documentos, identificando como o Poder Judiciário brasileiro vem incorporando a IA em sua estrutura decisória e administrativa e quais valores, pressupostos e riscos permeiam esse processo.

Os critérios de seleção dos documentos consideram:

- a) pertinência direta com o tema da pesquisa, isto é, materiais que abordem o uso, a regulação, a implementação ou os impactos éticos e sociais da IA no Judiciário;
- b) em especial, mas não restritivamente, o período temporal compreendido entre 2018 e 2025, correspondente à fase de institucionalização das políticas de transformação digital e de maior difusão das ferramentas de IA no sistema de justiça;
- c) origem institucional reconhecida, priorizando documentos oficiais do CNJ e dos tribunais, além de publicações acadêmicas avaliadas por pares.

Foram excluídos materiais sem autoria identificável e publicações não verificáveis.

1.5 Técnicas de coleta de dados

Para alcançar o objetivo geral desta pesquisa, conduziu-se uma pesquisa bibliográfica e documental, seguida da análise do material mediante técnica de análise de conteúdo conforme sistematizada por Bardin (2016).

A pesquisa bibliográfica e documental foi realizada em duas etapas basilares para fundamentar teórica e empiricamente o objeto investigado (Marconi e Lakatos, 2002), conforme abaixo descrito.

1.5.1 Levantamento bibliográfico

O levantamento bibliográfico constituiu uma etapa fundamental da pesquisa, voltada à construção dos referenciais teórico-conceituais que sustentam a análise crítica proposta. Essa etapa envolveu um mapeamento sistemático e analítico da literatura científica, com foco na identificação, seleção e organização de publicações relevantes para a compreensão das interseções entre IA, questões de gênero, governança algorítmica e Direito, com ênfase na aplicação dessas temáticas ao contexto do Poder Judiciário brasileiro.

A busca bibliográfica foi conduzida em bases de dados acadêmicas de alta relevância e reconhecida credibilidade científica, incluindo o Portal de Periódicos da CAPES, a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), a SciELO, o Google Scholar, a Scopus, além das bibliotecas física e digital da Pontifícia Universidade Católica do Paraná (PUCPR) e da Universidade Estadual de Ponta Grossa (UEPG). O uso combinado dessas fontes permitiu abranger tanto produções nacionais quanto internacionais, assegurando uma visão comparativa e interdisciplinar sobre o fenômeno investigado.

Os critérios de seleção adotados para o levantamento consideraram:

- a) relevância temática, priorizando trabalhos que discutem a aplicação da IA no sistema de justiça, a relação entre tecnologia e desigualdades de gênero, os desafios éticos da governança algorítmica e as dimensões jurídicas da regulação tecnológica;
- b) atualidade, com foco em publicações produzidas preferencialmente entre 2018 e 2025, período que marca a intensificação do debate sobre automação e viés algorítmico no campo jurídico; e
- c) pertinência epistemológica, valorizando produções que dialogam com as Ciências Sociais Aplicadas sob perspectivas críticas, interseccionais e feministas.

Foram incluídos artigos científicos, livros, capítulos de livros, dissertações, teses, relatórios técnicos e pareceres institucionais, desde que apresentassem rigor metodológico e

contribuição teórica significativa. Após a coleta, as obras foram organizadas em eixos analíticos, agrupando autores e correntes de pensamento conforme sua abordagem dos principais temas da tese – tais como Poder Judiciário e tecnologia; IA, ética e viés algorítmico; gênero e desigualdade; “neutralidade” tecnológica; e feminismo e epistemologias críticas.

A análise bibliográfica foi conduzida de forma interpretativa e crítica, articulando contribuições oriundas do campo jurídico, da filosofia política, da sociologia da tecnologia e dos estudos de gênero. Esse procedimento permitiu compreender a complexidade conceitual, histórica e epistemológica dos debates contemporâneos sobre IA e justiça, situando a pesquisa em um campo interdisciplinar que envolve a mediação entre direito, IA, poder e desigualdade de gênero.

Por fim, o levantamento bibliográfico não se limitou à mera revisão de literatura, mas configurou-se como um instrumento de reflexão teórico-crítica, indispensável para a formulação das categorias analíticas utilizadas ao longo da tese. Esse processo assegurou que a fundamentação teórica não apenas sustentasse a análise empírica dos documentos judiciais e institucionais, mas também contribuisse para revelar as tensões e disputas simbólicas que permeiam o discurso sobre neutralidade, eficiência e justiça algorítmica no contexto brasileiro.

1.5.2 Análise documental direta e indireta

A análise consistiu na seleção e exame detalhado de documentos oficiais, legislações pertinentes, relatórios técnicos sobre a implementação tecnológica no Judiciário e desenvolvimento e aplicação do Protocolo para Julgamento com Perspectiva de Gênero (Conselho Nacional de Justiça, 2021a), bem como em estudo de livros e artigos científicos a respeito do tema. A análise documental direta forneceu evidências concretas sobre como as práticas institucionais abordam (ou não) as questões relacionadas ao viés de gênero e à IA, facilitando uma compreensão crítica e aprofundada do contexto estudado.

1.6 Procedimento para análise dos dados

A análise qualitativa dos dados coletados foi orientada pela técnica de análise de conteúdo conforme sistematizada por Bardin (2016), a qual possibilita a interpretação crítica e estruturada de materiais textuais, permitindo identificar padrões, significados e relações entre os discursos analisados. Essa técnica mostrou-se adequada ao objetivo da pesquisa, que busca

compreender de que forma o uso da IA no Poder Judiciário brasileiro pode reproduzir ou mitigar vieses de gênero, articulando as dimensões normativas, institucionais e discursivas envolvidas.

O processo analítico foi operacionalizado em etapas interdependentes, que se complementam e conferem consistência à interpretação dos resultados:

- a)** Pré-análise: nesta etapa inicial, realizou-se a seleção rigorosa e a organização sistemática dos documentos coletados – incluindo decisões judiciais, relatórios técnicos, atos normativos, artigos científicos, livros, dissertações, teses e documentos institucionais como o “Protocolo para Julgamento com Perspectiva de Gênero” (Conselho Nacional de Justiça, 2021a) do CNJ. A leitura flutuante do material permitiu uma visão global preliminar e favoreceu a definição de categorias iniciais de análise, alinhadas aos objetivos da pesquisa e aos referenciais teóricos adotados.
- b)** Exploração do material: com o corpus devidamente delimitado e organizado, procedeu-se à leitura detalhada e exaustiva dos textos, seguida da codificação e categorização sistemática do conteúdo. As categorias de análise foram previamente definidas a partir do quadro teórico e das hipóteses da pesquisa, contemplando os seguintes eixos: Poder Judiciário e tecnologia; IA e viés algorítmico; gênero e desigualdade; “neutralidade” tecnológica; e feminismo e epistemologias críticas. A categorização permitiu estruturar o material em núcleos temáticos, facilitando a identificação de padrões discursivos, convergências e tensões entre os documentos analisados.
- c)** Tratamento dos resultados e interpretação: nessa fase, os dados categorizados foram submetidos a uma interpretação crítica e inferencial, com base nos referenciais teóricos previamente estabelecidos. A análise buscou compreender os sentidos produzidos pelos discursos institucionais e jurídicos acerca da IA, examinando como conceitos como eficiência, objetividade e neutralidade tecnológica são construídos, legitimados e eventualmente tensionados por perspectivas feministas e críticas.

A utilização detalhada dessas técnicas permitiu um aprofundamento analítico substancial, essencial para responder ao objetivo geral da pesquisa: examinar como o Direito

pode atuar de forma eficaz na prevenção e mitigação dos vieses de gênero produzidos ou amplificados por sistemas de IA no contexto do Judiciário brasileiro.

Dessa forma, o procedimento metodológico adotado garante não apenas a rastreabilidade e transparência da análise, mas também a coerência entre as etapas empíricas e teóricas do trabalho, assegurando uma leitura interdisciplinar que articula tecnologia, gênero e justiça sob uma perspectiva crítica e aplicada às Ciências Sociais.

1.7 Aspectos éticos

A pesquisa não envolve diretamente participantes humanos, o que a dispensa de submissão ao Comitê de Ética. No entanto, respeita rigorosamente os princípios éticos da pesquisa científica, especialmente o uso responsável das fontes, a transparência metodológica, a preservação da integridade intelectual e o reconhecimento de limitações interpretativas.

A perspectiva de gênero também é assumida como princípio ético, buscando visibilizar desigualdades e propor caminhos para sua mitigação.

1.8 Relevância do método adotado

O método qualitativo adotado se revelou adequado para atender plenamente aos objetivos específicos da presente pesquisa, ao permitir explorar em profundidade a reprodução e amplificação de discriminações de gênero provocadas pelos vieses algorítmicos e que podem colocar em xeque a implementação ética da IA no Poder Judiciário brasileiro. Ao analisar criticamente documentos, relatórios e protocolos judiciais, pôde-se compreender as dinâmicas internas e externas desses fenômenos sociais complexos no contexto específico do Judiciário brasileiro.

O método qualitativo também possibilitou investigar as limitações e potencialidades do Direito como ferramenta regulatória para o enfrentamento dos vieses. Ao realizar uma análise qualitativa detalhada das legislações, decisões judiciais e documentos institucionais, foi possível identificar claramente os limites e as possibilidades do ordenamento jurídico vigente diante das questões relacionadas à IA e ao gênero.

Por meio dessa abordagem, foi possível propor diretrizes concretas e fundamentadas empiricamente para a implementação ética e inclusiva da IA no âmbito judicial. A análise

qualitativa permitiu identificar estratégias práticas, contextualmente apropriadas, para mitigar vieses de gênero e promover uma governança algorítmica mais justa e equitativa.

Assim, o método qualitativo proporcionou uma base sólida para o desenvolvimento de uma análise crítica e profunda das relações entre IA, gênero e Direito, contribuindo significativamente para avançar nas reflexões e práticas relativas à governança algorítmica e à implementação ética das tecnologias digitais no Judiciário brasileiro.

1.9 Limitações metodológicas

Reconhece-se como limitação desta pesquisa a subjetividade inerente à interpretação qualitativa. Essa subjetividade pode influenciar as análises e conclusões, considerando que os dados coletados, principalmente por meio de documentos e decisões judiciais, envolvem interpretações que podem variar conforme a perspectiva do pesquisador. Além disso, dado o caráter exploratório e descritivo, os achados e inferências estão condicionados ao contexto específico do estudo de caso selecionado, que é o Poder Judiciário brasileiro.

Outro fator limitante é o acesso restrito ou parcial a determinadas informações ou documentos institucionais, o que pode impactar na abrangência da análise realizada. Esta limitação pode afetar diretamente a compreensão plena dos mecanismos pelos quais os vieses de gênero, algorítmicos ou não, são reproduzidos e enfrentados na prática judiciária.

Para mitigar essas limitações, o processo de análise foi realizado com transparência metodológica, permitindo um acompanhamento detalhado das etapas percorridas, oferecendo maior confiabilidade e validade aos resultados apresentados. A conscientização das limitações apontadas contribui para um entendimento mais realista das potencialidades e restrições das conclusões obtidas nesta tese, bem como para a identificação de caminhos para futuras pesquisas na área.

CAPÍTULO 2 – A INTELIGÊNCIA ARTIFICIAL E O DESAFIO DO VIÉS DE GÊNERO

A evolução tecnológica exerce um impacto profundo e, muitas vezes, inesperado, sobre nossa sociedade. Por um lado, a tecnologia oferece libertação, promovendo emancipações (Friedrich; Philippi, 2020), ampliando capacidades (Sen, 1985, p. 201) e elevando significativamente a qualidade de vida. Por outro lado, ela gera desafios e problemas sociais, econômicos e políticos que, em muitos casos, são imprevisíveis ou ainda carecem de soluções claras (Sadin, 2019). Nesse contexto, reflexões filosóficas, sociológicas e epistemológicas se tornam essenciais para compreender o impacto da crescente digitalização e da tecnologia na vida cotidiana, buscando formas de utilizá-las de maneira mais equilibrada, comedida e responsável.

As questões sociais, econômicas e políticas, direta ou indiretamente afetadas pela evolução tecnológica, remontam à Revolução Industrial, embora superem esse marco histórico. Com o advento da Era Digital (Castells, 2011), surgiram novas possibilidades que permitiram à tecnologia adentrar áreas do conhecimento e das interações sociais de maneira muito mais veloz, transformando profundamente a condição da vida humana tal como a conhecíamos. Atualmente, a digitalização de diversos setores cria oportunidades únicas para a criação de novas tecnologias, gerações de conhecimentos inéditos e novas formas de relacionamento.

Entre as conquistas da digitalização está a criação de espaços de poder comunicativo. Com isso, definiu-se quem detém a capacidade de fala, quem será ouvido e quem será marginalizado nesse novo cenário de interações tecnológicas. À medida que a ciência acompanha essas transformações sociais, a maneira e a rapidez com que o conhecimento é produzido também se alteram, acompanhando o ritmo das novas tecnologias (Latour, 1994).

Diante do imenso impacto histórico da civilização tecnológica, com o surgimento de inteligências artificiais, nanotecnologias, robótica, internet das coisas, smartphones, *smartwatches*, e-commerce e outras ferramentas digitais, torna-se impossível identificar um campo que não seja tocado pela velocidade, diversidade e volume da produção e disseminação de dados. O atual estágio de desenvolvimento tecnológico, veloz e sistêmico, favorece interações recíprocas entre sociedade e tecnologia (Schwab, 2016, p. 11).

O presente capítulo explora a interseção entre IA e o viés de gênero, um tema de crescente relevância em múltiplas esferas da sociedade contemporânea. A incorporação da IA em diversas áreas, incluindo o campo do Direito, traz consigo promessas de eficiência, inovação

e otimização de processos. No entanto, também suscita preocupações significativas sobre a potencial perpetuação e amplificação de desigualdades estruturais, sobretudo aquelas relacionadas ao gênero. O viés de gênero, nesse contexto, emerge como um desafio crítico, demandando uma análise aprofundada das ferramentas tecnológicas que influenciam decisões jurídicas e sociais.

Embora as técnicas de aprendizado de máquina tenham avançado de forma notável, sua aplicação em domínios como o Direito frequentemente reflete e reforça desigualdades sociais pré-existentes, incluindo os vieses de gênero. Isso ocorre porque os sistemas de IA são alimentados por dados históricos que carregam marcas de discriminações passadas, reproduzindo padrões desiguais em suas decisões automatizadas. Assim, a discussão sobre a neutralidade da tecnologia torna-se central, questionando-se até que ponto a IA pode ser verdadeiramente imparcial em um mundo marcado por assimetrias de poder.

O capítulo estrutura-se em três momentos principais. Inicialmente, são apresentadas as definições de IA e de aprendizado de máquina, estabelecendo uma base conceitual sólida para a compreensão do tema (subcapítulo 2.1). Em seguida, examinam-se as principais técnicas de aprendizado de máquina atualmente em uso e suas aplicações no campo do Direito (subcapítulo 2.2), destacando como essas tecnologias estão transformando a prática jurídica e, ao mesmo tempo, gerando novos dilemas éticos e sociais.

Em um terceiro momento, aprofunda-se a análise sobre a presença do viés humano tanto no processo cognitivo e decisório quanto nos sistemas de IA, questionando a viabilidade da neutralidade nas decisões algorítmicas (subcapítulo 2.3). Esse segmento explora como os vieses humanos se manifestam e são incorporados aos sistemas de IA, desconstruindo a noção de que a tecnologia é intrinsecamente imparcial (subcapítulos 2.3.1 e 2.3.2). A discussão inclui uma reflexão crítica sobre os mecanismos pelos quais os vieses de gênero são reproduzidos e amplificados, seja por meio dos dados utilizados para treinar os algoritmos, seja pela própria estrutura dos sistemas tecnológicos.

Ao longo deste capítulo, busca-se não somente compreender as complexidades envolvidas na interação entre tecnologia, tomada de decisão e gênero, mas também contribuir para um debate mais amplo sobre justiça social no contexto da IA. A análise proposta visa a identificar caminhos para o desenvolvimento e a implementação de sistemas tecnológicos mais éticos e inclusivos, capazes de enfrentar os desafios impostos pelos vieses de gênero e promover uma sociedade mais equitativa.

2.1 Definindo inteligência artificial e aprendizagem de máquina

As ideias e trabalhos que lançaram as bases para os conceitos da IA têm como figura paterna Alan Turing, matemático e lógico britânico nascido em 1912 que concebeu a “máquina de Turing”, modelo teórico fundamental para a computação moderna, e o “teste de Turing”, que é um critério de avaliação da inteligência de uma máquina com base em sua capacidade de imitar a inteligência humana. Foi em 1950 que Turing publicou o artigo *Computing Machinery and Intelligence* (Turing, 1950) que trouxe o questionamento “máquinas podem pensar?” e que propôs o teste de Turing como avaliação do comportamento inteligente computacional.

McCarthy, Minsky, Rochester e Shannon, responsáveis pela criação do termo “inteligência artificial” na *Dartmouth Summer Research Project on Artificial Intelligence*, no ano de 1956, descrevem a IA como técnica de fazer máquinas se comportarem de maneira inteligente como seres humanos¹. Conforme Bellman (1978, p. 2, tradução nossa²), a IA é “a automatização de atividades que associamos à cognição humana, tais como a tomada de decisões, resolução de problemas e o aprendizado”.

Dentre os variados conceitos de IA na literatura, pode ela ser caracterizada como a ciência de fazer máquinas inteligentes. Os autores Russell e Norvig (2010, p. 2) explicam as diversas abordagens da IA no Quadro 1, mas entendem que a melhor definição é aquela do “agente racional”, ou seja, a que trata sistemas inteligentes como máquinas que executam funções que exigem inteligência (racionalidade) quando são desempenhadas por seres humanos.

Quadro 1. Definições de inteligência artificial.

Definições de inteligência artificial	
Pensando como ser humano “O empolgante novo esforço para fazer os computadores pensarem... máquinas com mentes, no sentido pleno e literal” (Haugeland, 1985). “A automação de atividades que associamos ao pensamento humano, atividades como tomada de decisão, resolução de problemas e aprendizado...” (Bellman, 1978).	Pensando racionalmente “O estudo das faculdades mentais por meio do uso de modelos computacionais”. (Chamiak; McDermott, 1985) “O estudo computacional que torna possível perceber, raciocinar e agir” (Winston, 1992).

¹ No original: “[...] the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving.” (McCarthy et al., 1955).

² No original: “The automation of activities that we associate with human thinking, activities such as decision-making, problem solving, and learning.”

Agindo como ser humano “A arte de criar máquinas que realizam funções que requerem inteligência quando executadas por pessoas” (Kurzweil, 1990). “O estudo de como fazer os computadores realizarem tarefas nas quais, no momento, as pessoas são melhores” (Rich; Knight, 1991)	Agindo racionalmente “A inteligência computacional é o estudo do <i>design</i> de agentes inteligentes” (Poole et al., 1998). “A IA se preocupa com o comportamento inteligente em artefatos” (Nilsson, 1998).
---	---

Fonte: Russell e Norvig, 2010, p. 2, adaptação e tradução nossas³.

Para Badia e González (2017, p. 8), a IA pode ser rotulada como uma ciência e uma engenharia que permite projetar e programar computadores para a realização de tarefas que exigem inteligência, permitindo replicar a inteligência humana por meio de máquinas.

A IA compreende um conjunto de métodos que permitem a computadores resolver tarefas complexas, as quais exigiriam inteligência humana para serem resolvidas. Através da IA, um computador se torna um instrumento técnico “pensante” (Hoffmann-Riem, 2020, p. 434). Ela diz respeito, portanto, à reprodução digitalmente estruturada da capacidade de tomada de decisão semelhante à humana.

O conceito de IA trazida pela Comissão Europeia de *experts* em IA no “Draft Ethics Guidelines for Trustworthy AI” retrata a essência da engenhosidade dos sistemas inteligentes, bem como elenca as suas subáreas científicas (High-Level Expert Group on Artificial Intelligence, 2018, p. iv, tradução nossa⁴):

Thinking Humanly “The exciting new effort to make computers think . . . <i>machines with minds</i> , in the full and literal sense.” (Haugeland, 1985) “[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . .” (Bellman, 1978)	Thinking Rationally “The study of mental faculties through the use of computational models.” (Charniak and McDermott, 1985) “The study of the computations that make it possible to perceive, reason, and act.” (Winston, 1992)
Acting Humanly “The art of creating machines that perform functions that require intelligence when performed by people.” (Kurzweil, 1990) “The study of how to make computers do things at which, at the moment, people are better.” (Rich and Knight, 1991)	Acting Rationally “Computational Intelligence is the study of the design of intelligent agents.” (Poole et al., 1998) “AI . . . is concerned with intelligent behavior in artifacts.” (Nilsson, 1998)
Figure 1.1 Some definitions of artificial intelligence, organized into four categories.	

³ No original:

⁴ No original: “Artificial intelligence (AI) refers to systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from this data and deciding the best action(s) to take (according to pre-

A inteligência artificial (IA) refere-se a sistemas projetados por humanos que, diante de um objetivo complexo, agem no mundo físico ou digital ao perceberem seu ambiente, interpretarem os dados estruturados ou não estruturados coletados, raciocinarem com base no conhecimento derivado desses dados e decidirem as melhores ações a serem tomadas (de acordo com parâmetros predefinidos) para alcançar o objetivo proposto. Sistemas de IA também podem ser projetados para aprender a adaptar seu comportamento ao analisar como o ambiente é afetado por suas ações anteriores.

Como uma disciplina científica, a IA inclui várias abordagens e técnicas, como aprendizado de máquina (do qual o aprendizado profundo e o aprendizado por reforço são exemplos específicos), raciocínio de máquina (que abrange planejamento, agendamento, representação e raciocínio do conhecimento, busca e otimização) e robótica (que inclui controle, percepção, sensores e atuadores, bem como a integração de todas as outras técnicas em sistemas ciberfísicos).

No presente trabalho, a definição adotada aborda a IA como um ramo da computação que tem por finalidade, dentre outras, desenvolver o aprendizado de máquina por meio de diversas técnicas computacionais que visam imitar o raciocínio humano, permitindo que computadores resolvam diversas tarefas. São exemplos dessas tarefas a compreensão de linguagens, o reconhecimento de objetos e sons, a identificação de padrões e tendências em dados, a tomada de decisões, o planejamento e agendamento de tarefas de maneira otimizada e a resolução de problemas complexos.

Vale dizer, o objetivo da IA não é replicar a totalidade da inteligência humana com todas as suas capacidades e recursos. Em vez disso, busca-se reproduzir ou emular certas funções específicas, como habilidades de resolução de problemas mediante processos inferenciais. Isso visa alcançar desempenhos qualitativos comparáveis aos da mente humana, porém com vantagens em termos de quantidade e velocidade. Em vez de apenas inteligência, podemos considerar essas operações como racionalidade, otimizando os resultados e minimizando os tempos de execução, além de ampliar drasticamente o número de problemas analisados e soluções identificadas por unidade de tempo. Além disso, pode-se afirmar que a IA assume funções que hoje são exclusivas da inteligência humana, mas que no futuro podem ser rotineiramente executadas por sistemas computacionais inteligentes.

Embora as técnicas de IA tenham sido objeto de discussão desde 1950, após um longo período de lentidão em suas inovações, houve um considerável avanço nesse campo ao longo da última década. Hoje, diversos elementos desempenham papel crucial no aprimoramento dos

defined parameters) to achieve the given goal. AI systems can also be designed to learn to adapt their behaviour by analysing how the environment is affected by their previous actions.

As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems)."

sistemas de IA. Isso inclui a enorme quantidade de dados que agora são gerados e disponibilizados em ambientes online, a capacidade das máquinas de aprender com as informações fornecidas a elas e o progresso dos computadores, que experimentaram uma redução de custos e o desenvolvimento de novas modalidades de algoritmos voltados para a compreensão e a simulação da capacidade cognitiva humana.

Dados e algoritmos são os insumos para o desenvolvimento da IA. Conforme avança o processo de digitalização da sociedade, uma vasta quantidade de dados é gerada. Imagens, vídeos, textos e outros tipos de informação são disseminados pelo espaço virtual através de computadores, smartphones e outras tecnologias. Esses dados, que se destacam não somente pela sua quantidade massiva, mas também pela rapidez com que são transmitidos, exigem uma análise que vai além das correlações concebidas pela mente humana, necessitando do auxílio de programas tecnológicos mais sofisticados.

Um algoritmo é uma sequência finita de passos bem definidos, estruturados e representados matematicamente, para resolver um problema ou executar uma tarefa. Em outras palavras, algoritmos são compostos por uma sequência de códigos matemáticos que servem de instrução visando alcançar um objetivo, um resultado. Seu funcionamento consistirá numa escolha dentre várias possibilidades de resultados (Gillespie, 2016, p. 18-30). De modo mais simples, são diretrizes, instruções, seguidas por uma máquina.

Os algoritmos funcionam construindo um modelo com base em amostras de entrada, ou seja, quando um computador ou dispositivo recebe comandos ou sinais de fontes externas. Nesse processo, a criação de um algoritmo de IA envolve diversas etapas manuais, desde a coleta de dados até a seleção de quem irá realizá-la. Desde a origem dos dados até a análise sobre como o problema em questão está sendo representado e potencialmente resolvido pelo algoritmo. Portanto, via de regra, a intervenção humana na elaboração dessa tecnologia é significativa, pelo menos inicialmente. Ressaltamos isso porque algoritmos, aprendizado de máquina e IA não são termos intercambiáveis. É crucial esclarecer essa distinção para evitar qualquer confusão, especialmente considerando o nosso papel nesta análise, que é realizar uma investigação sociojurídica sobre o tema.

São várias as definições de aprendizagem de máquina. Uma delas é apresentada por Mitchell (1997, p. XV, tradução nossa⁵), como sendo o campo da computação preocupado com o desenvolvimento de “programas de computador [algoritmos] que podem melhorar

⁵ No original: “The field of machine learning is concerned with the question of how to construct computer programs that automatically improve with experience”.

automaticamente através da experiência”. De maneira geral, pode-se afirmar que é com a aprendizagem de máquina que computadores são programados para aprender com experiências passadas.

A definição trazida por Wolkart (2018, p. 665) reflete a função de um algoritmo de aprendizagem de máquina. Segundo o autor, o sistema de aprendizagem de máquina é aquele no qual os algoritmos “são capazes de prever ou generalizar padrões apreendidos a partir de um conjunto de dados utilizados para treinar o sistema”.

De acordo com Faceli et al. (2011, p. 3), a aprendizagem de máquina emprega um princípio de inferência denominado indução, por meio do qual “se obtêm conclusões genéricas a partir de um conjunto particular de exemplos”. Esses algoritmos aprendem “a induzir uma função ou hipótese capaz de resolver um problema a partir de dados que representam instâncias do problema a ser resolvido”. Esses dados formam um conjunto, chamado de “conjunto de dados” (Faceli et al., 2011, p. 3). Um conjunto de dados pode representar um objeto físico ou uma noção abstrata e, via de regra, esses objetos são representados por um vetor de características que os descreve, também chamado de atributos, campos ou variáveis (Faceli et al., 2011, p. 12).

Conforme Mulholland (2020, p. 331), a técnica “se configura como qualquer metodologia e conjunto de técnicas que utilizam dados em larga escala (*input*) para criar conhecimento e padrões originais e, com base neles, gerar modelos que são usados para predição a respeito dos dados tratados (*output*)”. Através desta, a IA adquire a capacidade de expandir suas experiências, adquirindo conhecimento delas, por meio de um ciclo de aprendizado contínuo e progressivo. A IA atinge sua máxima eficiência ao incorporar esses métodos de análise de dados.

Assim, pode-se dizer que o aprendizado de máquina se refere à prática de utilizar algoritmos para coletar dados, aprender com eles e, em seguida, fazer uma determinação ou previsão sobre algo. Nesse processo, a máquina é treinada para uma tarefa específica utilizando uma grande quantidade de dados e são os algoritmos que conferem a ela a capacidade de aprender a executá-la. Os algoritmos são entendidos como uma sequência finita de instruções que buscam resolver um problema. Eles são passos essenciais para a realização de uma tarefa, mas não necessariamente se referem a um programa de computador, já que sua implementação pode ser realizada por um computador, uma pessoa ou até mesmo outro tipo de autômato.

A aprendizagem de máquina utiliza o *big data* para a criação de algoritmos. O termo *big data* diz respeito a uma coleção de dados de grande volume, diversidade e de complexidade,

que podem ser estruturados (com padrão pré-definido e com estrutura definida e rígida), não estruturados (sem padrão pré-definido e sem estrutura estabelecida) ou semiestruturados (com representação estrutural heterogênea e parcialmente estruturados). Ele é um “ecossistema complexo de tecnologias, metodologias e processos usados para capturar, armazenar, gerenciar e analisar vastos volumes de dados diversificados” (Badman; Kosinski, 2024).

O objetivo da aprendizagem de máquina pode ser descritivo (o resultado seria informações a respeito da própria base de dados) ou prescritivo (o resultado seria a classificação e predição sobre determinada coisa com a utilização da ampla gama de dados). De acordo com Faleci et al. (2011, p. 6), nas tarefas de previsão, “a meta é encontrar uma função [...] a partir dos dados de treinamento que possa ser utilizada para prever um rótulo ou valor que caracterize um novo exemplo, com base nos valores de seus atributos de entrada.” Neste tipo de tarefa, cada objeto do conjunto de treinamento tem atributos de entrada e de saída. Os algoritmos utilizados para cumprir com tarefas preditivas seguem, via de regra, o modelo de aprendizagem de máquina supervisionada.

Já a tarefa descrição, como o próprio nome diz, tem por objetivo explorar ou descrever um conjunto de dados. Os algoritmos que são responsáveis pela realização dessa tarefa não usam do atributo de saída e por isso seguem o paradigma de aprendizado de máquina não supervisionado.

Tendo em vista a importância da IA como base para sistemas autônomos, é essencial compreender as técnicas de aprendizado de máquina que possibilitam sua aplicação prática, especialmente em áreas como o Direito, onde tais sistemas têm impacto significativo. A aprendizagem de máquina pode ser subdividida em supervisionada, não supervisionada e por reforço, questões que serão abordadas na subseção a seguir.

2.2 Principais técnicas de aprendizagem de máquina e algumas de suas aplicações no campo do Direito

A aprendizagem de máquina supervisionada utiliza modelos treinados com exemplos rotulados; o banco de dados que serve de treinamento de tais sistemas contém a resposta desejada para a solução do problema (Igual e Seguí, 2017, p. 70; Lewis, 2016, p. 17; Lundemir, 2021, p. 88). A saída desejada para o problema já é conhecida. O objetivo é aprender uma função que mapeie os dados de entrada para as saídas correspondentes, de modo a permitir que o modelo faça previsões ou tome decisões sobre novos dados, estes ainda não rotulados.

Um exemplo de aprendizagem de máquina supervisionada para o campo do Direito seria a classificação de um determinado conjunto de documentos jurídicos previamente rotulados como “fraude” ou “não fraude”. O treinamento utilizando esta base de dados teria como entrada os próprios documentos, sendo as etiquetas de “fraude” e “não fraude” as saídas, de forma a ensinar ao computador tal classificação, permitindo que novos documentos, não rotulados, sejam posteriormente classificados de maneira automática.

Entre as abordagens mais conhecidas para solucionar problemas de aprendizado supervisionado, destacam-se a regressão linear⁶, a regressão logística⁷, as redes neurais artificiais (RNNs)⁸, as máquinas de suporte vetorial (SVMs)⁹, as árvores de decisão¹⁰, os k-vizinhos mais próximos (KNN)¹¹ e o Naive Bayes¹². No entanto, neste trabalho não serão dados detalhes sobre cada uma dessas técnicas, já que isso ultrapassaria o escopo delineado. Em vez disso, a pesquisa é concentrada na aplicação dessas estratégias de IA no campo do Direito, ilustrando os avanços tecnológicos e a necessidade de uma abordagem interdisciplinar para uma compreensão completa desse tema.

Seguindo o raciocínio traçado, no que diz respeito à aprendizagem não supervisionada, os modelos são treinados com dados não rotulados (Igual; Seguí, 2017, p. 70; Lewis, 2016, p. 79; Lundemir, 2021, p. 88). Não há categorias pré-definidas. O objeto de tal técnica de aprendizagem é extrair informações úteis dentre os dados, bem como identificar padrões intrínsecos nos dados de entrada sem a orientação de saídas conhecidas.

Entre as aplicações da aprendizagem não supervisionada, temos o *clustering* (análise de agrupamento de dados), uma técnica de prospecção de dados cuja finalidade é realizar

⁶ Regressão Linear: É um método estatístico usado para modelar a relação entre uma variável dependente e uma ou mais variáveis independentes. A regressão linear busca encontrar a linha reta que melhor se ajusta aos dados observados.

⁷ Regressão Logística: É uma modalidade de análise de regressão usada para prever a probabilidade de uma variável categórica dependente. É comumente utilizada em problemas de classificação binária, onde a saída desejada é de natureza qualitativa.

⁸ Redes Neurais Artificiais (RNNs): São modelos computacionais compostos por unidades interconectadas, chamadas neurônios artificiais, que trabalham em conjunto para processar informações. Elas são capazes de aprender e reconhecer padrões complexos nos dados.

⁹ Máquinas de Suporte Vetorial (SVMs): São algoritmos de aprendizado de máquina que são usados para análise de dados de classificação e regressão. As SVMs encontram o hiperplano que melhor separa os dados em classes diferentes, maximizando a margem de separação.

¹⁰ Árvores de Decisão: São modelos de aprendizado de máquina que representam uma árvore de decisão onde cada nó interno representa um “teste” em um atributo, cada ramificação representa o resultado do teste e cada nó folha representa uma classe.

¹¹ K-Vizinhos Mais Próximos (KNN): É um algoritmo de aprendizado supervisionado usado para classificação e regressão. No KNN, os dados são classificados com base nas classes das instâncias vizinhas. O “k” representa o número de vizinhos mais próximos a serem considerados.

¹² Naive Bayes: É um algoritmo de classificação probabilístico baseado no teorema de Bayes, que assume independência entre os recursos. É amplamente utilizado em problemas de classificação de texto e mineração de texto.

agrupamentos automáticos de dados de acordo com o seu grau de semelhança (Igual; Seguí, 2017, p. 112). O *clustering* também é utilizado para a redução de dimensionalidade dos dados e para detecção de anomalias (detectar ações que fogem do padrão do conjunto de dados).

Nesta modalidade de aprendizagem, o algoritmo fica responsável por encontrar estruturas e padrões na base de dados e, a partir deles, classificar resultados, estabelecer regras associativas, fazer recomendações, detectar anomalias ou reduzir a dimensionalidade da base de dados.

Na área do Direito, um exemplo de aprendizagem de máquina não supervisionada seria a análise de documentos jurídicos para identificar temas ou tópicos comuns. Em um grande volume de documentos legais (como decisões judiciais ou petições), pode-se usar a técnica para entender quais são os principais assuntos abordados nesses documentos, sem que se necessite categorizá-los manualmente. Algoritmos de aprendizagem de máquina não supervisionada, como *clustering*, podem ser aplicados para agrupar documentos em *clusters* com base em similaridades textuais, auxiliando os profissionais da área jurídica a identificarem tendências e encontrar temas relevantes sem que seja necessária uma rotulação manual prévia.

Já a aprendizagem de máquina por reforço é a técnica que visa ensinar ao computador qual ação priorizar diante de determinada situação por uma vinculação de recompensas e punições a possíveis resultados (Lundemir, 2021, p. 89). Em outras palavras, a meta desse modelo de aprendizagem é recompensar uma ação considerada positiva e punir uma ação negativa (Faceli et al., 2011, p. 7).

A aprendizagem de máquina por reforço é um paradigma no qual um agente aprende a tomar decisões sequenciais em um ambiente dinâmico, visando maximizar uma recompensa acumulada ao longo do tempo. Nesse contexto, o agente interage com o ambiente, tomando ações e observando as consequências dessas ações, sendo recompensado ou penalizado com base nos resultados obtidos.

O processo de aprendizagem de máquina por reforço envolve três elementos principais: o agente, o ambiente e a função de recompensa. O agente é responsável por tomar decisões, o ambiente representa o mundo no qual o agente opera e a função de recompensa fornece feedback ao agente sobre a qualidade de suas ações.

No Direito, a técnica poderia ser utilizada para otimizar estratégias legais em um processo judicial. Por exemplo, temos um sistema inteligente que atua como um agente, representando um advogado, interagindo com um ambiente simulado que reproduz as etapas de um julgamento. O agente pode aprender a tomar decisões, como formular argumentos,

apresentar evidências, indicar provas ou fazer objeções, com o objetivo de maximizar a probabilidade de se sair vencedor na ação judicial. A função de recompensa poderia ser definida com fundamento no resultado do julgamento, onde vitórias teriam recompensas positivas e derrotas penalidades negativas.

Com o passar do tempo, o agente aprenderia a adaptar suas estratégias com base nas interações com o ambiente, melhorando sua eficácia na defesa ou acusação em processos judiciais (aprenderia a tomar decisões de maneira mais eficaz). Isso poderia potencialmente auxiliar advogados na preparação de casos e na tomada de decisões durante audiências e julgamentos.

Ainda, no contexto dos algoritmos empregados na área de IA, vale destacar a ampla utilização das RNNs, as quais têm alcançado resultados relevantes (Lundemir, 2021, p. 89). As redes artificiais buscam simular a estrutura e o funcionamento do cérebro humano (Salomão, 2022, p. 30; IBM, [20-?]), ao invés de realizar análises estatísticas convencionais. Enquanto os métodos estatísticos tradicionais assumem que as diferentes variáveis de entrada têm efeitos simples e independentes sobre as variáveis de saída que precisam ser explicadas, as RNNs operam distintamente: elas combinam as entradas de maneiras complexas para criar representações do fenômeno em questão, as quais, por sua vez, determinam outras características implícitas, culminando no resultado.

Uma RNN tem a capacidade de lidar com abstrações complexas, usando suas camadas intermediárias para aprender sobre a presença de características abstratas nos dados. Uma RNN simples consiste em múltiplas camadas de neurônios (Stryker, [20-?]). Os dados entram na rede através da camada de entrada e atravessam várias camadas intermediárias, onde a informação é processada. Finalmente, uma camada de saída produz os resultados desejados. Cada neurônio na rede possui um conjunto de pesos e uma função de ativação. Nas redes com camadas intermediárias, os sinais são combinados utilizando diferentes pesos e passados para a próxima camada. Nesse contexto, vale dizer que o aumento das camadas pode levar a uma maior precisão no resultado, embora sejam também necessários mais dados rotulados e maior infraestrutura computacional para processamento. Um maior número de camadas intermediárias indica a capacidade da rede de reconhecer aspectos mais detalhados e abrangentes nos dados de entrada.

O que se chama hoje de *deep learning* ou aprendizado profundo utiliza redes neurais artificiais com diversas camadas (redes neurais profundas) para aprender e representar padrões complexos nos dados (IBM, [20-?]; Stryker, [20-]). A sua arquitetura é mais complexa, com

muitas camadas de neurônios interconectados. A capacidade de ter múltiplas camadas permite que o modelo aprenda representações de dados mais complexas e abstratas. O objetivo dessa abordagem de aprendizagem de máquina é permitir que computadores aprendam automaticamente a representar e entender os dados de forma hierárquica, ou seja, aprendendo características cada vez mais abstratas e complexas à medida que avançam nas camadas da rede neural. Tais modelos, por exemplo, têm o poder de realizar tarefas como reconhecimento de imagens, processamento de linguagem natural, tradução automática, entre outras, sem a necessidade de programação manual de regras específicas.

Em resumo, a aprendizagem de máquina é uma modalidade de IA que envolve a criação de algoritmos que podem aprender, de forma automática, a partir de dados. O objetivo no desenvolvimento de uma IA é que seus algoritmos sejam capazes de resolver problemas de maneira tão eficiente quanto um ser humano e, por vezes, superando a capacidade humana.

Tratando sobre esta modalidade de IA, importante é a observação de Almeida e Doneda (2016), segundo a qual o avanço do poder computacional e o crescimento contínuo dos volumes de dados tornam possível que os algoritmos executem tarefas de grande escala e complexidade, frequentemente ultrapassando as capacidades humanas de controle e compreensão. Em certos casos, essa complexidade é tamanha que até mesmo os próprios programadores têm dificuldade em prever ou explicar completamente os resultados gerados.

Embora as técnicas de aprendizado de máquina sejam amplamente aplicadas no Direito, é crucial reconhecer que a qualidade e a neutralidade de suas decisões dependem diretamente dos dados utilizados e dos algoritmos desenvolvidos, que nem sempre estão isentos de viés.

2.3 O viés humano e da inteligência artificial: a neutralidade existe?

Neste subcapítulo, debruça-se sobre a complexa relação entre viés humano e IA, questionando a noção de neutralidade que frequentemente acompanha as tecnologias emergentes. A premissa de que sistemas de IA são isentos de preconceitos e discriminações é um mito que merece análise crítica, em especial quando se busca estabelecer códigos éticos e morais para um bom uso da IA.

Embora as tecnologias de IA sejam muitas vezes vistas como ferramentas objetivas, puramente geridas por códigos matemáticos, os algoritmos que a sustentam são desenvolvidos

por seres humanos, imbuídos de suas próprias experiências, crenças e, inevitavelmente, preconceitos.

Assim, analisa-se como os vieses presentes nas decisões humanas se infiltram nos dados e nos algoritmos, moldando as saídas (resultados) da IA, correndo-se elevado risco de perpetuação (e até mesmo criação de novas formas) de desigualdades, analisando, particularmente em termos de gênero.

O estudo é dividido em duas partes: primeiro, explora-se o que caracteriza e como se manifestam os vieses humanos nos processos de tomada de decisão, e em seguida, examina-se o impacto desses vieses nos sistemas de IA, fundamentando a crítica à suposta neutralidade dessas tecnologias. Ao longo desta análise, evidenciam-se as implicações éticas e sociais que emergem da interação entre viés humano e IA, propondo reflexões sobre a necessidade de abordagens mais conscientes e inclusivas no desenvolvimento e implementação de soluções tecnológicas nos mais variados campos da vida. Esse entendimento é essencial para garantir a aplicação da IA também no Direito, tal como se propõe em capítulo subsequente desta tese.

Adota-se a definição de “decisão” dada por Horta (2019, p. 66), uma vez que a explicação trazida pelo autor considera a literatura interdisciplinar que envolve teorias do julgamento e da tomada de decisão, economia comportamental, etologia, neurociências, biologia e psicologia evolucionista. Segundo Horta (2019a, p. 66), a “decisão” se caracteriza por um conjunto de processos cognitivos que visam avaliar e selecionar opções de acordo com um determinado curso de ação. Os processos de tomada de decisão podem ser alocados em quatro categorias (Horta, 2019a, p. 66):

Num primeiro momento, (i) a informação é disponibilizada, ou por meio da percepção sensorial, ou da memória. A seguir, (ii) ocorre o “julgamento”, que consiste na avaliação ou discriminação dessa informação ou estado de coisas. Com base nas suas preferências, o indivíduo (iii) escolhe, selecionando, a partir de diferentes opções, o curso de ação desejado, o que configura a “decisão”, no sentido estrito. A partir dessa decisão, resulta (iv) um comportamento ou ação no mundo exterior. [...] existem condicionantes de várias ordens em cada uma dessas etapas, derivadas de causas proximais e distais (vide Seção 3), que ajudam a explicar por que ocorrem vieses cognitivos.

Ainda, ressalta-se que a exposição trazida neste subcapítulo a respeito dos processos de tomada de decisão e vieses humanos empresta elementos e conceituações da Psicologia Comportamental e da Economia Comportamental. Isso se dá com a finalidade de esclarecer o que são e como ocorrem os vieses humanos, para então explicar o que são e como acontecem os vieses maquínicos.

2.3.1 O que são e como ocorrem os vieses humanos

A pesquisa no campo da Psicologia Comportamental sobre julgamento e tomada de decisão humana remonta às décadas de 1940 e 1950. Um dos marcos iniciais relevantes foi o trabalho desenvolvido por Von Neumann e Morgenstern, que atingiram notoriedade por sua contribuição ao campo da “Teoria dos Jogos”. Como explica Horta (2019, p. 69), em uma primeira fase, os estudos sobre julgamento e tomada de decisão humana se dividiam em duas áreas principais, que hoje estão mais interligadas: uma delas se concentrava em como as pessoas processam informações e fazem inferências (o “julgamento”), enquanto a outra investigava como elas escolhem entre diferentes opções de ação (a “tomada de decisão”).

Na época, o foco estava em desenvolver paradigmas experimentais capazes de testar e medir, a partir de uma perspectiva comportamental, como as pessoas de fato fazem julgamentos e tomam decisões, em comparação com os padrões normativos sugeridos pela racionalidade da Economia neoclássica, pelos cálculos probabilísticos e pela lógica dedutiva clássica, entre outros modelos. Logo, os primeiros experimentos começaram a questionar as previsões elaboradas pelos modelos teóricos então estabelecidos sobre comportamento decisório. No final da década de 1950, já se sabia que, em situações experimentais de resolução de problemas, os participantes tinham grande dificuldade em realizar cálculos estatísticos com precisão, ou mesmo em executar operações dedutivas simples (Horta, 2019, p. 70; Goldstein; Hogarth, 1997).

Herbert Simon (1955), por sua vez, sugeriu que fosse dada atenção às limitações inerentes ao processo decisório e à própria estrutura cognitiva humana, propondo o conceito de “racionalidade limitada”. Contudo, somente com a dificuldade em aplicar na prática a “Teoria da Utilidade” é que essas críticas começaram a ganhar mais atenção. Na década de 1960, economistas passaram a se dedicar ao estudo dos processos cognitivos envolvidos nos fenômenos observados. Foi nesse contexto que se consolidou a Economia Comportamental (Horta, 2019, p. 69-70; Goldstein; Hogarth, 1997).

A Economia Comportamental revela que as decisões humanas geralmente não são tomadas de forma racional e objetiva. Pelo contrário, as pessoas são frequentemente influenciadas por tendências cognitivas e motivadas por vieses na hora de tomar decisões. Foi na Economia Comportamental que pesquisadores como Kahneman e Tversky (1982) e Thaler (2016) começaram a investigar a fundo quais estratégias cognitivas (ou “heurísticas”) as pessoas utilizam para resolver problemas complexos de maneira rápida e intuitiva, mas que,

muitas vezes, resultam em erros sistemáticos ou desvios em relação aos padrões esperados (os “vieses”). Esse movimento visava reformular a teoria econômica, adotando modelos mais realistas sobre como os indivíduos formam preferências, avaliam informações disponíveis e fazem escolhas (Horta, 2019, p. 70).

O termo “viés cognitivo” foi introduzido na década de 1970 por Tversky e Kahneman e é definido por eles como um “erro sistemático no julgamento e na tomada de decisões comum a todos os seres humanos, que pode ser atribuído a limitações cognitivas, fatores motivacionais e/ou adaptações a ambientes naturais”. Os autores atribuem os vieses a operações mentais de atalhos: “pessoas se atêm a um número limitado de princípios heurísticos, os quais reduzem a tarefa complexa de acessar probabilidades e prever valores a simples operações de julgamento” (Tversky; Kahneman, 1974, p. 1124).

Conforme explicam Kahneman e Tversky, raramente tomamos decisões de maneira completamente objetiva e analítica. Nosso processo de tomada de decisões sofre o impacto de uma série de fatores que prejudicam nossa capacidade de realizar um julgamento puramente racional. Os autores identificaram duas formas de pensamento: uma rápida e intuitiva e outra mais lenta, que requer maior atenção. No entanto, mesmo em situações que demandariam uma análise mais profunda, as decisões podem sofrer influência de estereótipos, e a reflexão por si só não seria suficiente para eliminá-los ou contorná-los (Kahneman; Tversky, 1981; Tversky; Kahneman, 1982). Em outras palavras, na maioria das vezes, optamos pela forma rápida e intuitiva de pensar e, mesmo quando buscamos tornar nossa tomada de decisão mais racional, inconscientemente, nossas experiências e ideias influenciam nossas escolhas.

No modelo teórico de Kahneman (2013), o Sistema 1, mais rápido e intuitivo, demanda pouco esforço cognitivo, sendo utilizado quando se dispõe de pouca informação. Tal Sistema cria atalhos cognitivos, também denominados de heurísticas. As heurísticas podem ser entendidas como mecanismos de reconhecimento de informações, que auxiliam a encontrar respostas simples, mesmo que imperfeitas, para questionamentos que necessitariam de maior reflexão e esforço cognitivo.

O Sistema 1 tem como ponto de partida informações pretéritas e utiliza tais informações para responder, de modo rápido e automático, em situações semelhantes. Busca-se coerência, embora se confunda, em alguns casos, “familiaridade e conforto cognitivo com ilusões de veracidade”, em razão da insensibilidade “à quantidade e à qualidade da informação que estiver na gênese das impressões e intuições” (Wojciechowski; Rosa, 2018, p. 28-29).

Já o Sistema 2 requer esforço e atenção, possuindo vínculo com processos lógicos, sequenciais e conscientes, não obstante sejam lentos e, às vezes, ineficientes. É esse sistema que utilizamos para pensar de maneira abstrata, programando a “memória para obedecer instruções que encarnam tarefas não habituais”. Após as tarefas passarem a ser pelo Sistema 2 e se tornarem repetitivas, habituais, são realocadas ao Sistema 1, permitindo que sejam executadas de modo automático (Wojciechowski; Rosa, 2018, p. 35).

Como se observa, os dois Sistemas trabalham de forma cooperada, atendendo a demandas distintas. Entretanto, há situações em que, mesmo diante de questões complexas, aplicamos o Sistema 1, endossando generalizações de maneira equivocada. São nesses momentos que simplificações podem trazer resultados indesejados, levando a decisões subótimas. As heurísticas e os vieses ocorrem, assim, quando se recorre ao Sistema 1, enquanto se crê estar a fazer uso do Sistema 2, em busca da “melhor” e “mais rápida” solução (Wojciechowski; Rosa, 2018, p. 39).

Para além dos estudos de Kahneman e Tversky, muitas definições sobre viés podem ser encontradas na literatura científica. De modo geral, vieses cognitivos resultam no uso de atalhos mentais durante o processo de tomada de decisão humana. Frequentemente, esses atalhos são eficazes. Não obstante, há comprovação científica de que, entre seus efeitos, o mais evidente seria a “limitação de processamento” de informações.

Como observa Eyghen (2022, p. 1), alguns entendem o viés cognitivo como epistêmico. Ou seja, “como um padrão sistemático de desvio da norma ou racionalidade no julgamento” (Hasseltan et al., 2005 apud Eyghen, 2022, p. 1, tradução nossa¹³), ou como uma percepção subjetiva da realidade (Kahneman; Tversky, 1972 apud Eyghen, 2022, p. 1, tradução nossa)¹⁴. Também há aqueles que compreendem o viés como forma de cognição automática, ou seja, como uma forma de “processamento automático de informações” (Shiffrin; Schneider, 1977 apud Eyghen, 2022, p. 1, tradução nossa)¹⁵ ou “processamento de informações sem atenção” (Payne; Gawronski, 2010 apud Eyghen, 2022, p. 1, tradução nossa).¹⁶

Haseltan et al. (2016, p. 969) defendem que vieses cognitivos podem ter três origens distintas. Segundo os autores, uma origem seria a própria seleção de opções de decisão, que pode favorecer atalhos mentais úteis que funcionam bem na maioria das situações, embora não atendam totalmente a certos padrões normativos (“heurísticas”); vieses também podem surgir

¹³ No original: “systematic pattern of deviation from norm or rationality in judgment”.

¹⁴ No original: “top-down, subjective directed perception”.

¹⁵ No original: “automatic information processing”.

¹⁶ No original: “information processing without attention”.

quando determinada tarefa não é compatível com os tipos de problemas para os quais a mente foi adaptada (“artefatos”); e, por último, os vieses podem ocorrer se a resposta viesada resultar em menores custos de erro do que respostas imparciais (“viés de gestão de erros”). O Quadro 2 expõe a taxonomia evolutiva do viés cognitivo de acordo com os autores.

Quadro 2. Taxonomia evolutiva do viés cognitivo.

Taxonomia evolutiva do viés cognitivo	
Tipos de vieses	Exemplos
<i>Artefato</i> : vieses aparentes e erros são artefatos das estratégias de pesquisa; eles resultam da aplicação inapropriada de padrões normativos ou do posicionamento de seres humanos em ambientes não naturais.	Casos de negligência da taxa base na previsão estatística. Casos de viés de confirmação.
<i>Viés de gerenciamento de erro</i> : viés de seleção favorecendo o erro menos custoso; embora as taxas de erro aumentem, os custos líquidos são reduzidos.	Percepção auditiva. Xenofobia. Superpercepção sexual por homens. Subpercepção de compromisso por mulheres.

Fonte: Haselton et al., 2016, p. 969, adaptação e tradução nossas¹⁷.

No entanto, a explicação mais comum para o viés talvez seja a que relaciona a circunstância com o que se entende por heurística. Nesse contexto, o viés é entendido como um subproduto necessário do processamento limitado da mente humana: “Como o tempo e a capacidade de processamento de informações são limitados, os seres humanos precisam usar atalhos ou regras gerais que são propensos a falhas de maneira sistemática” (Haselton et al., 2016, p. 970, tradução nossa¹⁸).

Evolutionary Taxonomy of Cognitive Biases	
Type of Bias	Examples
<i>Artifact</i> : Apparent biases and errors are artifacts of research strategies; they result from the application of inappropriate normative standards or placement of humans in unnatural settings.	Some instances of base-rate neglect in statistical prediction Some instances of confirmation bias
<i>Error Management Bias</i> : Selection favored bias toward the less costly error; although error rates are increased, net costs are reduced.	Auditory looming Xenophobia Sexual overperception by men Commitment underperception by women

¹⁷ No original:

¹⁸ No original: “Perhaps the most commonly invoked explanation for bias is as a necessary by-product of processing limitations—because information processing time and ability are limited, humans must use shortcuts or rules of thumb that are prone to breakdown in systematic ways”.

De acordo com Peer e Gamliel (2013, p. 114), as heurísticas correspondem a estratégias mentais simplificadas, utilizadas como regras práticas que permitem às pessoas avaliar situações e decidir sem analisar todas as informações disponíveis. Assim, elas se apoiam em um conjunto restrito de indícios que tornam o processo de escolha mais ágil. Para os autores, essas estratégias surgem porque nossos recursos cognitivos e motivacionais são limitados, exigindo que sejam empregados de forma eficiente no cotidiano. No entanto, essa dependência de apenas parte das informações pode gerar vieses sistemáticos e previsíveis, conduzindo a decisões menos adequadas.

Existem diferentes tipos de heurísticas, entre as quais se destacam: a) da disponibilidade; b) da ancoragem e ajustamento; c) da representatividade; e d) do afeto.

A heurística da disponibilidade pode ser descrita como “o processo de avaliar a frequência de algo com base na facilidade com que exemplos ou ocorrências vêm à mente” (Kahneman, 2012, p. 166). Quando pensamos na prática jurídica, por exemplo, diversas influências podem afetar o julgamento de um juiz durante o processo decisório. Um exemplo é a influência de casos que recebem grande atenção da mídia, o que pode afetar como um magistrado os encara.

Quanto à heurística da ancoragem, ela pode ser entendida como “a tendência das pessoas a se fiar a um valor particular – o primeiro valor disponível a elas – para estimar uma quantidade desconhecida, ajustando a partir desse valor. Em outras palavras, esse valor particular funciona como âncora – um padrão de referência” (Wojciechowski; Rosa, 2018, p. 39-40). Segundo Kahneman (2012, p. 153), a ancoragem pode ter o efeito de sugestão (*priming*) e também pode resultar em ajustes insuficientes. No contexto jurídico, essa heurística poderia se manifestar, por exemplo, se um juiz decidisse um caso muito complexo e volumoso antes de julgar outro caso mais simples. A atenção e os recursos dispendidos no primeiro processo poderiam ser muito mais intensos, enquanto no segundo poderiam ser substancialmente menores (Bertoni e Amaral, 2018).

Já a heurística da representatividade refere-se à tendência das pessoas a avaliar a probabilidade de um evento “B” com base na semelhança de “A” com “B” (Tversky; Kahneman, 1974). Sob essa ótica, a heurística está relacionada à utilização de estereótipos para formar julgamentos.

A heurística do afeto, por sua vez, refere-se à maneira como os indivíduos tomam decisões com base em suas emoções ou crenças subjetivas sobre o mundo.

Além disso, a decisão humana – e, por consequência, também a decisão de um magistrado – está sujeita a diversos vieses cognitivos que podem comprometer a imparcialidade e a racionalidade esperadas no processo decisório. Um exemplo é o viés de confirmação, no qual o indivíduo tende a buscar, selecionar e valorizar apenas informações que sustentem suas crenças, hipóteses ou expectativas já formadas, ao mesmo tempo em que desconsidera ou minimiza dados que poderiam contrariá-las (Bertoni; Amaral, 2018). Esse mecanismo cognitivo funciona como uma espécie de filtro seletivo da realidade, que reforça convicções pré-existentes e dificulta a abertura para perspectivas divergentes, podendo levar a julgamentos parciais e pouco fundamentados.

Outro viés recorrente é o viés retrospectivo, caracterizado pela crença de que, após o desfecho de uma situação, sempre se soube qual seria o resultado. Esse fenômeno, conhecido como efeito “eu sempre soube”, cria uma falsa sensação de previsibilidade e inevitabilidade, distorcendo a percepção sobre a complexidade real do processo decisório. No âmbito judicial, tal distorção pode fazer com que um magistrado ou mesmo a sociedade interpretem determinadas decisões como óbvias ou esperadas, quando, na prática, foram fruto de escolhas incertas e baseadas em informações limitadas.

De maneira geral, pode-se afirmar que as heurísticas são mecanismos cognitivos usados para simplificar o processo decisório. No entanto, o uso excessivo ou inadequado dessas ferramentas pode gerar vieses que acabam distorcendo as escolhas. O Quadro 3 ilustra os vieses que podem surgir em função do uso de heurísticas.

Quadro 3. A relação entre vieses e heurísticas.

Relação entre vieses e heurísticas	
Viés	Descrição
<i>Vieses derivados da Heurística da Disponibilidade</i>	
1. Facilidade de recordação	Indivíduos julgam eventos que são mais facilmente lembrados da memória, com base na vividez ou na atualidade, como sendo mais numerosos do que eventos de igual frequência cujos exemplos são mais difíceis de recordar.
2. Recuperabilidade	Indivíduos são tendenciosos em suas avaliações sobre a frequência de eventos com base em como a estrutura de sua memória afeta o processo de busca.
<i>Vieses derivados da Heurística da Representatividade</i>	

3. Insensibilidade à taxa base	Ao avaliar a probabilidade de eventos, os indivíduos tendem a ignorar as taxas base se qualquer outra informação descritiva for fornecida, mesmo que seja irrelevante.
4. Insensibilidade ao tamanho da amostra	Ao avaliar a confiabilidade das informações de uma amostra, os indivíduos frequentemente falham em reconhecer o papel do tamanho da amostra.
5. Mitos sobre o acaso	Indivíduos esperam que uma sequência de dados gerada por um processo aleatório pareça “aleatória”, mesmo quando a sequência é muito curta para que essas expectativas sejam estatisticamente válidas.
6. Regressão à média	Indivíduos tendem a ignorar o fato de que eventos extremos tendem a regressar à média em tentativas subsequentes.
7. A falácia da conjunção	Indivíduos julgam erroneamente que conjunções (dois eventos ocorrendo simultaneamente) são mais prováveis do que um conjunto mais amplo de ocorrências das quais a conjunção é um subconjunto.
<i>Vieses derivados da Heurística da Confirmação</i>	
8. A armadilha da confirmação	Indivíduos tendem a buscar informações que confirmem o que acreditam ser verdadeiro e falham em procurar evidências que possam contradizer suas crenças.
9. Ancoragem	Indivíduos fazem estimativas de valores com base em um valor inicial (derivado de eventos passados, atribuição aleatória ou qualquer informação disponível) e, geralmente, fazem ajustes insuficientes a partir dessa âncora ao estabelecer um valor final.
10. Viés de eventos conjuntivos e disjuntivos	Indivíduos apresentam um viés ao superestimar a probabilidade de eventos conjuntivos (onde dois ou mais eventos acontecem juntos) e subestimar a probabilidade de eventos disjuntivos (onde ao menos um dos eventos ocorre).
11. Visão retrospectiva e a maldição do conhecimento	Após descobrir se um evento ocorreu ou não, os indivíduos tendem a superestimar o grau em que teriam previsto o resultado correto. Além disso, falham em ignorar as informações que possuem

	e que outros não têm ao prever o comportamento dos outros.
12. Excesso de confiança	Indivíduos tendem a ser excessivamente confiantes na infalibilidade de seus julgamentos ao responder a questões moderadamente ou extremamente difíceis.

Fonte: Barzeman e Moore, 2013, p. 58, adaptação e tradução nossas¹⁹.

Bias	Description
Biases Emanating from the Availability Heuristic	
1. Ease of recall	Individuals judge events that are more easily recalled from memory, based on vividness or recency, to be more numerous than events of equal frequency whose instances are less easily recalled.
2. Retrievability	Individuals are biased in their assessments of the frequency of events based on how their memory structures affect the search process.
Biases Emanating from the Representativeness Heuristic	
3. Insensitivity to base rates	When assessing the likelihood of events, individuals tend to ignore base rates if any other descriptive information is provided—even if it is irrelevant.
4. Insensitivity to sample size	When assessing the reliability of sample information, individuals frequently fail to appreciate the role of sample size.
5. Misconceptions of chance	Individuals expect that a sequence of data generated by a random process will look “random,” even when the sequence is too short for those expectations to be statistically valid.
6. Regression to the mean	Individuals tend to ignore the fact that extreme events tend to regress to the mean on subsequent trials.
7. The conjunction fallacy	Individuals falsely judge that conjunctions (two events co-occurring) are more probable than a more global set of occurrences of which the conjunction is a subset.
Biases Emanating from the Confirmation Heuristic	
8. The confirmation trap	Individuals tend to seek confirmatory information for what they think is true and fail to search for disconfirmatory evidence.
9. Anchoring	Individuals make estimates for values based upon an initial value (derived from past events, random assignment, or whatever information is available) and typically make insufficient adjustments from that anchor when establishing a final value.
10. Conjunctive and disjunctive events bias	Individuals exhibit a bias toward overestimating the probability of conjunctive events and underestimating the probability of disjunctive events.
11. Hindsight and the curse of knowledge	After finding out whether or not an event occurred, individuals tend to overestimate the degree to which they would have predicted the correct outcome. Furthermore, individuals fail to ignore information they possess that others do not when predicting others’ behavior.
12. Overconfidence	Individuals tend to be overconfident of the correctness of their judgments, especially when answering difficult questions.

¹⁹ No original:

Então, a heurística cognitiva, muito embora represente um mecanismo de redução de tempo e esforço no processo de julgamento e na tomada de decisões no dia a dia, o processo de simplificação do pensamento pode levar a erros e vieses (Tonetto et al., 2006, p. 187).

Em grande parte das descrições do viés, os seres humanos frequentemente permanecem inconscientes de suas operações. Pode-se afirmar, portanto, que o viés cognitivo se manifesta sem que o ser humano tenha plena consciência da sua ocorrência. Ademais, embora seres humanos possam ser conscientizados da existência e ocorrência de vieses, por meio da introspecção ou de informações externas, por exemplo, explicar por qual motivo sua cognição está inclinada em uma direção específica, muitas vezes, se revela desafiador.

Há, contudo, divergências nas explicações dos vieses cognitivos. Abordagens filogenéticas podem ser encontradas na literatura da psicologia evolutiva e, em resumo, defendem que muitos vieses são inerentes à estrutura cerebral humana. Sugere-se que a maioria deles resulta da estruturação da mente humana em toda a espécie. Ou seja, em razão da maneira como cérebros humanos são estruturados, diversos vieses se originam. Nessa concepção, apesar de alguns vieses serem desencadeados por estímulos sensoriais, a resposta a esses estímulos é “pré-programada” ou predeterminada pela composição cognitiva humana. Nesse contexto, o estímulo sensorial não é o principal fator determinante para o surgimento do viés (Eyghen, 2022, p. 1).

Já na abordagem ontogenética, que prevalece na literatura da psicologia evolutiva (Haselton et al., 2016, p. 968-969), sugere-se que grande parte dos vieses não nascem com o ser humano, ou seja, não são inatos à sua estrutura cerebral. Os vieses seriam adquiridos (desenvolvidos) a partir da interação da mente humana com o ambiente. Por exemplo, vieses culturais teriam, incontestavelmente, origem do ambiente em que vivemos. Eles não estariam enraizados em estruturas cognitivas inatas, mas dependem de estímulos externos. Em outras palavras, essa modalidade de viés depende da exposição do ser humano a experiências sensoriais (Eyghen, 2022, p. 2).

Observa-se assim, que alguns vieses são inatos, enquanto outros se manifestam gradualmente à medida que nossos cérebros amadurecem e interagem com o ambiente em que vivemos. Tanto na abordagem filogenética quanto na ontogenética, tem-se que os vieses apresentam uma fenomenologia semelhante, com padrões recorrentes na cognição que são difíceis de superar. O que difere são as observações empíricas e “enquanto a explicação ontogenética sugere uma maior variação de vieses dependendo do ambiente cultural, ela também prevê mais variação individual dentro dos grupos” (Eyghen, 2022, p. 2). Assim, a

interação entre fatores inatos e ambientais parece ser crucial para a manifestação e a diversidade dos vieses cognitivos.

Ainda, há estudos que apontam que a percepção individual daqueles que ocupam posições de grande poder em hierarquias sociais é menos precisa do que a de aqueles que estão no escalão social mais baixo (Fiske, 1993). Aqueles que estão no topo da hierarquia possuem maior probabilidade de endossar estereótipos sobre outras pessoas do que se atentar para a individualização de informações relacionadas especificamente à pessoa que está sendo avaliada, o que poderia tomar sua tomada de decisão mais precisa (Goodwin et al., 2000).

Nos últimos anos, diversos estudos mostram que os vieses cognitivos não são somente um fenômeno individual, mas também têm implicações sociais. Pesquisadores têm destacado as consequências da racionalidade viesada no âmbito coletivo (Horta, 2019, p. 71; Thaler; Sunstein, 2008).

O que se vê é que os diferentes modelos de julgamento e tomada de decisão têm em comum a ideia de que existem parâmetros normativos para a realização de inferências, como a lógica dedutiva clássica, a aritmética e as regras do cálculo probabilístico. No entanto, esses parâmetros não são, de fato, os que as pessoas utilizam em suas decisões no cotidiano. Em outras palavras, ao analisarmos o comportamento humano descritivamente, podemos observar que a racionalidade das decisões foge dos padrões normativos esperados (Horta, 2019, p. 70-71; Baron, 2008, p. 31-34).

Neste trabalho, adotamos o conceito de viés cognitivo como percepções distorcidas de fatos ou conclusões distorcidas a seu respeito em razão de crenças ou percepções subjetivas, desconsiderando a sua origem, se ontogenética ou filogenética, por exemplo. Dito em outras palavras, adota-se o viés como uma inclinação injusta ou preconceito a favor ou contra uma pessoa ou grupo.

Para que vieses sejam, de certa forma, superados, inicialmente é imprescindível o seu reconhecimento. Somente é possível realizar escolhas mais racionais e informadas quando se reflete a respeito das variáveis envolvidas e dos preconceitos infiltrados no processo de tomada de decisão. É por esse motivo que o presente trabalho serve de alerta: se nenhuma perspectiva pode, de fato, assumir a pretensão de neutralidade e ser puramente imparcial, visto que do ponto de vista daquele que formula a decisão, a formula com base em suas experiências e vivências, deve-se incentivar a reflexão sobre os mais variados tipos de vieses e seus impactos reais.

Compreender as limitações dos processos cognitivos humanos é fundamental para analisar os vieses presentes em algoritmos. A acurácia de um sistema de IA está diretamente

relacionada à qualidade dos dados utilizados em seu treinamento; assim, informações viesadas levarão a resultados igualmente falhos. Dependendo de como são programados, os algoritmos não apenas replicarão o comportamento decisório humano (seja ele viesado ou não). Para além disso, poderão também desenvolver seus próprios vieses e até intensificar determinadas distorções.

2.3.2 *A inteligência artificial e o mito da neutralidade: vieses nos dados e nos algoritmos*

No subcapítulo anterior, verificou-se que o conceito de viés cognitivo é a tendência humana de perceber fatos ou tirar conclusões distorcidas a respeito deles. Ele é um “preconceito a favor ou contra algo ou alguém, que pode resultar em decisões injustas” (High-Level Expert Group on Artificial Intelligence, 2018, p. iv, tradução nossa²⁰).

O viés humano reside nas associações implícitas que afetam a tomada de decisões, e essa questão se estende aos sistemas de IA. Uma decisão tendenciosa revela uma perspectiva unilateral, influenciada pela visão de mundo, experiência, valores e até mesmo intuição de um indivíduo ou grupo em relação ao contexto em que estão inseridos. Indivíduos podem desenvolver vieses a favor ou contra diversos elementos, como uma característica pessoal específica, um grupo étnico, uma orientação sexual, uma nação, uma religião, uma classe social, uma ideologia, entre outros. Nesse sentido, os vieses algorítmicos representam o fenômeno pelo qual as pessoas incorporam suas visões de mundo e, muitas vezes, preconceitos às tecnologias.

Em uma definição estatística, o viés seria a “diferença entre o valor estimado ou previsto e o valor real”. Em muitos casos, o viés estatístico presente na IA está associado a resultados socialmente tendenciosos, que refletem e até reforçam desigualdades pré-existentes na sociedade (Accessnow.Org, 2018, p. 11). Essas distorções podem impactar negativamente a tomada de decisões automatizadas, gerando consequências injustas e perpetuando discriminações.

Um sistema computacional viesado pode ser definido como aquele que de maneira “sistemática e injusta discrimina certos indivíduos ou grupos em favor de outros”. A discriminação ocorre quando o sistema nega uma oportunidade ou um benefício, ou se atribui “um resultado indesejável a um indivíduo ou a um grupo com base em motivos que são irrazoáveis ou inadequados” (Friedman; Nissenbaum, 1996, p. 332, tradução nossa²¹).

²⁰ No original: “Bias is a prejudice for or against something or somebody, that may result in unfair decisions”.

²¹ No original: “[...] we use the term bias to refer to computer systems that *systematically* and *unfairly discriminate* against certain individuals or groups of individuals in favor of others. A system discriminates unfairly if it denies

É incontestado que seres humanos estão sujeitos a tomar decisões viesadas. Assim, se sistemas inteligentes são projetados por seres humanos, tais sistemas também não estão livres de vieses. Preconceitos e discriminações podem, mesmo que de forma não intencional, ser inseridos na IA através de algoritmos ou dados desbalanceados, não inclusivos (High-Level Expert Group on Artificial Intelligence, 2018, p. iv; Freitas, 2025, p. 178-198).

A IA pode carregar consigo erros de diversas origens. Não se pode deixar de lado que seus resultados são probabilidades e, como tal, há incertezas. Dentre os principais problemas trazidos com o avanço da aplicação das técnicas de IA, têm-se a discriminação e o viés dos sistemas, o risco à privacidade e a opacidade algorítmica (Monteiro, 2021, p. 47-48). A crescente utilização da IA pode aumentar significativamente os efeitos negativos de erros que, aparentemente, podem parecer pequenos. Resultados equivocados de sistemas de IA podem violar direitos humanos de várias formas (Human Rights Council, 2021).

O tema sobre vieses algoritmos ganha especial relevância ao se discutir soluções de problemas por meio de IA. Algoritmos de aprendizagem de máquina, por exemplo, fazem previsões e tomam decisões a partir de seu aprendizado. Na abordagem da aprendizagem de máquina, vieses podem ser originados por diversos fatores durante a etapa de seu desenvolvimento. Eles podem ocorrer em razão da seleção de dados para treinamento (*inputs*) e também podem surgir na aplicação da técnica final em contextos distantes da sua função originária, por exemplo. Eles podem ser inconscientes ou até mesmo intencionais.

Quando afirmamos que o aprendizado de máquina se baseia no raciocínio indutivo, que consiste em extrair regras e padrões de grandes conjuntos de dados, é comum associá-lo à estatística, o que poderia sugerir uma certa neutralidade social, já que se apoia em princípios matemáticos para resolver problemas. No entanto, as previsões baseadas em dados nem sempre resultam em decisões neutras. Pelo contrário, frequentemente observamos o oposto: a tecnologia não apenas pode perpetuar preconceitos e discriminações, mas também ampliá-los, dada sua escalabilidade.

Diversos sistemas de IA têm mostrado tendência a reproduzir e até amplificar vieses presentes na sociedade, muitas vezes como resultado de dados históricos desbalanceados ou discriminatórios. Um exemplo ocorre no campo do reconhecimento facial, onde algoritmos têm demonstrado desempenho inferior ao tentar identificar pessoas de pele escura ou mulheres. Esse viés é geralmente atribuído ao fato de que muitos desses sistemas foram treinados com

an opportunity or a good or if it assigns an undesirable outcome to an individual or group of individuals on grounds that are unreasonable or inappropriate”.

conjuntos de dados que contêm uma representação desproporcional de rostos brancos e masculinos, o que leva a uma maior precisão na identificação de indivíduos dessas características e falhas ao lidar com outras etnias ou gêneros. Isso resulta não somente em imprecisão, mas também em uma discriminação implícita, já que o sistema acaba favorecendo um grupo em detrimento de outros (Buolamwini; Gebru, 2018).

Outro exemplo de viés ocorre no setor financeiro, especificamente em sistemas de concessão de crédito e empréstimos. Algoritmos usados por instituições financeiras para decidir sobre o crédito frequentemente refletem e amplificam desigualdades históricas, como discriminação racial ou socioeconômica. Quando os dados históricos usados para treinar esses sistemas incluem informações que discriminam minorias raciais ou pessoas de classes sociais mais baixas, os algoritmos podem reproduzir esses mesmos padrões, resultando em decisões de crédito injustas. Ou seja, esses sistemas podem, inadvertidamente, negar crédito a indivíduos com base em características como raça ou origem social, perpetuando a exclusão econômica e dificultando o acesso a oportunidades para grupos marginalizados (O'Neil, 2020).

Um exemplo diretamente relacionado ao campo do Direito diz respeito ao sistema de justiça criminal. Algoritmos empregados para avaliar o risco de reincidência ou determinar a concessão de liberdade condicional também podem incorporar viés racial, principalmente quando alimentados por dados históricos de criminalização de determinados grupos, como negros. Muitas vezes, esses sistemas são baseados em informações de prisões anteriores ou de práticas de policiamento, desproporcionais em relação a certas populações. Como resultado, pessoas pertencentes a esses grupos podem ser avaliadas de forma mais punitiva, com maior probabilidade de receber sentenças mais severas ou de ser negada a liberdade condicional, refletindo um ciclo de discriminação que remonta à desigualdade racial no sistema judicial (Angwin et al., 2016; Skeem; Lowenkamp, 2016).

Nesse contexto, o *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS) é um exemplo de aprendizagem de máquina aplicada ao Direito. Desenvolvido pela empresa NORTHPOINTE em 2012, o sistema gerou ampla discussão ao estabelecer critérios para prever a probabilidade de reincidência criminal de indivíduos. Embora tenha sido implementado nos Estados Unidos, o COMPAS foi rejeitado pela Suprema Corte devido a evidências de discriminação racial. Um estudo realizado por Larson et al. (2016) revelou que as chances de reincidência para pessoas negras eram, em média, duas vezes mais altas do que para os demais grupos. Apesar de o sistema ter sido auditado e não utilizar explicitamente a cor da pele como variável, constatou-se que ele ainda atribuía probabilidades

mais altas de reincidência a indivíduos negros. Esse viés pode ser explicado por dois fatores principais: (a) o sistema considerava informações sobre o local de residência do indivíduo, e, nos Estados Unidos, áreas predominantemente habitadas por pessoas negras tendem a ter características que facilitam essa segmentação, e (b) surgiram suspeitas de viés na equipe responsável pelo pré-processamento e limpeza dos dados, que, de certa maneira, pode ter incorporado informações relacionadas à cor da pele em outras variáveis utilizadas.

Em Leavy (2018), se identificam vieses de gênero na sobrerrepresentação de homens no desenvolvimento de sistemas de IA. No artigo, destaca-se que algoritmos de IA podem perpetuar e amplificar desigualdades de gênero existentes na sociedade, especialmente quando treinados com dados históricos que refletem tais desigualdades. Argumenta-se que a falta de diversidade nas equipes de desenvolvimento e a ausência de uma compreensão crítica das questões de gênero contribuem para a reprodução desses vieses nos sistemas de IA.

Esses exemplos demonstram como o viés em IA pode ser introduzido de variadas formas, de maneira sutil, mas poderosa, reforçando desigualdades existentes em várias áreas da sociedade.

A inclusão de viés nos algoritmos pode ocorrer de diversas maneiras, desde o treinamento de dados que refletem desigualdades sociais e históricas até a presença de termos mais associados a determinados grupos. Quando o viés se encontra nos dados ou ocorre no nível de *input* do sistema inteligente, ele pode ocorrer, por exemplo, em razão do uso de dados históricos tendenciosos, da falta de representatividade nos dados ou também em função da má rotulagem dos dados (Accessnow.Org, 2018, p. 12):

- a)** Uso de dados históricos tendenciosos: uma vez que os sistemas de aprendizagem de máquina usam base de dados para identificar padrões, quaisquer vieses encontrados nesses dados serão reproduzidos pela máquina.
- b)** Falta de representatividade nos dados: quando os dados de entrada do sistema não são representativos da população-alvo, ou seja, quando há viés de seleção. O sistema treinado com viés de seleção naturalmente apresentará recomendações que favorecerão determinado grupo em relação a outro.
- c)** Dados mal selecionados ou não rotulados da maneira correta: quando os dados que comporão determinado sistema são incompletos,

incorretos ou desatualizados, os resultados fornecidos pela máquina serão imprecisos.

O que se verifica é que muitas das ocorrências de viés estão relacionadas aos dados coletados e ao modo em que são realizados o processamento e o treinamento de dados. A título exemplificativo, traz-se o Quadro 4, que identifica as principais formas pelas quais o viés pode ser incorporado a um sistema de IA:

Quadro 4. Exemplos de incorporação de viés em sistemas de IA.

Exemplos de incorporação de viés em sistemas de IA	
Viés nos dados de treinamento do sistema	Os sistemas de IA aprendem a partir de grandes volumes de dados. Se os dados de treinamento forem viesados, o modelo também será. Esse viés pode ocorrer devido a: a) <i>Seleção de dados não representativos</i>: se o conjunto de dados usado para treinar a IA não for representativo da população ou contexto em questão, o modelo pode produzir resultados tendenciosos. Por exemplo, se um algoritmo de reconhecimento facial for treinado predominantemente com imagens de pessoas de uma única etnia, ele pode não funcionar corretamente para outras etnias, resultando em falhas ou discriminação. b) <i>Dados históricos tendenciosos</i>: em muitos casos, os dados utilizados para treinar modelos de IA refletem práticas e decisões históricas, que podem conter preconceitos ou discriminações existentes na sociedade. Por exemplo, sistemas de IA usados em processos seletivos podem aprender a priorizar candidatos com base em históricos de contratações que, muitas vezes, favorecem certos grupos (como homens para cargos em tecnologia) ou negligenciam outros (como mulheres ou minorias raciais). c) <i>Rotulagem de dados</i>: a forma como os dados são rotulados também pode introduzir viés. Se os rótulos forem atribuídos de maneira subjetiva ou com base em preconceitos humanos, isso afetará diretamente os resultados do modelo. Por exemplo, se um sistema de moderação de conteúdo for treinado com

	exemplos rotulados de maneira tendenciosa (como classificar certos tipos de discurso como “ofensivos” com base em estereótipos), ele pode replicar e amplificar esses preconceitos.
Viés de modelagem do sistema (viés algorítmico)	O viés pode também ser inserido na própria construção dos modelos de IA, durante a escolha de algoritmos, parâmetros e métodos de aprendizado. Isso inclui: a) <i>Escolha de algoritmos ou modelos inadequados</i>: a escolha de certos algoritmos pode impactar como os dados são processados e interpretados. Por exemplo, algoritmos que não são bem calibrados para lidar com dados desbalanceados (onde uma classe é muito mais representada que outra) podem criar modelos que favorecem a classe majoritária, ignorando a minoria. b) <i>Vieses implícitos nos parâmetros</i>: mesmo quando um modelo é desenvolvido para ser o mais “neutro” possível, parâmetros e decisões como a escolha de <i>features</i> (variáveis) podem introduzir viés. Por exemplo, se um modelo for projetado para avaliar crédito com base em dados históricos e inclui variáveis como raça ou gênero, isso pode perpetuar discriminação.
Viés na interpretação e na implementação dos resultados	Mesmo que o modelo de IA tenha sido treinado de forma objetiva, o viés pode surgir durante a implementação ou interpretação dos resultados: a) <i>Feedback loop (ciclo de retroalimentação)</i>: quando as decisões do modelo de IA influenciam os dados de entrada para modelos futuros, pode ocorrer um ciclo de retroalimentação que amplifica o viés. Por exemplo, se um sistema de recomendação de filmes ou produtos sugere conteúdos predominantemente de um grupo demográfico específico, esses usuários podem interagir mais com o sistema, criando mais dados sobre esse grupo e reforçando o viés existente. b) <i>Falta de transparência e explicabilidade</i>: quando o processo de decisão de um sistema de IA não é transparente ou não pode ser explicado claramente (uma característica conhecida como

	<i>caixa-preta</i>), é mais difícil identificar e corrigir vieses. Isso é especialmente problemático em áreas como justiça criminal, saúde ou recrutamento, onde decisões automatizadas podem ter consequências significativas na vida das pessoas.
Viés por interpretação de contextos culturais, políticos ou sociais	A IA também pode ser influenciada por contextos culturais, sociais ou políticos, que, se não forem bem compreendidos ou considerados, podem resultar em decisões viesadas. Isso é particularmente relevante em sistemas que operam em contextos globais ou em áreas sensíveis, como redes sociais ou saúde pública, onde normas culturais, políticas ou sociais podem afetar as escolhas e o comportamento de um modelo.

Fonte: a autora (com uso do ChatGPT para organização em quadro, atendendo ao disposto no “Guia orientador: Reflexão sobre o uso da inteligência artificial na pesquisa acadêmica” (UNIVERSIDADE ESTADUAL DE PONTA GROSSA, [2025?])).

Bastante didática é a categorização feita por Friedman e Nissenbaum (1996, p. 334-335) em *Bias em Computer Systems*, de onde se retira o Quadro 5, que classifica os vieses em sistemas computacionais em três tipos principais: pré-existent, técnicos e emergentes.

Conforme apresentado no referido quadro, os vieses pré-existent têm origem em práticas, atitudes e estruturas sociais que antecedem a criação do sistema. Eles podem ser incorporados de forma consciente ou inconsciente por indivíduos diretamente envolvidos em seu desenvolvimento, como desenvolvedores ou clientes, ou ainda refletir padrões sociais mais amplos, como aqueles relacionados a gênero, raça ou classe social.

O viés técnico, por sua vez, decorre de limitações ou decisões de ordem técnica tomadas ao longo do processo de desenvolvimento do sistema. Essas limitações podem envolver restrições de hardware ou software, algoritmos que tratam de forma desigual diferentes grupos em determinados contextos, falhas em geradores de números aleatórios ou, ainda, tentativas de formalizar conceitos humanos complexos – como julgamentos legais ou práticas sociais – em regras computacionais rígidas, o que pode gerar distorções relevantes.

Já o viés emergente manifesta-se no uso efetivo do sistema, geralmente após sua implementação, em virtude de mudanças no conhecimento social, na composição do público ou nos valores culturais. Esse tipo de viés ocorre, por exemplo, quando os usuários apresentam níveis de conhecimento, hábitos ou valores culturais distintos daqueles presumidos no momento

do *design*, o que pode comprometer tanto a eficácia quanto a equidade das funcionalidades e interfaces do sistema.

Quadro 5. Categorias de viés em sistemas computacionais.

Categorias de viés em sistemas computacionais
Essas categorias descrevem maneiras pelas quais o viés pode surgir no <i>design</i> de sistemas computacionais. Os exemplos ilustrativos retratam casos plausíveis de viés. <i>1. Viés pré-existente</i> O viés pré-existente tem suas raízes em instituições sociais, práticas e atitudes. Quando sistemas computacionais incorporam vieses que existem independentemente, e geralmente antes da criação do sistema, então o sistema exemplifica o viés pré-existente. O viés pré-existente pode entrar em um sistema tanto através dos esforços explícitos e conscientes de indivíduos ou instituições, quanto implícita e inconscientemente, mesmo apesar das melhores intenções. <i>1.1. Individual</i> Viés que se origina de indivíduos que têm uma contribuição significativa no <i>design</i> do sistema, como o cliente que encomenda o <i>design</i> ou o <i>designer</i> do sistema (por exemplo, um cliente incorpora vieses raciais pessoais nas especificações em software de aprovação de empréstimos). <i>1.2. Societal</i> Viés que se origina da sociedade em geral, como de organizações (por exemplo, indústria), instituições (por exemplo, sistemas legais) ou da cultura em geral (por exemplo, vieses de gênero presentes na sociedade em geral que levam ao desenvolvimento de software educacional que, em geral, atrai mais meninos do que meninas). <i>2. Viés técnico</i> O viés técnico surge de restrições técnicas ou considerações técnicas. <i>2.1. Ferramentas computacionais</i> Viés que se origina de uma limitação da tecnologia computacional, incluindo hardware, software e periféricos (por exemplo, em um banco de dados para correspondência de doadores de órgãos com potenciais receptores de transplante, certos indivíduos recuperados e exibidos nas telas iniciais são sistematicamente favorecidos para uma correspondência em relação aos indivíduos exibidos em telas posteriores). <i>2.2 Algoritmos descontextualizados</i> Viés que se origina do uso de um algoritmo que não trata todos os grupos de forma justa em todas as condições significativas (por exemplo, um algoritmo de

agendamento que agenda aviões para decolagem se baseia na listagem alfabética das companhias aéreas para ordenar os voos prontos em um determinado tempo).

2.3 Geração de números aleatórios

Viés que se origina de imperfeições na geração de números pseudoaleatórios ou no uso incorreto de números pseudoaleatórios (por exemplo, uma imperfeição em um gerador de números aleatórios usado para selecionar receptores para um medicamento escasso leva sistematicamente a favorecer indivíduos no final do banco de dados).

2.4 Formalização de constructos humanos

Viés que se origina de tentativas de tornar constructos humanos como discurso, julgamentos ou intuições passíveis de serem tratados por computadores: quando quantificamos o qualitativo, discretizamos o contínuo ou formalizamos o não formal (por exemplo, um sistema jurídico expert aconselha réus se devem ou não aceitar um acordo, assumindo que a lei pode ser definida de maneira inequívoca que não está sujeita a interpretações humanas e humanizadas em contexto).

[...]

3. *Viés emergente*

O viés emergente surge em um contexto de uso com usuários reais. Esse viés geralmente surge algum tempo após a conclusão do *design*, como resultado de mudanças no conhecimento social, na população ou nos valores culturais. As interfaces do usuário são particularmente propensas ao viés emergente porque, por *design*, buscam refletir as capacidades, características e hábitos dos usuários potenciais. Assim, uma mudança no contexto de uso pode criar dificuldades para um novo conjunto de usuários.

3.1 Novo conhecimento social

Viés que se origina do surgimento de novos conhecimentos na sociedade que não podem ser ou não são incorporados ao *design* do sistema (por exemplo, um sistema especialista médico para pacientes com AIDS não tem mecanismo para incorporar descobertas médicas de ponta que afetam como indivíduos com certos sintomas devem ser tratados).

3.2 Incompatibilidade entre usuários e *design* do sistema

Viés que se origina quando a população que usa o sistema difere em alguma dimensão significativa da população assumida como usuária no *design*.

3.2.1 Diferente expertise

Viés que se origina quando o sistema é usado por uma população com uma base de conhecimento diferente daquela assumida no *design* (por exemplo, um caixa eletrônico com uma interface que faz uso extensivo de instruções escritas – “coloque

o cartão, lado da fita magnética para baixo, na abertura à sua esquerda” – é instalado em um bairro com uma população predominantemente não alfabetizada).

3.2.2 Diferentes valores

Viés que se origina quando o sistema é usado por uma população com valores diferentes daqueles assumidos no *design* (por exemplo, um software educacional para ensinar conceitos matemáticos é incorporado em uma situação de jogo que recompensa estratégias individualistas e competitivas, mas é usado por alunos com um background cultural que em grande parte evita a competição e, em vez disso, promove esforços cooperativos).

Fonte: Friedman e Nissenbaum, 1996, p. 334-335, adaptação e tradução nossas²².

Table I. Categories of Bias in Computer System Design

These categories describe ways in which bias can arise in the design of computer systems. The illustrative examples portray plausible cases of bias.

1. Preexisting Bias

Preexisting bias has its roots in social institutions, practices, and attitudes.

When computer systems embody biases that exist independently, and usually prior to the creation of the system, then the system exemplifies preexisting bias. Preexisting bias can enter a system either through the explicit and conscious efforts of individuals or institutions, or implicitly and unconsciously, even in spite of the best of intentions.

1.1. Individual

Bias that originates from individuals who have significant input into the design of the system, such as the client commissioning the design or the system designer (e.g., a client embeds personal racial biases into the specifications for loan approval software).

1.2 Societal

Bias that originates from society at large, such as from organizations (e.g., industry), institutions (e.g., legal systems), or culture at large (e.g., gender biases present in the larger society that lead to the development of educational software that overall appeals more to boys than girls).

2. Technical Bias

Technical bias arises from technical constraints or technical considerations.

2.1 Computer Tools

Bias that originates from a limitation of the computer technology including hardware, software, and peripherals (e.g., in a database for matching organ donors with potential transplant recipients certain individuals retrieved and displayed on initial screens are favored systematically for a match over individuals displayed on later screens).

2.2 Decontextualized Algorithms

Bias that originates from the use of an algorithm that fails to treat all groups fairly under all significant conditions (e.g., a scheduling algorithm that schedules airplanes for take-off relies on the alphabetic listing of the airlines to rank order flights ready within a given period of time).

2.3 Random Number Generation

Bias that originates from imperfections in pseudorandom number generation or in the misuse of pseudorandom numbers (e.g., an imperfection in a random-number generator used to select recipients for a scarce drug leads systematically to favoring individuals toward the end of the database).

2.4 Formalization of Human Constructs

Bias that originates from attempts to make human constructs such as discourse, judgments, or intuitions amenable to computers: when we quantify the qualitative, discretize the continuous, or formalize the nonformal (e.g., a legal expert system advises defendants on whether or not to plea bargain by assuming that law can be spelled out in an unambiguous manner that is not subject to human and humane interpretations in context).

²² No original:

A fim de melhor visualizar as formas de vieses presentes em sistemas inteligentes, apresentam-se as Figuras 1 a 4, que resumem de maneira didática algumas das explicações trazidas acima.

Table I. *Continued*

These categories describe ways in which bias can arise in the design of computer systems. The illustrative examples portray plausible cases of bias.

3. Emergent Bias

Emergent bias arises in a context of use with real users. This bias typically emerges some time after a design is completed, as a result of changing societal knowledge, population, or cultural values. User interfaces are likely to be particularly prone to emergent bias because interfaces by design seek to reflect the capacities, character, and habits of prospective users. Thus, a shift in context of use may well create difficulties for a new set of users.

3.1 New Societal Knowledge

Bias that originates from the emergence of new knowledge in society that cannot be or is not incorporated into the system design (e.g., a medical expert system for AIDS patients has no mechanism for incorporating cutting-edge medical discoveries that affect how individuals with certain symptoms should be treated).

3.2 Mismatch between Users and System Design

Bias that originates when the population using the system differs on some significant dimension from the population assumed as users in the design.

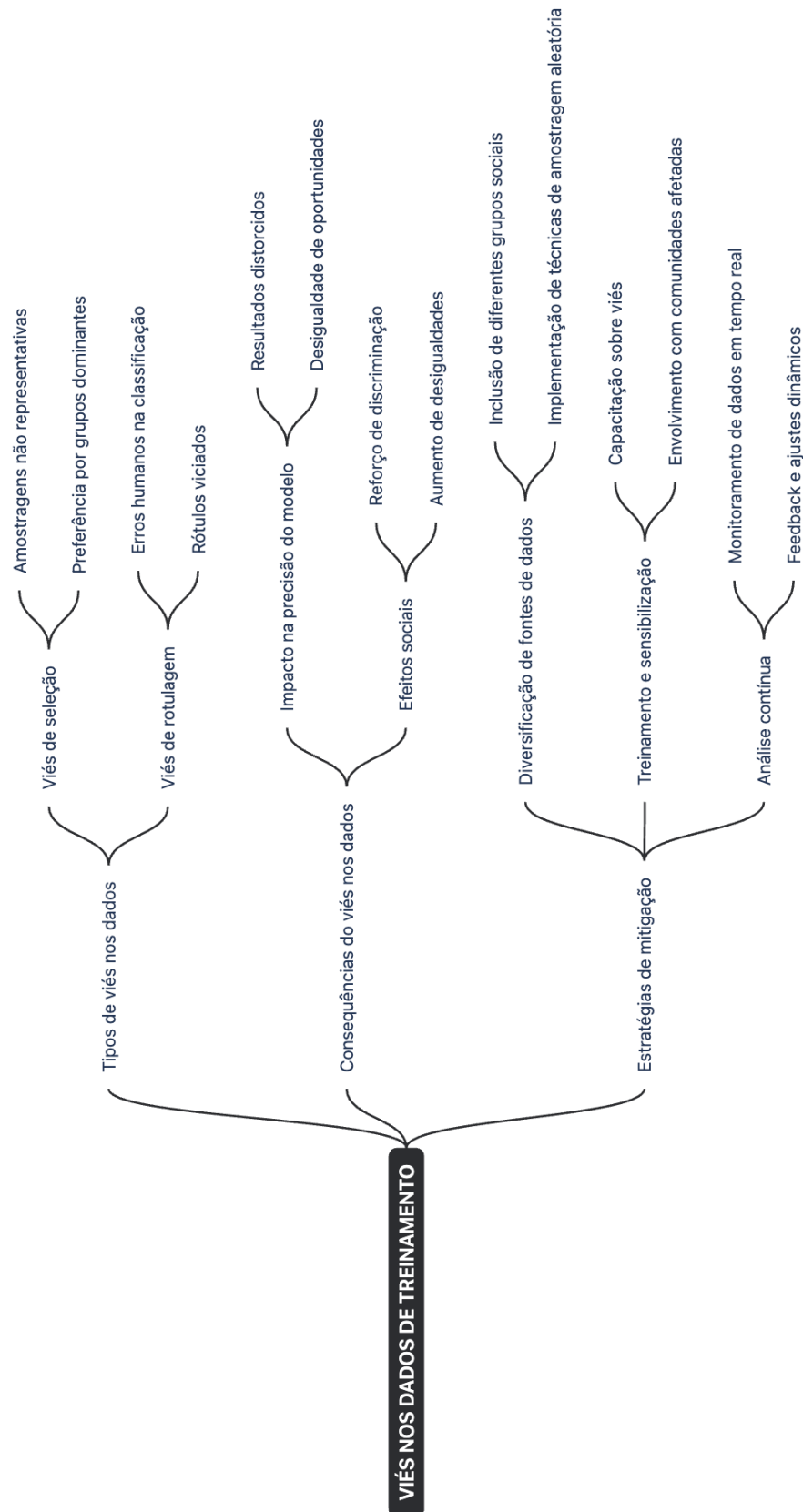
3.2.1 Different Expertise

Bias that originates when the system is used by a population with a different knowledge base from that assumed in the design (e.g., an ATM with an interface that makes extensive use of written instructions—"place the card, magnetic tape side down, in the slot to your left"—is installed in a neighborhood with primarily a nonliterate population).

3.2.2 Different Values

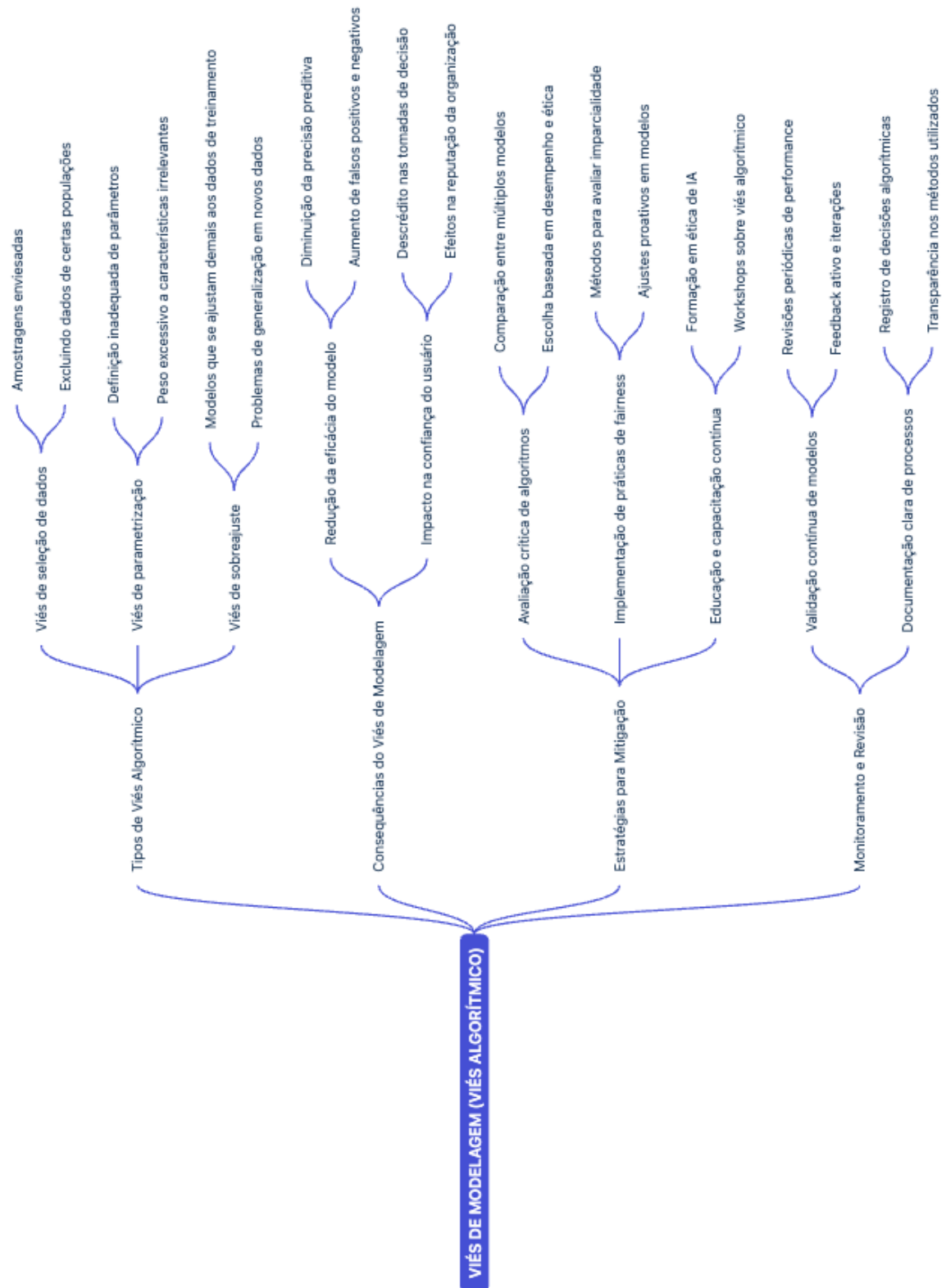
Bias that originates when the system is used by a population with different values than those assumed in the design (e.g., educational software to teach mathematics concepts is embedded in a game situation that rewards individualistic and competitive strategies, but is used by students with a cultural background that largely eschews competition and instead promotes cooperative endeavors).

Figura 1. Mapa mental sobre viés nos dados de treinamento.



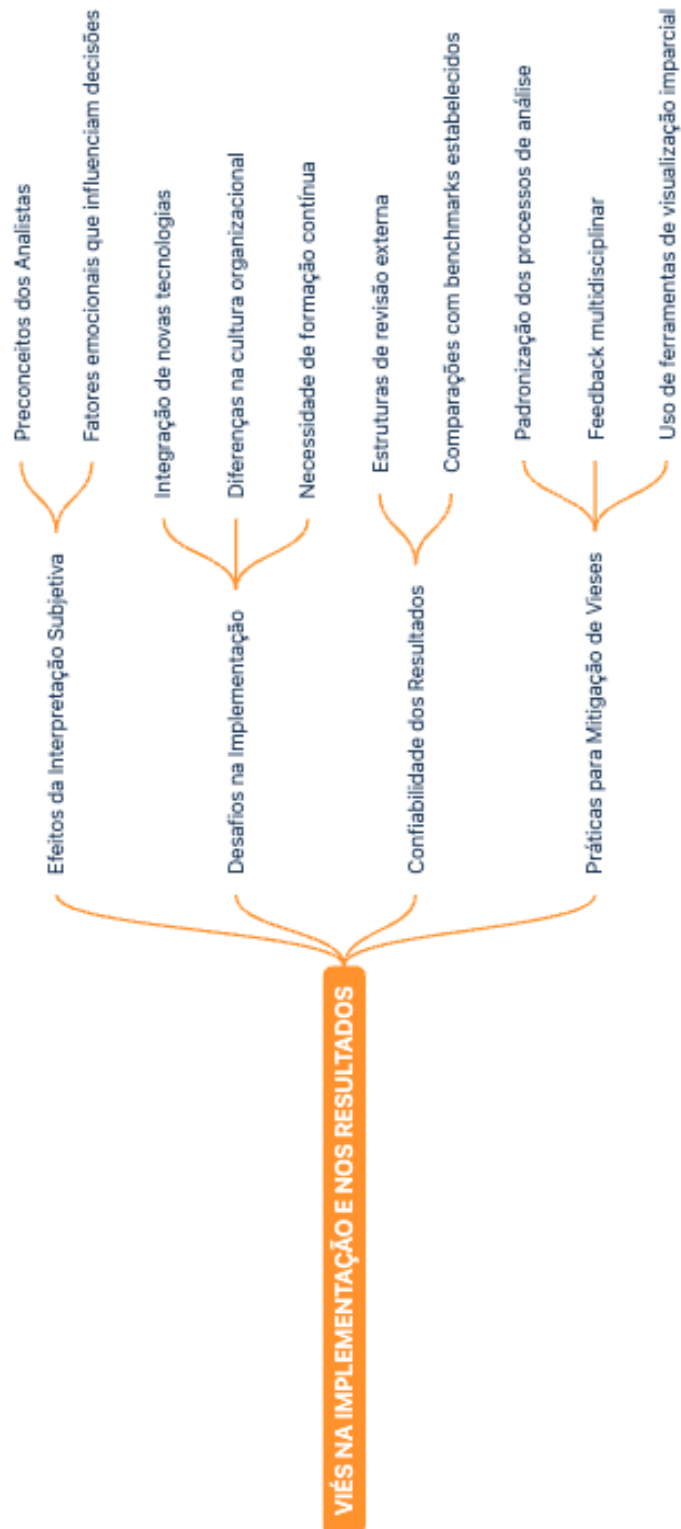
Fonte: a autora (com uso da IA Monica.im para a criação do mapa mental, observando-se o que dispõe o “Guia orientador: Reflexão sobre o uso da inteligência artificial na pesquisa acadêmica” (UNIVERSIDADE ESTADUAL DE PONTA GROSSA, [2025?])).

Figura 2. Mapa mental sobre viés de modelagem ou viés algorítmico.



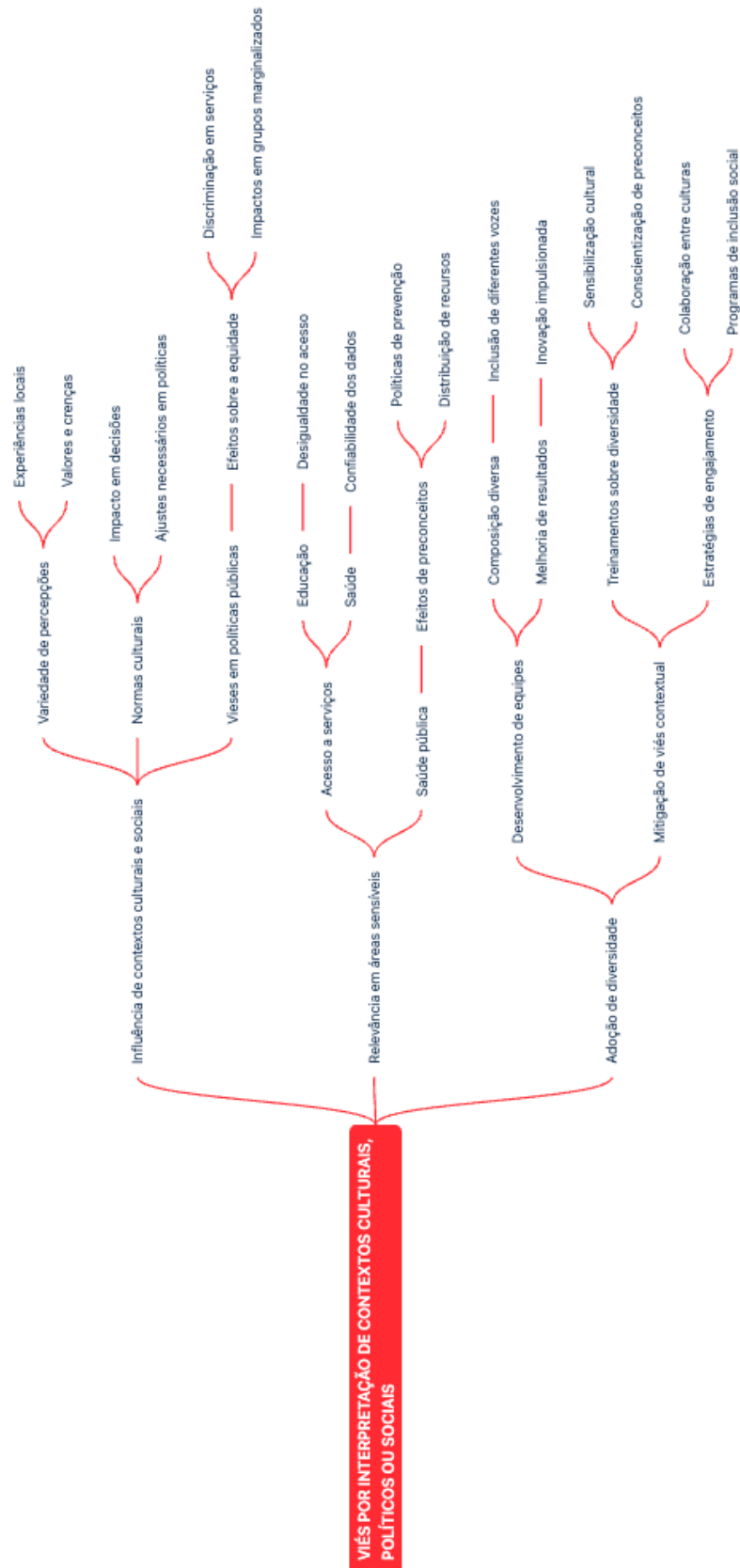
Fonte: a autora (com uso da IA Monica.im para a criação do mapa mental, observando-se o que dispõe o “Guia orientador: Reflexão sobre o uso da inteligência artificial na pesquisa acadêmica” (UNIVERSIDADE ESTADUAL DE PONTA GROSSA, [2025?])).

Figura 3. Mapa mental sobre viés na implementação e nos resultados.



Fonte: a autora (com uso da IA Monica.im para a criação do mapa mental, observando-se o que dispõe o “Guia orientador: Reflexão sobre o uso da inteligência artificial na pesquisa acadêmica” (UNIVERSIDADE ESTADUAL DE PONTA GROSSA, [2025?])).

Figura 4. Mapa mental sobre viés por interpretação de contextos culturais, políticos ou sociais.



Fonte: a autora (com uso da IA Monica.im para a criação do mapa mental, observando-se o que dispõe o “Guia orientador: Reflexão sobre o uso da inteligência artificial na pesquisa acadêmica” (UNIVERSIDADE ESTADUAL DE PONTA GROSSA, [2025?])).

É fundamental afastar a suposição de que os modelos matemáticos empregados no processamento de dados são inteiramente racionais e neutros. Como observa O’Neil (2020), tais modelos expressam opiniões codificadas matematicamente, sendo inevitavelmente influenciados pelas subjetividades dos programadores – desde a seleção e rotulagem dos dados até as categorias empregadas pelos algoritmos para organizar as informações. Um exemplo disso pode ser observado em modelos de recomendação de plataformas de *streaming*, que podem privilegiar determinados gêneros cinematográficos conforme as preferências de quem os desenvolveu, em vez de oferecer uma diversidade real de opções ao usuário.

No Direito, por exemplo, em um sistema de análise preditiva que avalia a probabilidade de reincidência criminal, as escolhas feitas pelos desenvolvedores – como quais fatores considerar, como histórico criminal, demografia ou até mesmo dados socioeconômicos – podem refletir preconceitos existentes. Se o modelo for treinado com dados que contêm viés racial, ele poderá, inadvertidamente, reforçar discriminações ao avaliar indivíduos de grupos minoritários como mais propensos a cometer novos crimes, em vez de fornecer uma avaliação justa e imparcial.

É correta a afirmativa de que modelos matemáticos são simplificações da realidade que buscam realizar análises estatísticas, mas frequentemente (e, às vezes, não intencionalmente) ignoram informações relevantes. A IA é suscetível a erros, uma vez que é desenvolvida e utilizada por seres humanos que, consciente ou inconscientemente, incorporam suas visões de mundo tanto nos dados e na programação quanto no momento de utilizar o sistema. Por exemplo, a falta de compreensão matemática pode impedir que os usuários entendam o que os modelos realmente representam e quais estruturas os fundamentam.

Não bastasse a IA ser suscetível a erros graves e com impacto social significativo, há também a dificuldade de acesso às estruturas subjacentes desses modelos. Nem sempre os vieses são facilmente identificados, pois podem passar despercebidos nos códigos algorítmicos (Friedman; Nissenbaum, 1996, p. 331). Isso resulta em um problema generalizado de transparência no uso de algoritmos e tecnologias de *big data*. Muitas vezes não se tem acesso aos algoritmos que compõem esses modelos, devido a barreiras de propriedade intelectual que separam usuários de desenvolvedores (Lindoso, 2018, p. 374-375).

Além disso, a qualidade dos dados selecionados é crucial para garantir representações fiéis do que se pretende analisar, evitando que certas características sejam sub-representadas ou super-representadas. A subjetividade se manifesta não apenas no desenvolvimento dos

algoritmos, mas também na fase de seleção de dados e no aprendizado do modelo, refletindo valores e preferências de quem controla o processo (Lindoso, 2018, p. 376).

Adicionalmente, as correlações geradas muitas vezes surgem de um processamento de dados opaco, onde dados pessoais sensíveis são tratados da mesma forma que dados anônimos. Essa abordagem pode distorcer a importância de conteúdos potencialmente discriminatórios, negligenciando a devida atenção e cuidado que deveriam ser aplicados.

A noção de neutralidade frequentemente atribuída às tecnologias, sob a alegação de que os algoritmos são meramente cálculos estatísticos, acaba absolvendo empresas, pesquisadores, desenvolvedores e órgãos públicos envolvidos na criação de algoritmos e sistemas de qualquer responsabilidade.

Por exemplo, na criação de um modelo de IA, são escolhidos aspectos da realidade que o sistema considerará, descartando-se os demais aspectos que serão “irrelevantes”. Esse processo de separação de importâncias revela a opinião e as prioridades dos criadores do sistema, frutos de suas ideologias. Nesse contexto, torna-se comum a “sobressimplificação” de processos que são complexos, criando-se anomalias, simplificações errôneas. Quando se considera a implementação da IA em áreas como o Direito, simplificações como essas podem desconsiderar aspectos importantes da vida humana e tornar a definição de sucesso do modelo bastante arbitrária. O modelo pode apresentar um bom funcionamento aos olhos de seus criadores e de algumas pessoas, mas não do ponto de vista das pessoas afetadas negativamente por ele (Boeing, 2019, p. 62).

Em resumo, o recorte desta pesquisa concentra-se nos vieses algorítmicos em sentido amplo (viés nos dados, na modelagem do sistema inteligente e na sua interpretação), pois eles traduzem, em termos técnicos e mensuráveis, os processos sociais e cognitivos de produção do viés humano. O viés algorítmico em sentido amplo representa a tradução tecnológica das distorções humanas, potencializadas pela escala e pela aparente neutralidade da IA. Enquanto os vieses cognitivos afetam decisões individuais, os vieses algorítmicos ampliam e automatizam tais assimetrias, transformando erros humanos em padrões estruturais de discriminação, especialmente no contexto judicial. Essa escolha metodológica permite observar como distorções de julgamento – muitas vezes invisíveis em contextos humanos – se tornam detectáveis e passíveis de regulação quando incorporadas a sistemas inteligentes.

Como se viu anteriormente, no plano cognitivo, as heurísticas – como disponibilidade, representatividade, ancoragem e confirmação – são descritas como atalhos mentais que simplificam a tomada de decisão. Elas permitem que o ser humano formule julgamentos

rápidos, mas também introduzem distorções sistemáticas, conhecidas como vieses cognitivos. Kahneman e Tversky (1974, 1981) demonstram que, ao recorrer ao pensamento rápido e intuitivo (Sistema 1), o indivíduo tende a superestimar a frequência de eventos facilmente lembrados, ignorar probabilidades base ou buscar apenas informações que confirmem suas crenças – fenômenos que levam a decisões parciais e subjetivas.

Essa limitação humana é replicada nos sistemas de IA, pois os algoritmos são concebidos, treinados e ajustados por pessoas sujeitas às mesmas distorções cognitivas. Como mostra a tese, os vieses algorítmicos surgem quando as associações implícitas, percepções seletivas e padrões de julgamento humano são incorporados aos dados e modelos computacionais. Assim, se os sistemas inteligentes são projetados por seres humanos, tais sistemas também não estão livres de vieses, reproduzindo discriminações de forma automatizada e, muitas vezes, ampliada.

Em outras palavras, os vieses humanos e as heurísticas que distorcem o julgamento individual encontram um paralelo direto nas falhas de *design* e de treinamento de algoritmos. O que, na mente humana, decorre da economia cognitiva e da necessidade de simplificação, na IA manifesta-se na dependência de dados incompletos, desbalanceados ou enviesados. Por exemplo, um algoritmo de predição judicial pode reproduzir o viés de confirmação ao reforçar padrões históricos de condenações, ou o viés de disponibilidade ao supervalorizar tipos de casos mais recorrentes em seus dados de treinamento.

Dessa forma, a tese evidencia que há continuidade entre os vieses humanos e os vieses algorítmicos: ambos emergem de mecanismos de simplificação que buscam eficiência cognitiva ou computacional, mas que, ao fazê-lo, sacrificam a neutralidade e a equidade. No contexto do Judiciário, essa relação é particularmente preocupante, pois a automatização de decisões potencializa a reprodução de preconceitos estruturais – especialmente de gênero – transformando distorções cognitivas individuais em injustiças sistêmicas mediadas por tecnologia.

Seja como for, entende-se aqui que vieses são (quase sempre) prejudiciais, limitando a participação de grupos na sociedade e minando o potencial da aplicação de sistemas inteligentes em várias áreas, gerando desconfiança ao produzir resultados distorcidos e injustos. É comum encontrar dados e parâmetros viesados (Friedman; Nissenbaum, 1996; Lima, 2022; García-Ull; Melero-Lázaro, 2023), já que são produzidos por seres humanos e, portanto, refletem os preconceitos inerentes à natureza humana. Nesse sentido, ideologias, preferências e discriminações, ao serem embutidas em um algoritmo de maneira “descuidada”, são

revestidas de autoridade científica, passam despercebidas, pois não são identificáveis para a maioria dos indivíduos. Com isso, tais opiniões (dos programadores ou embutidas nos próprios dados que compõem o sistema de IA) são perpetuadas e disfarçadas numa espécie de “neutralidade” robótica e de “verdade científica”.

Independentemente da origem do viés, a evidência científica é clara: os vieses presentes em sistemas inteligentes frequentemente resultam em consequências indesejadas, muitas vezes discriminatórias (Buolamwini; Gebru, 2018; Noble, 2018; Obermeyer et al., 2019; Ahmad et al., 2022; Bhatnagar; Gajjar, 2024). Esta realidade destaca a necessidade de abordagens para lidar com e reduzir esses vieses (Mittelstadt et al., 2016). Combater e reduzir o viés discriminatório em sistemas de IA tornou-se uma urgência inegável. Os pesquisadores estão ativamente explorando maneiras de enfrentar esse desafio, investigando inclusive a possibilidade de treinar sistemas de aprendizado de máquina para aprender sem viés (Shrestha e Das, 2022; González; Rampino, 2024).

O que se defende não é que a IA deixe de ser utilizada (o que seria uma utopia), mas que quanto mais elevado for o risco para os direitos humanos, mais rigorosos devem ser os requisitos legais para a utilização desses sistemas. Assim, aqueles que desenvolvem e utilizam sistemas de IA devem garantir que esses sistemas melhorem a tomada de decisões e sejam responsáveis na busca e redução de vieses. A compreensão e a medição da justeza dos sistemas inteligentes tornam-se essenciais.

A questão de quando um sistema é justo o suficiente para ser implementado requer uma abordagem multidisciplinar, incorporando perspectivas éticas e sociais. Modelos explicativos e algoritmos transparentes surgem como ferramentas fundamentais no combate aos vieses, contribuindo para um ecossistema de IA mais equitativo e responsável.

Como visto, dados distorcidos podem deturpar resultados e dar causa a danos aos indivíduos. Assim, é imprescindível uma análise cuidadosa em relação aos dados que servirão de *input* a determinado sistema inteligente, bem como ser especialmente diligente quanto às propriedades do código/algoritmo, monitorando seu funcionamento e estudando seus resultados. Sem isso, corre-se o risco de reverberações e perpetuações que ameaçam direitos humanos. Em outras palavras, como os algoritmos são criados por humanos, há um alto potencial de absorção de preconceitos e discriminações já presentes na sociedade. São diversos os casos em que a tecnologia se torna tão racista, sexista e opressiva quanto a própria sociedade.

Quando empregados pela Administração Pública, algoritmos devem aderir a certos padrões mínimos de transparência e responsabilidade, além dos princípios legais que orientam

a ação administrativa. No entanto, isso não implica que os modelos não possam ser ajustados para se alinharem com sistemas democráticos. O processo de desenvolvimento e aplicação responsável dos modelos pode ser dividido em pelo menos duas etapas, mesmo que se queira manter a integridade de seus códigos. Na fase anterior à elaboração dos códigos, é crucial que a sociedade esteja informada sobre quais aspectos da realidade serão considerados pelo modelo, o que se reflete nos dados que serão utilizados como entradas. Isso possibilita, por exemplo, questionar os critérios que levaram certos dados a serem considerados relevantes e entender por que outros foram desconsiderados. Após a criação dos códigos, é possível realizar auditorias em seus resultados para identificar possíveis vieses ou distorções (Boieng, 2019, p. 65-66).

A inclusão de diferentes setores da sociedade na criação de algoritmos é fundamental para que as definições de sucesso e as funções para as quais eles são utilizados sejam estabelecidas da forma mais benéfica para o maior número possível de pessoas, ou, quando isso não for viável, para minimizar seus impactos negativos. Isso reduz a probabilidade de que tais algoritmos se tornem ferramentas prejudiciais para as pessoas ou que se envolvam em ciclos de retroalimentação negativos. É importante notar que um mesmo algoritmo, treinado com os mesmos dados, pode causar mais ou menos danos dependendo de como é utilizado (Boieng, 2019, p. 66).

Nesse contexto, é certo que a formação de uma nova cultura baseada na comunicação multimodal e no processamento digital de informações trouxe consigo uma série de desigualdades sociais, econômicas e tecnológicas (Castells, 2011). Diante desses novos desafios, deve-se analisar e identificar as principais tendências sociais e os maiores hiatos da Era Digital, especialmente no que se refere à criação de novas identidades e às disparidades sociais emergentes, dentre elas, a repercussão que o viés de gênero causa quando o tema é aprendizagem de máquina.

2.4 A inteligência artificial e papel do feminismo no enfrentamento do viés de gênero

A crescente implementação da IA nas mais diversas esferas da sociedade tem levantado questões críticas sobre seus impactos e implicações, especialmente no que tange à perpetuação de desigualdades de gênero. As tecnologias, que deveriam ser ferramentas de inovação e progresso, muitas vezes reproduzem e amplificam os preconceitos e estereótipos presentes nas sociedades humanas, afetando de maneira desproporcional as mulheres e outros grupos marginalizados. Esse fenômeno é amplificado pela falta de diversidade nas equipes de

desenvolvimento e pela escassez de uma análise crítica das implicações sociais e éticas dessas tecnologias (Lima, 2013; Lima, 2016; Moreira et al., 2018; Chan, [20–]; Cernadas; Calvo-Iglesias, 2021; Nunes et al., 2023).

O problema da discriminação contra mulheres é uma questão profundamente enraizada mundialmente, que se manifesta em diversos aspectos sociais, econômicos e políticos. Ele reflete séculos de desigualdade de gênero e discriminação estrutural. Atualmente, como observam a UNESCO e a IRCAI (2024, p. 5), muitos obstáculos à equidade de gênero ainda persistem, incluindo a violência baseada no gênero, as disparidades salariais e a sub-representação feminina em cargos de liderança, entre outros. O documento desenvolvido através do Programa das Nações Unidas para o Desenvolvimento (PNUD), intitulado *2023 Gender Social Norms Index*, destaca que a discriminação de gênero é uma questão global que ainda prevalece e que cobre 85% da população mundial: “9 em cada 10 homens e mulheres possuem preconceitos fundamentais contra as mulheres” (United Nations Development Programme, 2023).

Este subcapítulo explora o papel do feminismo no enfrentamento do viés de gênero presente nos sistemas de IA.

Segundo a UNESCO (2014, p. 60), gênero pode ser conceituado como o significado social atribuído a ser homem ou a ser mulher. São características socialmente impostas a homens e mulheres, servindo de parâmetro para definir o que se pode e o que se deve fazer enquanto homem e mulher, determinando os comportamentos, os papéis, as expectativas e os direitos de homens e mulheres.

O termo se refere às normas e expectativas ligadas à expressão em comportamentos, gestos, linguagem, sentimentos e aparência física. Assim, o gênero está relacionado a características (estereotipadas) que se conectam a expectativas sociais sobre comportamentos, atitudes e preferências adequadas ou inadequadas (Pillinger, 2019, p. 5). Embora se reconheça que sexo e gênero se influenciam mutuamente e que o gênero é geralmente visto como uma construção social, Butler (1990) leva essa argumentação além, afirmando que tanto o sexo quanto o gênero são construções sociais. Ela cita Beauvoir, que declarou: “Não se nasce mulher, torna-se mulher”.

Para a UNESCO (2014, p. 11), a igualdade de gênero é um conceito fundamental que se refere à igualdade de direitos, responsabilidades e oportunidades entre mulheres e homens. Esse princípio implica que os interesses, necessidades e prioridades de homens e de mulheres sejam cuidadosamente considerados, reconhecendo a diversidade existente entre os diferentes

grupos. Essa diversidade também deve incluir fatores como raça, classe social, etnia e orientação sexual, que podem influenciar as experiências de cada indivíduo. Nessa condição, a igualdade de gênero não é apenas um princípio de direitos humanos; é uma condição indispensável para o desenvolvimento sustentável que coloca as pessoas no centro das políticas públicas e é, ao mesmo tempo, um objetivo a ser alcançado em si.

Para discorrer sobre o tema, iniciaremos com uma análise técnica a respeito de identidade, gênero e feminismo. A partir dessa base, a discussão se desvia para o campo da tecnologia, abordando como o feminismo tem se posicionado em relação à criação e ao uso de tecnologias como a IA, com ênfase no tecnofeminismo.

Busca-se não somente destacar os desafios impostos pelo viés de gênero na IA, mas também iluminar algumas das estratégias e perspectivas que o feminismo, especialmente o tecnofeminismo, oferece para a construção de um futuro tecnológico mais inclusivo e igualitário.

2.4.1. O feminismo, a construção de identidades e os espaços de poder

O conceito de “feminismo” está diretamente relacionado à identificação e superação das diferenças construídas em múltiplos domínios sociais. Ao abordar este tema, Perrot (2005, p. 11) discute as identidades e as diferenças, bem como a criação de estereótipos, vinculando-os à luta feminista: “A questão dos papéis que foram e ainda são atribuídos a determinados gêneros envolve as lutas, as discussões e as resistências do feminismo até os dias atuais”. Nesse sentido, o debate sobre identidade, gênero e feminismo torna-se essencial para a análise da sociedade contemporânea e seus sujeitos (Cornell; Pearse, 2015).

Dentro dessa discussão, a estruturação do pensamento binário moderno, em que o homem é considerado o sujeito do conhecimento e a referência, enquanto a mulher é posicionada como o “outro”, o oposto, é um ponto fundamental. Como observam Silva e Ribeiro (2014, p. 254), esse sistema binário é central na lógica patriarcal que limita as possibilidades de autoexpressão das mulheres, ao mesmo tempo em que define seu ser, seu comportamento e suas possibilidades de realização.

A questão de gênero, abordada por Haraway (1991, p. 219), também destaca a relação entre os termos “gênero” e “diferença”. Para Haraway, o conceito de gênero está associado à construção de sistemas de diferenciação e, ao trabalhar com a teoria feminista sob uma perspectiva epistemológica, é necessário compreender essa diferença. Ao analisar categorias

como sexo, corpo, biologia, raça e natureza, o feminismo busca questionar os binarismos que tentam universalizar esses conceitos (Haraway, 1991, p. 250). Nesse sentido, o feminismo não só desafia a ideia de uma identidade fixa e imutável, mas também trabalha para abrir espaço para uma multiplicidade de experiências e a fluidificação das identidades.

A construção das identidades de gênero, como esclarece Scott (1995), ocorre por meio de interações sociais e convivências, de maneira que os discursos sociais repetidos se tornam verdades incontestáveis, gerando marcas identitárias como “feminilidade” e “masculinidade”. A criação dessas identidades pressupõe, portanto, a criação de diferenças. Ambas são social e culturalmente produzidas dentro de relações de poder (Silva, 2005).

É nesse cenário que o feminismo desempenha um papel crucial, oferecendo uma lente crítica para questionar e desconstruir as construções sociais e culturais que moldam as identidades. Tradicionalmente, essas identidades foram moldadas por normas patriarcais que impuseram papéis rigidamente definidos para os indivíduos, especialmente para as mulheres. Ao contestar essas normas, o feminismo busca permitir que as mulheres e outros grupos possam se reconhecer de maneira autônoma e plural, sem as limitações impostas pelo patriarcado.

Na perspectiva sociológica, Hall (2011, p. 13; 15; 38-39) argumenta que a identidade é continuamente moldada e transformada em um processo dinâmico, flexível e fluido, que ocorre por meio das interações sociais e é influenciado pelas representações culturais e pela forma como os indivíduos são interpretados pelos sistemas culturais que os rodeiam. Isso se reflete nas sociedades modernas, caracterizadas pela constante mudança (Bauman, 2021), onde a identidade é moldada por processos inconscientes de identificação e pela maneira como os indivíduos se posicionam em relação às categorias sociais e culturais estabelecidas.

Silva (2000) acrescenta que a identidade contemporânea está intimamente ligada à diferença, de modo que um conceito só pode ser entendido em relação ao outro. A identidade, portanto, define-se pela criação de fronteiras e distinções entre o que significa ser ou não ser, pertencer ou não pertencer, estar incluído ou excluído. Esses processos de inclusão e exclusão estão profundamente enraizados nas dinâmicas culturais e nas relações de poder que definem quem é considerado “dentro” ou “fora” da sociedade.

Para Foucault (2006), as identidades são constituídas discursivamente em um campo de relações de poder disseminadas pela sociedade. A formação das identidades não é um processo neutro, mas sim determinado por forças e relações de poder que operam constantemente. Foucault (1995, p. 235) sugere que, ao se tornar sujeito, o indivíduo está sujeito a alguém pelo controle e pela dependência, sendo preso à sua própria identidade através de uma

consciência ou autoconhecimento que, por sua vez, são produtos de uma dinâmica de poder. Ao questionar a “sexualidade”, Foucault (1990) argumenta que ela deve ser entendida dentro dessas relações de poder que a moldam ao longo da história, sugerindo que as práticas sexuais, assim como a subjetividade humana, são influenciadas pelas estruturas sociais e políticas que governam os comportamentos humanos.

O sexismo, nesse contexto, pode ser compreendido como um sistema de dominação institucionalizada, no qual o homem ocupa o centro do poder decisório. Nesse sistema, os homens definem os papéis e o destino das mulheres, mantendo-as em uma posição subalterna e dependente. Estar oprimida nesse sistema significa ser privada da capacidade de fazer escolhas autônomas, de definir seu próprio caminho e sua identidade. O feminismo, então, surge como uma força de resistência contra esse patriarcado, questionando e contestando essas formas de dominação e buscando a erradicação do sistema patriarcal.

A luta feminista está intimamente ligada à transformação dos espaços de poder, pois para que as mulheres possam conquistar a autonomia e liberdade de suas identidades, é necessário reconfigurar as estruturas sociais que as restringem. Como Hooks (2015, p. 5) destaca, essa luta acontece em qualquer contexto em que homens e mulheres se oponham à exploração e opressão social, visando erradicar o patriarcado. Nesse contexto, a questão da identidade está diretamente conectada à desconstrução dos estereótipos e das expectativas pré-estabelecidas, que limitam as possibilidades de autoexpressão e autodefinição das mulheres.

Embora o movimento feminista não tenha se desenvolvido linearmente, já que é um campo marcado por disputas e hierarquias com diversas vertentes, como aponta Cockburn (2006), a segunda onda do feminismo teve um papel importante ao analisar a ausência das mulheres em espaços de poder, incluindo os campos da ciência e da tecnologia (Wajcman, 2004, p. 16). Essas ausências evidenciam a necessidade de repensar e reconfigurar os espaços de poder para garantir que as mulheres, assim como outros grupos marginalizados, possam ocupar e definir suas identidades de maneira autônoma, sem as limitações impostas pelo patriarcado.

2.4.2 A emergência do (tecno)feminismo em contraposição ao viés de gênero na inteligência artificial

A crescente implementação tecnológica nas últimas décadas tem transformado a sociedade de maneira significativa, impulsionando a Quarta Revolução Industrial e criando uma

demanda sem precedentes por profissionais qualificados em áreas como tecnologia da informação, IA, engenharia e ciência de dados. No entanto, um dos grandes desafios dessa transformação é a sub-representação das mulheres nessas disciplinas.

Embora a tecnologia desempenhe um papel central na economia moderna e na criação de empregos do futuro, as mulheres continuam a ser uma minoria nesses campos cruciais (Bello et al., 2021, p. 25). Em grande parte das sociedades, há desigualdades entre mulheres e homens em responsabilidades atribuídas, atividades realizadas, quanto ao acesso a direitos e a recursos, bem como no que diz respeito a oportunidades de tomada de decisão. Essa lacuna quanto à equidade de gênero é preocupante porque a falta de diversidade pode limitar a inovação e a eficácia das soluções tecnológicas.

No que diz respeito à produção do conhecimento científico, por exemplo, pode-se afirmar que a ciência moderna é predominantemente masculina, uma vez que a maioria da população cientista é composta por homens (Simmel apud Homey, 1926; Estébanez, 2004; Keller, 2004; Aquino, 2006). Na ciência moderna, percebe-se que as características e as habilidades valorizadas em seu desenvolvimento são as associadas ao masculino (Wajcman, 1991; Bourdieu, 2002; Witzel, 2011). Nesse sentido, a dominação masculina é assimétrica, vertical e desigual (Britto; Cruz, 2023).

Em 2019, a UNESCO elaborou um relatório que identifica que somente 29,3% dos cargos de pesquisa científica no mundo pertencem a mulheres (UNESCO Institute for Statistics, 2019, p. 2). Na ciência como um todo, as mulheres são consideradas “fora de lugar”, um argumento que reflete a hegemonia masculina, caracterizada por um viés androcêntrico que define a legitimidade e objetividade do conhecimento (Silva; Ribeiro, 2014, p. 457; Britto; Cruz, 2023).

Especificamente no domínio da IA, a desigualdade de gênero também representa uma preocupação significativa. A maneira como os sistemas de IA são elaborados pode levar a problemas relacionados ao viés de gênero. Isso porque tanto os dados podem conter discriminações (Noble, 2018; Allhutter, 2018; Alexander, 2018), como os próprios desenvolvedores dos algoritmos podem não ter consciência de seus preconceitos e, de maneira implícita, transferir suas crenças de gênero enraizadas na sociedade para os robôs.

Numa perspectiva negativa, a sub-representação feminina nas áreas de ciência e tecnologia, especialmente na criação de ferramentas baseadas em IA, não só contribui para a perpetuação de discursos sexistas dentro da própria tecnologia, como também é capaz de reverter décadas de avanços em igualdade de gênero (Bray, 2007; Franqueira, 2019; Costa,

2020; Britto; Cruz, 2023, UNESCO et al., 2023, p. 22). A predominância masculina na criação dessas tecnologias pode, de forma silenciosa, reforçar e ampliar desigualdades de gênero, criando um ciclo que tende a perpetuar esses vieses ao longo do tempo. Isso ocorre porque, como as mulheres representam uma porcentagem consideravelmente menor da força de trabalho tecnológica (Simonite, 2018; Young et al., 2021, p. 2, 10; Kumar; Choudhury, 2022, p. 2), há o risco de que as tecnologias, em especial as de aprendizado de máquina, acabem codificando e refletindo vieses de gênero (Bass; Huet, 2017; Leavy, 2018; Von Ende; Oliveira, 2020).

Além disso, o fato de existir um número significativamente maior de homens programadores do que mulheres programadoras pode contribuir para que as máquinas aprendam a partir de algoritmos (e também de dados) predominantemente carregados de estereótipos de gênero (Caliskan et al. 2017; Britto; Cruz, 2023; UNESCO; IRCAI, 2024). Isso faz com que a aplicação dessas tecnologias, muitas vezes, continue a reproduzir tais vieses, aprofundando a desigualdade de gênero. O relatório *Where are the women? Mapping the gender job gap in AI*, publicado por Young et al. (2021), destaca que a escassez de mulheres na ciência de dados resulta em ciclos de retroalimentação que introduzem ainda mais o viés de gênero nos sistemas de IA e aprendizado de máquina.

Tal como os demais vieses, o viés de gênero pode ser inserido na IA em diferentes fases do seu desenvolvimento e aplicação. Primeiramente, os dados usados para treinar algoritmos podem não representar adequadamente o grupo feminino, devido às decisões tomadas anteriormente sobre que dados coletar e como organizá-los. Em seguida, o processo de modelagem ou análise pode carregar vieses devido a suposições ou escolhas feitas pelos desenvolvedores, refletindo seus próprios valores e prioridades (sejam conscientes ou inconscientes), ou resultando de uma compreensão inadequada dos dados. Até mesmo as decisões sobre quais sistemas de IA devem ser criados podem carregar preconceitos. Como O’Neil (2020, p. 35) observa claramente: “Modelos são opiniões embutidas em matemática”. Se a maior parte das decisões sobre IA é tomada por homens brancos, é natural que a tecnologia, supostamente “neutra”, acabe refletindo preferências masculinas (Zou; Schiebinger, 2018). Quanto menor for a presença de mulheres no setor de IA, maior o risco de os sistemas automatizados futuros perpetuarem e reforçarem o viés de gênero, criando um ciclo contínuo de desigualdade.

Esse problema se agrava pelo fato de que os algoritmos e os dados que alimentam essas tecnologias não são neutros ou objetivos, mas refletem o mundo e as estruturas sociais

em que vivemos. Vale ressaltar que a cada dia a IA se torna mais ubíqua, tornando imperioso um estudo detalhado do *gap* de gênero existente na área de computação e, em especial, da IA e da ciência de dados (Young et al. 2021, p. 2).

Diversos produtos de IA já tiveram resultados discriminatórios no que diz respeito ao gênero. Para citar alguns exemplos: um algoritmo de recrutamento criado pela Amazon, iStock, foi identificado como discriminatório em relação a candidatas mulheres (Dastin, 2018; Reuters, 2018); um chatbot em redes sociais precisou ser desligado após começar a propagar discursos de ódio racistas e sexistas (Kwon; Yun, 2021); os algoritmos de geração de imagens treinados com a base de dados da ImageNet, iGPT, SimCLRv2, Lensa e Dall-E 2 são carregados de vieses (Steed; Caliskan, 2021; Mahdawi, 2021; Heikkilä, 2022; Heikkilä, 2023); ao ser testado, o Midjourney revelou discriminação de gênero em relação ao trabalho feminino (Thomson e Thomas, 2023); algoritmos que anunciam vagas científicas são vistos de maneira desproporcional entre homens e mulheres (Maron, 2018; Lambrecht; Tucker, 2019); sistema de reconhecimento facial que identifica rostos de homens brancos, mas não identifica rosto de mulheres negras (Boulamwini; Gebru, 2018); GPT-2 e Llama 2, modelos de linguagem da OpenAI e da Meta, respectivamente, também são carregados de vieses sociais, incluindo de gênero (UNESCO; IRCAL, 2024, p. 3); e até um programa de reconhecimento facial em Oslo que sugeria opções de saladas para mulheres e pizzas para homens em um restaurante (Brandão, 2021).

No Brasil, a desigualdade de gênero no setor de tecnologia é bastante evidente. Um estudo da Serasa Experian, intitulado “Perfil das Mulheres em TI no Brasil” e realizado em fevereiro de 2024, revelou que menos de 1% das mulheres no país estão envolvidas com tecnologia. A pesquisa apontou que a presença feminina nesse campo representa apenas 0,07%, o que equivale a 69,8 mil profissionais, com base em uma amostra de mais de 93,3 milhões de mulheres; em comparação aos homens, 0,33% trabalham em TI, conforme os dados obtidos a partir de uma análise de 92,4 milhões de homens (Serasa Experian, 2024).

No cotidiano, em que o patriarcalismo e a lógica capitalista global favorecem práticas machistas e sustentam representações sociais discriminatórias de gênero, a tecnologia tende a reproduzir e até intensificar essas mesmas desigualdades. No caso da IA, o viés de gênero incorporado tanto na codificação dos sistemas quanto nos dados utilizados pode levar a decisões automatizadas que não apenas refletem, mas também amplificam assimetrias já existentes, configurando uma forma algorítmica de opressão de gênero. Nessa perspectiva, a tecnologia

acaba por se tornar uma manifestação concreta do poder masculino (Caliskan et al., 2017; Britto; Cruz, 2023).

Em síntese, a sub-representação feminina nas áreas de ciência e tecnologia, combinada com a presença de vieses de gênero nos dados e algoritmos, ameaça a promoção da equidade no desenvolvimento tecnológico. A predominância masculina na criação e condução da IA pode, portanto, transformar essas ferramentas em expressões reforçadas do poder patriarcal, aprofundando desigualdades estruturais e restringindo as oportunidades para as mulheres nesse campo.

É nesse cenário que o papel do feminismo se torna central, consistindo em investigar a presença de preconceitos de gênero que permeiam as diversas áreas do conhecimento humano (Scavone, 2008). Especificamente, a abordagem tecnofeminista questiona como as construções de gênero influenciam concepções de conhecimento, pesquisa, produção científica e desenvolvimento tecnológico (Wajcman, 2004).

O tecnofeminismo pode ser compreendido como uma evolução natural do pensamento feminista, incorporando as complexidades da era digital (Silva, 2019). Enquanto o feminismo tradicional desconstrói estruturas patriarcais e desigualdades de gênero em contextos sociais, políticos e culturais, o tecnofeminismo amplia essa análise para a tecnologia. Reconhece-se que ferramentas e sistemas tecnológicos não são neutros, mas refletem relações de poder desiguais na sociedade.

Na perspectiva tecnofeminista, a análise vai além da presença feminina na ciência e tecnologia, abordando como as mulheres podem ser tratadas de maneira mais justa dentro e fora desses campos. O foco não se restringe ao acesso das mulheres à tecnologia, mas se estende aos processos que moldam seu desenvolvimento e aplicação, bem como à construção e reforço das normas de gênero.

Uma análise tecnofeminista (Wajcman, 2004) interseccional examina como as dinâmicas de poder relacionadas ao gênero e as significações atribuídas ao gênero impactam o processo de transformação tecnológica e ficam refletidas nas próprias tecnologias. Estudos indicam que essas tecnologias são associadas ao gênero tanto pelo contexto em que estão inseridas quanto pelo *design*. Assim, as “práticas de gênero” moldam a revolução digital e influenciam as estruturas políticas e socioeconômicas que sustentam os sistemas tecnológicos (Wajcman et al., 2020, p. 5). Como expressa Wajcman (2010, p. 149-150, tradução nossa²³):

²³ No original: “women’s identities, needs and priorities are configured together with digital technologies”.

“as identidades, necessidades e prioridades das mulheres são configuradas em conjunto com as tecnologias digitais”.

A IA é moldada por seus algoritmos e pelos dados que a compõem, carregando experiências e pontos de vista humanos, inseridos consciente e inconscientemente. Se “a tecnologia, por si só, não é nem intrinsecamente patriarcal nem claramente libertadora” (Wajcman et al., 2020, p. 5, tradução nossa²⁴), mas reflete o modo de “pensar” da sociedade, a análise feminista deve se concentrar nas circunstâncias econômicas, sociais e políticas que influenciam o desenvolvimento e o uso das tecnologias digitais.

Dessa forma, é imprescindível abordar como as tecnologias incorporam e reproduzem normas de gênero impostas socialmente. Além disso, é necessário utilizar as tecnologias digitais para impulsionar a igualdade de gênero, evitando reforçar padrões discriminatórios. A aplicação de conceitos e técnicas da epistemologia feminista pode potencializar a revolução digital como meio para promover a justiça social e a equidade de gênero (Wajcman et al., 2020, p. 5).

Seja qual for a área do conhecimento, independentemente do aspecto abstrato ou prático, a implementação de uma perspectiva de gênero se torna crucial, pois envolve a avaliação minuciosa das implicações de qualquer ação planejada, seja uma nova legislação, políticas públicas ou programas sociais, para mulheres e homens.

O tecnofeminismo, além de identificar problemas, propõe práticas e políticas para garantir que a IA seja desenvolvida e implementada de forma justa e equitativa. Uma de suas principais contribuições é a identificação e correção de vieses de gênero nos sistemas de IA. O feminismo tradicional já alertava sobre a presença desses vieses em diversas instituições sociais, como educação, trabalho e mídia. O tecnofeminismo amplia essa análise, demonstrando como esses preconceitos estão embutidos nos dados e algoritmos da IA (Romero, 2012; Wajcman; Frade, 2023).

Pode-se encaixar na abordagem tecnofeminista várias propostas para o enfrentamento do viés de gênero na IA. Por exemplo, a realização de auditorias regulares nos conjuntos de dados utilizados para treinar sistemas de IA, garantindo que eles sejam representativos e livres de estereótipos de gênero, bem como a promoção da diversidade nas equipes de desenvolvimento de IA, pois perspectivas diversas ajudam a identificar e corrigir vieses que poderiam passar despercebidos.

Outro princípio fundamental que se adapta à abordagem é a transparência algorítmica, que exige clareza e possibilidade de revisão nos processos de tomada de decisão dos algoritmos.

²⁴ No original: “technology as such is neither inherently patriarchal nor unambiguously liberating”.

A inclusão e representação também são pilares centrais do feminismo, que sempre defendeu a presença equitativa das mulheres em espaços de poder e decisão (Correia, 2020). No contexto da IA, isso significa incentivar a participação de meninas e mulheres em áreas como ciência, tecnologia, engenharia e matemática desde cedo, por meio de programas de mentoria e bolsas de estudo (Britto; Cruz, 2023). Além disso, promover mulheres para cargos de liderança em empresas de tecnologia e comitês de ética de IA é essencial para garantir que suas vozes influenciem as decisões estratégicas. O desenvolvimento de tecnologias deve considerar as necessidades e perspectivas de todos os gêneros, evitando soluções que reforcem estereótipos ou excluam determinados grupos.

Estudos demonstram que equipes diversificadas, compostas por homens e mulheres, não apenas são mais eficazes na identificação de dados tendenciosos, mas também têm maior capacidade de antecipar questões com sérias consequências sociais (Kumar; Choudhury, 2022, p. 2). Como apontado em Quiros et al. (2018, tradução nossa²⁵), “a tecnologia reflete os valores de seus criadores... Está claro que equipes mais diversas, envolvidas no desenvolvimento dessas tecnologias, podem ajudar a identificar e prevenir vieses”.

A diversidade nas equipes de desenvolvimento não só favorece a identificação de preconceitos ocultos, mas também fortalece a criação de soluções mais inclusivas, que atendem a um espectro mais amplo de necessidades e perspectivas. Isso permite a avaliação, sob uma perspectiva de gênero, dos dados, das decisões de *design* e do contexto social que envolvem a tomada de decisões algorítmicas (Yarger et al., 2019).

Defende-se, portanto, que a paridade de gênero no desenvolvimento de sistemas de IA é um dos pilares fundamentais para garantir que os produtos e tecnologias sejam mais equitativos, representando de forma justa todas as partes da sociedade. Essa é uma das abordagens estratégicas possíveis, que visa garantir que as preocupações e experiências de ambos os sexos sejam integradas de maneira significativa no *design*, implementação, monitoramento e avaliação de políticas, programas econômicos e sociais. Ao fazer isso, procura-se assegurar que tanto mulheres quanto homens possam se beneficiar equitativamente, contribuindo para a erradicação das desigualdades históricas que persistem na sociedade. O objetivo, portanto, é criar um ambiente onde a igualdade de gênero não somente seja promovida, mas se torne uma realidade tangível, capacitando todos os indivíduos a alcançarem seu pleno potencial. Dessa maneira, a luta pela igualdade de gênero não é apenas um desafio

²⁵ No original: “Technology reflects the values of its developers... It is clear that having more diverse teams working in the development of such technologies might help in identifying biases and prevent them”.

social; é uma oportunidade de transformar a sociedade em um lugar mais justo e inclusivo para todos (ECOSOC 1997 apud UNESCO, 2014, p. 15).

Portanto, é fundamental que iniciativas robustas sejam implementadas em níveis governamentais, acadêmicos e corporativos para atrair e, crucialmente, reter mulheres nas áreas de ciência e tecnologia. Sem esses esforços, o progresso na correção do desequilíbrio de gênero não somente ficará estagnado, mas também poderá comprometer o potencial da tecnologia de moldar um futuro mais inclusivo e inovador.

Ainda, o enfoque interseccional (Lamartine, 2023), herdado do feminismo tradicional, é outra diretriz fundamental do tecnofeminismo. Ele reconhece que o gênero não é a única dimensão da opressão, mas interage com fatores como raça, classe, sexualidade e outras categorias sociais (Machado e Negreiro, 2024). No desenvolvimento da IA, isso implica garantir que os conjuntos de dados reflitam a diversidade da população, contemplando diferentes gêneros, raças, classes sociais e orientações sexuais. Além disso, é essencial avaliar os impactos da IA sobre distintos grupos, sobretudo aqueles em situação de vulnerabilidade, como mulheres negras, pessoas LGBTQIA+ e indivíduos de baixa renda. Para mitigar possíveis impactos negativos, como discriminação algorítmica em contratações ou no acesso a serviços, devem ser implementadas políticas inclusivas que protejam esses grupos marginalizados.

A ética e a responsabilidade, quando aliadas ao tecnofeminismo, enfatizam a necessidade de relações de poder mais justas e transparentes. No contexto da IA, isso se traduz na criação de leis e regulamentações que obriguem as empresas a desenvolver sistemas de IA éticos, com mecanismos claros de responsabilização em casos de discriminação ou danos. A criação de comitês de ética multidisciplinares, incluindo especialistas em gênero, é fundamental para supervisionar o desenvolvimento e a implementação dessas tecnologias. Além disso, avaliações de impacto de gênero devem ser conduzidas antes do lançamento de novas tecnologias, garantindo que não perpetuem desigualdades.

Ademais, a promoção da autonomia e do empoderamento feminino são aspectos essenciais no desenvolvimento de IA sob uma perspectiva tecnofeminista. O feminismo historicamente busca garantir que as mulheres tenham controle sobre suas vidas e corpos e essa lógica se estende ao campo tecnológico. Isso implica criar ferramentas que promovam sua autonomia. Além disso, é fundamental assegurar que as tecnologias de IA respeitem a privacidade e o consentimento dos usuários, especialmente em áreas sensíveis como reconhecimento facial e coleta de dados. Para evitar a exclusão digital, o acesso equitativo às

tecnologias de IA deve ser uma prioridade, garantindo que mulheres e outros grupos marginalizados não fiquem à margem das inovações tecnológicas.

Vale ressaltar que o tecnofeminismo também desafia as estruturas de poder que perpetuam desigualdades, incluindo o capitalismo e o patriarcado. Isso significa promover modelos tecnológicos que não estejam exclusivamente concentrados em grandes corporações, mas que incluam comunidades, governos e organizações sem fins lucrativos. A tecnologia deve ser desenvolvida para o bem comum, priorizando o bem-estar social e a equidade de gênero, em vez do lucro e do controle. Nesse sentido, a participação cidadã, especialmente de grupos feministas, é elemento que garante que os sistemas inteligentes sejam democráticos e inclusivos.

Como destacam Wajcman et al. (2020, p. 3), para abordar as questões emergentes, as tecnologias digitais não devem ser vistas como ferramentas independentes e neutras em relação ao gênero, mas sim como elementos inseridos em um contexto socioeconômico e político mais amplo, que influencia seu *design*, objetivo e utilização. Nas palavras de Wajcman (2004, p. 35, tradução nossa): “Um sistema tecnológico nunca é meramente técnico: seu funcionamento no mundo real possui elementos técnicos, econômicos, organizacionais, políticos e até culturais”²⁶. Essa visão é relevante em um cenário onde ferramentas digitais são frequentemente apresentadas como soluções universais para problemas sociais.

O desenvolvimento e a implementação de novas tecnologias sempre envolvem escolhas e preferências, abrindo espaço para projetá-las de maneira a não somente evitar danos, mas também promover o bem. Como as tecnologias baseadas em dados impactam praticamente todos os aspectos da vida econômica, social e política na “Sociedade em Rede”, torna-se essencial direcionar a atenção para o modo como essas ferramentas digitais podem ser utilizadas para impulsionar a igualdade, com foco especial nos direitos das mulheres na Era Digital (Wajcman et al., 2020, p. 3).

Como exposto, o feminismo e o tecnofeminismo oferecem diretrizes concretas para enfrentar os desafios éticos e sociais no desenvolvimento de IA, ao combinar crítica teórica com ações práticas. Essas abordagens enfatizam a importância da inclusão, diversidade, transparência e responsabilidade, propondo um modelo de desenvolvimento tecnológico que não apenas previna desigualdades, mas promova ativamente a justiça social e a equidade de

²⁶ No original: “A technological system is never merely technical: its real-world functioning has technical, economic, organizational, political and even cultural elements.”

gênero. Seguir essas diretrizes possibilita a criação de tecnologias que beneficiem a todos, sem reproduzir ou ampliar as desigualdades existentes.

Enfim, outras estratégias de utilização da própria tecnologia a favor da equidade de gênero serão descritas adiante nesta tese, aproveitando-se como exemplo a crescente implementação tecnológica do Poder Judiciário brasileiro nos últimos anos, em especial, após a pandemia de COVID-19 (Otoni, 2020; Siqueira et al., 2023).

CAPÍTULO 3 – O PODER JUDICIÁRIO BRASILEIRO E O DESAFIO DO VIÉS DE GÊNERO

Este capítulo aborda a intersecção entre o Poder Judiciário brasileiro e o viés de gênero, um desafio premente que demanda atenção e reflexão crítica. À medida que a sociedade avança em direção à equidade, o sistema judiciário brasileiro, como guardião da justiça, não pode se eximir de uma análise profunda sobre sua própria composição e funcionamento. Ao explorar o tema, busca-se entender as características do perfil sociodemográfico dos magistrados no Brasil e as implicações que essa estrutura possui na aplicação do Direito.

Inicia-se a discussão com uma análise do perfil sociodemográfico do Poder Judiciário brasileiro, revelando a diversidade ou a falta dela dentro desse espaço de poder. Essa análise se justifica para compreender como a composição do Judiciário pode influenciar as decisões judiciais e refletir as desigualdades de gênero presentes na sociedade.

O que se investiga é se e como a disparidade de gênero na composição do Judiciário possui profundas implicações na neutralidade da tomada de decisões e na aplicação do Direito. O que se pressupõe, com base no estudo sobre viés de gênero, é que a predominância de um grupo específico pode resultar em visões de mundo e experiências limitadas, levando a julgamentos que não consideram adequadamente as particularidades e necessidades de todos os cidadãos. Essa falta de representatividade pode, portanto, comprometer a imparcialidade e a justiça das decisões, perpetuando desigualdades em vez de promovê-las.

Compondo a investigação, estuda-se a hermenêutica jurídica e a “neutralidade” do Direito, explorando como essas ideias interagem com a desigualdade de gênero e os vieses humanos que permeiam a prática judicial. Questiona-se a ideia de que o Direito é um campo isento de preconceitos, enfatizando como as construções sociais e culturais moldam as interpretações e decisões.

Por fim, aborda-se o “Protocolo para Julgamento com Perspectiva de Gênero” (Conselho Nacional de Justiça, 2021a) e o resgate feminista na magistratura brasileira, discutindo como essas abordagens podem contribuir para uma justiça mais equitativa. Ao longo deste capítulo, o objetivo é proporcionar uma compreensão abrangente dos desafios enfrentados pelo Poder Judiciário brasileiro em relação ao viés de gênero, com a finalidade última de propor caminhos para a transformação dessa realidade, que também sofre mudanças com a crescente implementação da IA no processo decisório judicial.

3.1 O perfil sociodemográfico do Poder Judiciário brasileiro

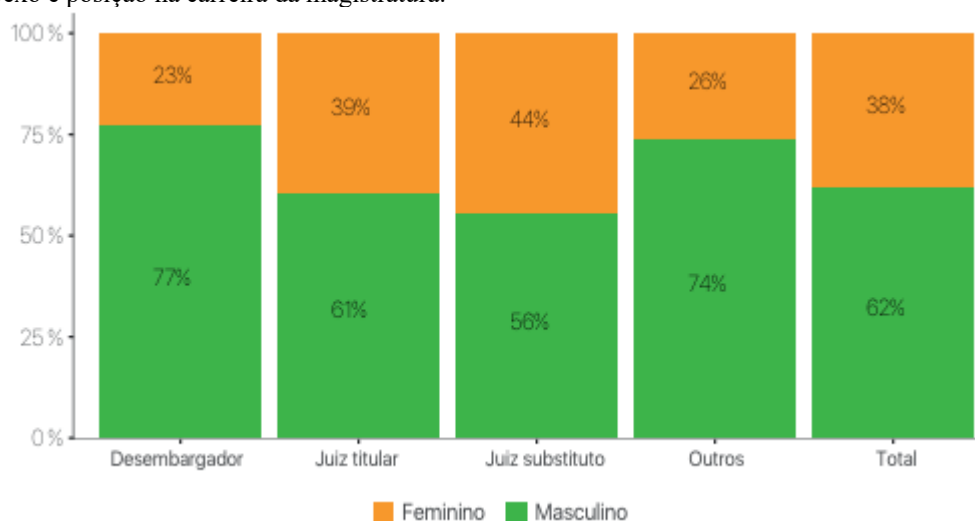
Em 2018, foi realizada uma pesquisa pelo CNJ, via questionário respondido pelos magistrados brasileiros de todos os 90 tribunais do país, buscando identificar as suas características demográficas, sociais e profissionais. Esta pesquisa foi intitulada “Perfil Sociodemográfico dos Magistrados Brasileiros”. De um total de 18.168, 1.348 magistrados ativos participaram da pesquisa (62,5% do total) (Poder Judiciário e Conselho Nacional de Justiça, 2018, p. 5).

Foi traçado o perfil sociodemográfico do Poder Judiciário à época, considerando sexo, idade, local de nascimento, estado civil, quantidade de filhos e etnia racial (cor ou raça declarada). No que diz respeito ao gênero, identificou-se que apenas 38% da magistratura brasileira era composta por mulheres, sendo a Justiça do Trabalho a que teria maior participação (47%). A Justiça Estadual seria composta por 36% de mulheres e a Justiça Federal, por 32% (Poder Judiciário e Conselho Nacional de Justiça, 2018, p. 8).

A análise da distribuição de gênero conforme o ano de ingresso na magistratura revela que, entre os magistrados ativos que começaram sua carreira até 1990, somente 25% eram mulheres. Para aqueles que ingressaram entre 1991 e 2000, a participação feminina sobe para 40%. Entre os que entraram de 2001 a 2010, as mulheres representam 41% dos novos magistrados, enquanto essa proporção cai para 37% entre os que ingressaram a partir de 2011. A pesquisa também destaca uma diminuição na representatividade feminina entre os novos magistrados a partir de 2011 em todos os segmentos (Poder Judiciário e Conselho Nacional de Justiça, 2018, p. 8).

Em relação à carreira da magistratura, as mulheres constituem 44% dos juízes substitutos, 39% dos juízes titulares e 23% dos desembargadores (Poder Judiciário e Conselho Nacional de Justiça, 2018, p. 8). A Figura 5 colacionada da pesquisa aqui em exposição, ilustra de maneira ímpar como a disparidade de gênero acontece consoante a posição na carreira (quanto mais elevada, menos mulheres se vê):

Figura 5. Sexo e posição na carreira da magistratura.



Fonte: Poder Judiciário e Conselho Nacional de Justiça, 2018, p. 10.

Além da faixa etária dos magistrados e da sua origem demográfica, a pesquisa revela que a maioria dos magistrados se identifica como branca (80,3%), seguida por 18,1% que se declaram negros (sendo 16,5% pardos e 1,6% pretos) e 1,6% de origem asiática (amarelo). Somente 11 magistrados se declararam indígenas (Poder Judiciário e Conselho Nacional de Justiça, 2018, p. 8-9).

O estudo intitulado “Diagnóstico da participação feminina no Poder Judiciário” (Poder Judiciário e Conselho Nacional de Justiça, 2019), publicado em 2019 pelo Poder Judiciário em conjunto com o CNJ corrobora com a identificação da ausência de equidade de gênero na história da composição do Judiciário brasileiro:

O Poder Judiciário brasileiro é composto em sua maioria por magistrados do sexo masculino, com apenas 38,8% de magistradas em atividade. A participação feminina na magistratura é ainda menor se considerar o total de magistrados que atuaram nos últimos 10 anos, com 37,6%.

Apesar de ainda baixo, houve evolução do percentual de magistradas em relação à composição do Poder Judiciário no ano de 1988, quando as mulheres ocupavam somente 24,6% dos cargos da magistratura, [...] (Poder Judiciário e Conselho Nacional de Justiça, 2023, p. 7).

A Figura 6, retirada do Diagnóstico supramencionado, informa que a predominância de homens no Poder Judiciário no Brasil ocorre desde 1988 e continua longe de ser substituída pela equidade com o gênero feminino. Em 1988, apenas 24,6% dos magistrados em atividade eram mulheres, enquanto 75% eram homens. Nos últimos 10 anos, a média de magistradas aumentou para 37,6%, indicando um crescimento na participação feminina. No entanto, em 2018, o percentual de magistradas teve um aumento de 0,8%, mostrando que, embora exista

uma tendência de aumento na representação feminina na magistratura, este aumento é muito devagar, fazendo com que a hegemonia nos tribunais prevaleça por muitos anos.

Figura 6. Percentual de magistradas no Poder Judiciário brasileiro.



Fonte: Poder Judiciário e Conselho Nacional de Justiça, 2019, p. 8.

Já no ano de 2023, o CNJ elaborou um relatório denominado “2º Censo do Poder Judiciário 2023”, que buscou atualizar as informações a respeito do perfil demográfico, social, econômico, dentre outros, do Poder Judiciário, que haviam sido colhidas no ano de 2013. Somente 31,44% dos servidores e 40,51% dos juízes efetivamente contribuíram com a pesquisa desenvolvida (Poder Judiciário; Conselho Nacional de Justiça, 2023, p. 12), sendo que destes, 40,3% se declararam como do sexo feminino, 59,3% do sexo masculino e 0,3% preferiram não informar (Poder Judiciário; Conselho Nacional de Justiça, 2023, p. 21).

No censo foram relacionadas as informações de sexo e raça, identificando-se que as mulheres brancas dominam a maioria dos cargos entre as magistradas, representando 83,2% das juízas titulares, 85% das juízas substitutas e 92,3% das desembargadoras (Poder Judiciário; Conselho Nacional de Justiça, 2023, p. 19).

As magistradas pardas correspondem a 12,6% entre as juízas titulares, 11,1% entre as substitutas e 11,2% das que ocupam cargos de desembargadoras. Por sua vez, as juízas negras representam apenas 1,3% entre as titulares, 2,1% entre as substitutas e 1,5% entre as desembargadoras. Já as juízas que se identificam como amarelas ocupam 1,7% dos cargos de juízas titulares, 0,9% dos cargos de juízas substitutas e 1% entre as desembargadoras (Poder Judiciário; Conselho Nacional de Justiça, 2023, p. 19).

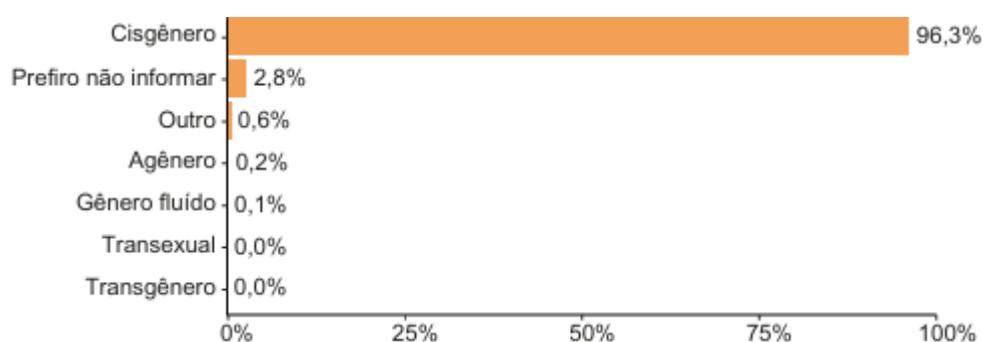
Em relação aos magistrados do sexo masculino, a cor branca é predominante em todas as categorias. Os magistrados brancos ocupam 81% dos cargos de juízes titulares, 80% dos juízes substitutos e 87,8% dos desembargadores. Entre os magistrados pardos, essa população representa 14,9% dos juízes titulares, 16,1% dos substitutos e 10,1% dos desembargadores.

Assim como entre as magistradas, os menores percentuais entre os magistrados do sexo masculino também são observados nas raças preta e amarela. Os juízes negros constituem 1,3% entre os titulares, 1,7% entre os substitutos e 1,1% entre os desembargadores. No caso dos juízes amarelos, eles representam 1,2% dos titulares, 1,3% dos substitutos e 0,6% dos desembargadores (Poder Judiciário; Conselho Nacional de Justiça, 2023, p. 19).

O mesmo censo traz alguns apontamentos a respeito do relatório de pesquisa “Discriminação e Violência contra a População LGBTQIA+”, que descreve a identidade de gênero como uma vivência pessoal e interna, que pode ou não coincidir com o sexo atribuído ao nascimento. Para classificar a identidade de gênero, o relatório utiliza as seguintes categorias: cisgênero, quando a identidade de gênero corresponde ao sexo designado ao nascimento; agênero, caracterizado pela ausência de identidade de gênero; gênero fluido, para aqueles que não se identificam com uma única identidade ou papel de gênero; transexual, que são indivíduos que se reconhecem com um gênero distinto do sexo biológico e podem optar por tratamentos hormonais ou cirúrgicos para alinhar-se a essa identidade; e transgênero, que se refere a pessoas cuja identidade de gênero não corresponde ao sexo atribuído ao nascimento (Poder Judiciário; Conselho Nacional de Justiça, 2023, p. 22).

Na pesquisa, foram identificados 96,3% magistrados que se identificaram como cisgênero, 2,8% preferiram não informar, 0,3% se declararam como agênero e 0,1% como gênero fluído, sendo elaborado o gráfico constante na Figura 7:

Figura 7. Identidade de gênero no Poder Judiciário - 2023.



Fonte: Poder Judiciário; Conselho Nacional de Justiça, 2023, p. 23.

Adicionalmente, há estudos como o da “Participação Feminina no Poder Judiciário”, realizado em 2019, atualizado em 2023 e publicado no relatório “Justiça em Números” produzido pelo CNJ. Nesse documento, são apresentadas informações recentes sobre a Política Nacional de Incentivo à Participação Feminina no Poder Judiciário, que foi estabelecida pelo

CNJ por meio da Resolução CNJ n. 255, de setembro de 2018, visando garantir a igualdade de gênero em todos os segmentos dos tribunais (Conselho Nacional de Justiça, 2018a).

A Resolução CNJ n. 255, de setembro de 2018, destaca a importância da construção de espaços democráticos e de igualdade entre homens e mulheres, indicando a assimetria na ocupação de cargos quanto à representatividade feminina no Poder Judiciário. O objetivo da criação da “Política Nacional de Incentivo à Participação Feminina no Poder Judiciário” é de oferecer proposições concretas para a ampliação da representação feminina no Judiciário (Conselho Nacional de Justiça, 2018a).

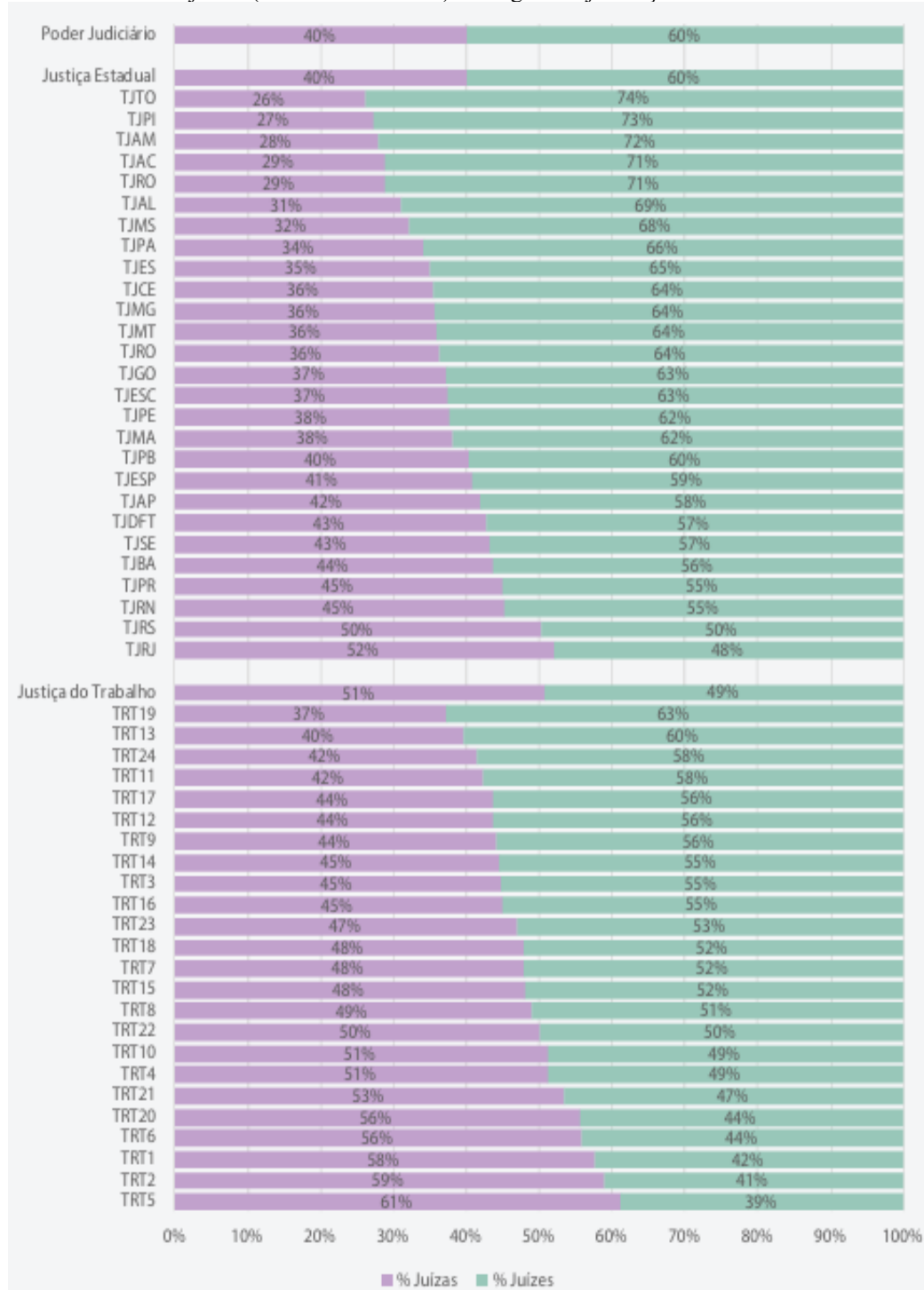
Subsequente à Resolução, em 2022, as magistradas e servidoras de todo o Brasil publicaram a “Carta de Brasília pela Igualdade de Gênero no Poder Judiciário” (2022) durante o “Seminário Mulheres na Justiça: Novos Rumos da Resolução CNJ n. 255”. Na carta, as mulheres que compõem parte do Poder Judiciário brasileiro propõem uma série de tarefas a fim de efetivar a igualdade substancial entre homens e mulheres no Poder Judiciário. O documento inicia destacando as desigualdades políticas, sociais, econômicas e salariais existentes entre homens e mulheres e, em especial, se preocupa com sub-representação das mulheres em espaços institucionais de poder, tais como nos Tribunais e no CNJ (Carta de Brasília pela Igualdade de Gênero no Poder Judiciário, 2022).

No ano de 2023, na “2ª Edição do Seminário Mulheres na Justiça: Novos Rumos da Resolução CNJ n. 255”, foi publicada uma nova carta com os mesmos propósitos, mas buscando novas ações concretas para a garantia da paridade de gênero na carreira da magistratura brasileira (Carta de Brasília pela Igualdade de Gênero no Poder Judiciário de 2023, 2023).

Quanto ao estudo da “Participação Feminina no Poder Judiciário” atualizado em 2023 (Conselho Nacional de Justiça, 2023a), identifica-se que, em 2022, embora a equidade seja também uma luta no 1º grau de jurisdição na maioria dos tribunais brasileiros, a participação de mulheres como ministras e desembargadoras se encontra com percentuais inferiores a 25%.

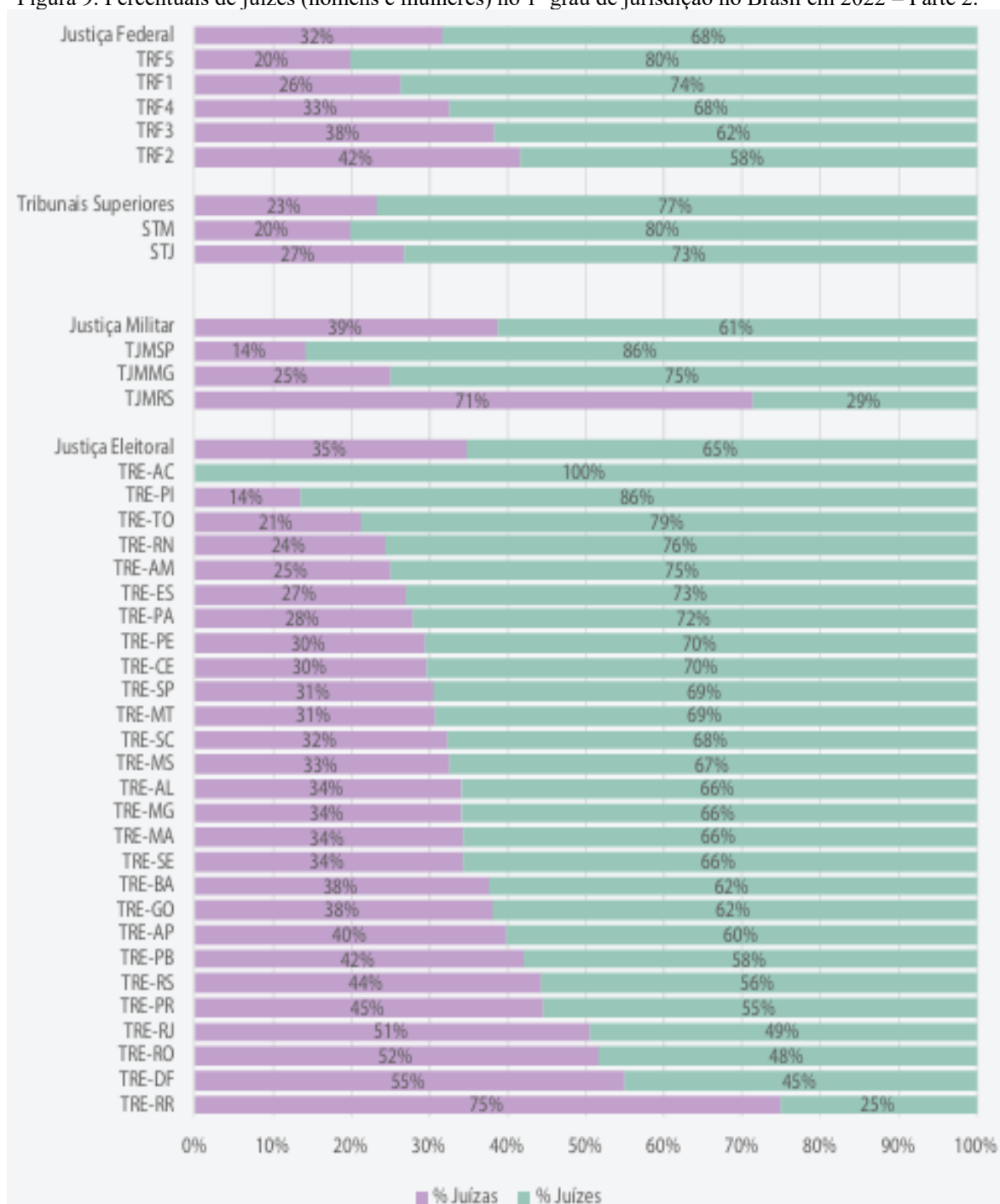
Confirmam-se as Figuras 5 a 8 retiradas do estudo, que ilustram de maneira ímpar a desigualdade entre homens e mulheres no Poder Judiciário brasileiro. Nas Figuras 8 e 9, temos uma perspectiva relacionada ao 1º grau de jurisdição. Nas Figuras 10 e 11, consta a comparação no que diz respeito aos cargos de ministros e desembargadores.

Figura 8. Percentuais de juízes (homens e mulheres) no 1º grau de jurisdição no Brasil em 2022 – Parte 1.



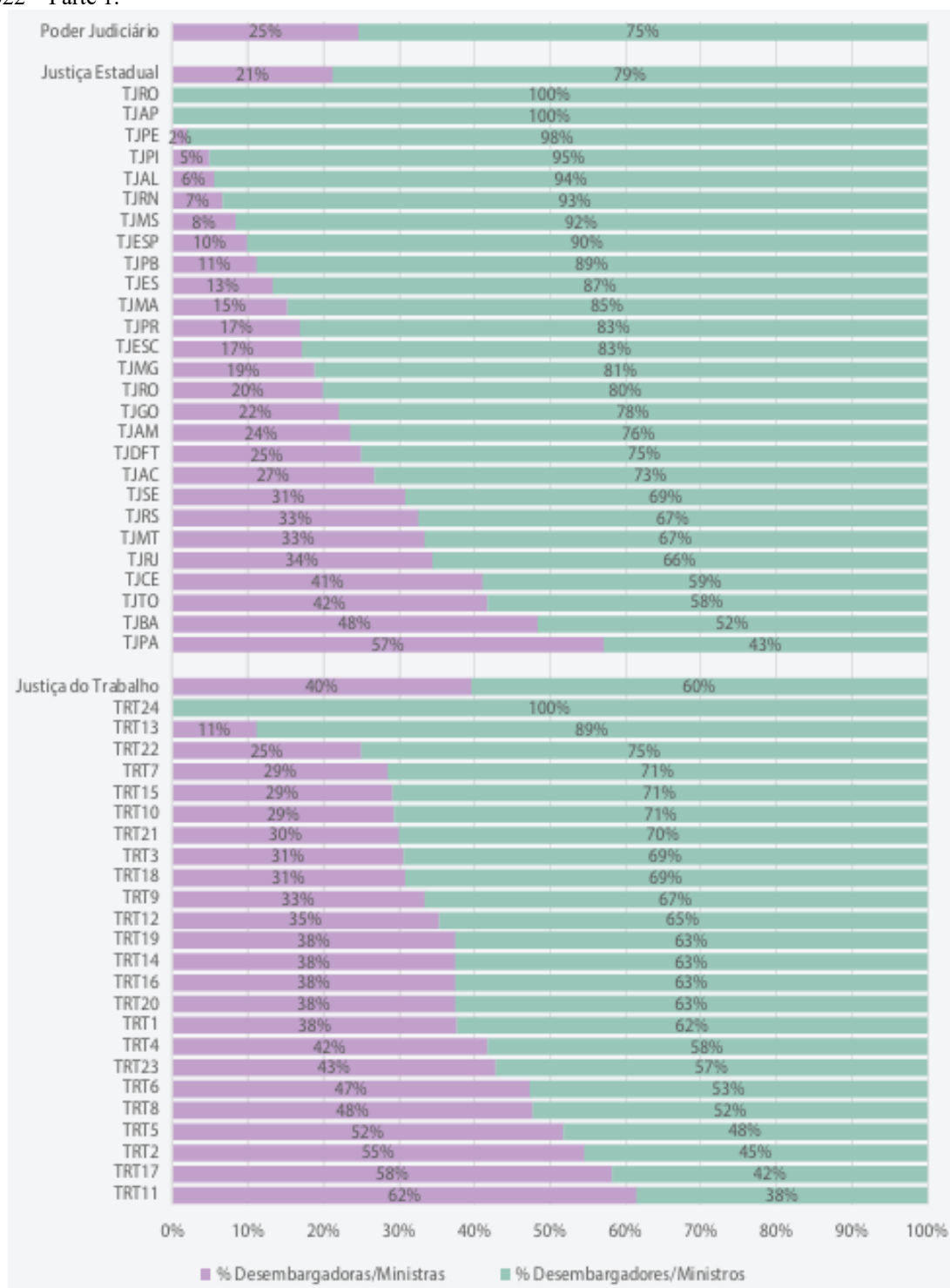
Fonte: Conselho Nacional de Justiça, 2023a, p. 12.

Figura 9. Percentuais de juízes (homens e mulheres) no 1º grau de jurisdição no Brasil em 2022 – Parte 2.



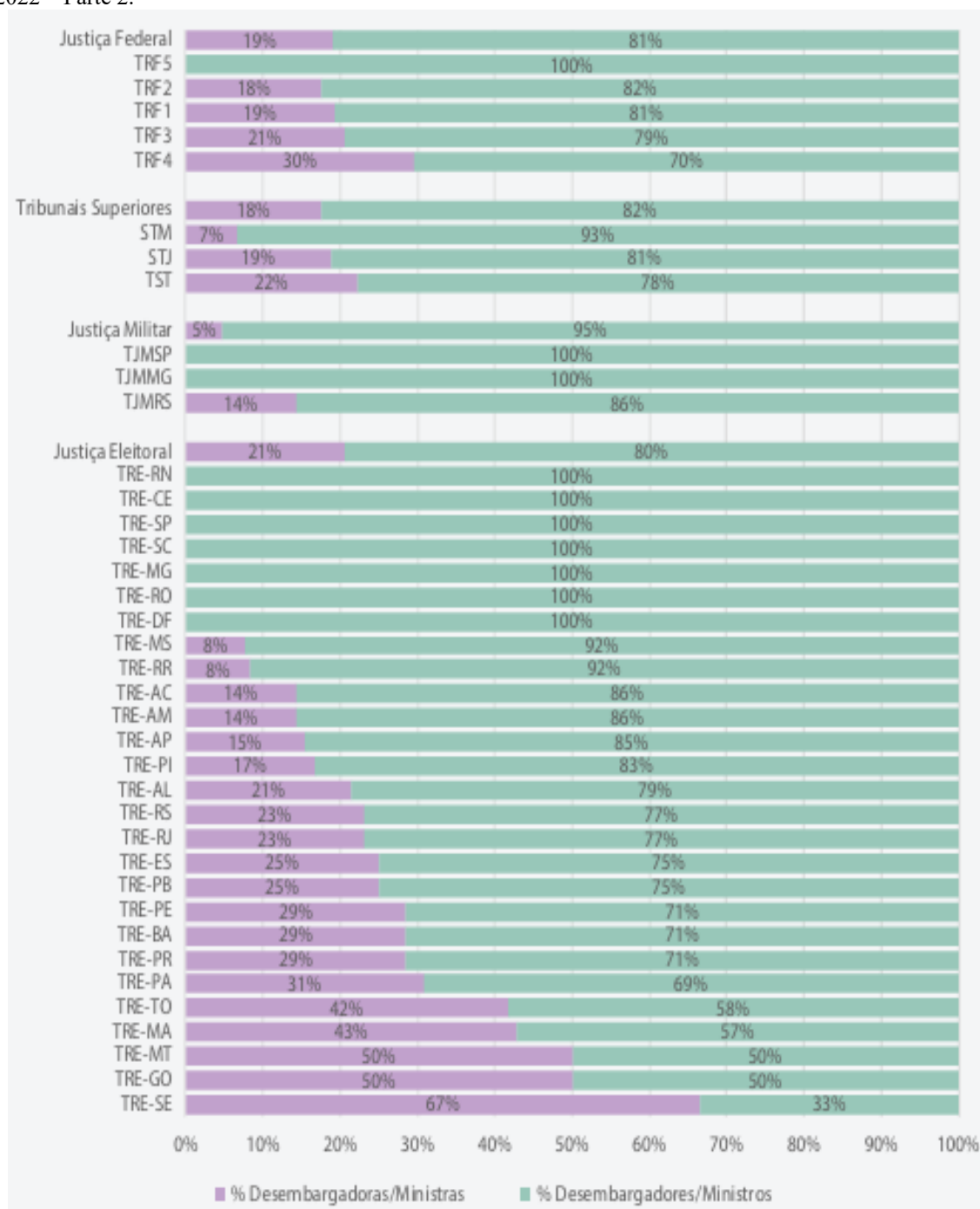
Fonte: Conselho Nacional de Justiça, 2023a, p. 13.

Figura 10. Percentuais de desembargadores e ministros (homens e mulheres) no Poder Judiciário brasileiro em 2022 – Parte 1.



Fonte: Conselho Nacional de Justiça, 2023a, p. 15.

Figura 11. Percentuais de desembargadores e ministros (homens e mulheres) no Poder Judiciário brasileiro em 2022 – Parte 2.



Fonte: Conselho Nacional de Justiça, 2023a, p. 16.

Da Figura 8, revela-se que na Justiça Estadual, a média geral é de 40% de juízas e 60% de juízes. A proporção de juízas varia consideravelmente entre os tribunais, indo de 26% no Tribunal de Justiça do Tocantins (TJTO) a 52% no Tribunal de Justiça do Rio de Janeiro (TJRJ). Por outro lado, a proporção de juízes varia de 48% no TJRJ a 74% no TJTO. Tribunais como o TJTO, Tribunal de Justiça do Piauí (TJPI) e Tribunal de Justiça do Amapá (TJAM) apresentam as menores proporções de mulheres, com 26%, 27% e 28% de juízas, respectivamente. Em

contraste, tribunais como o TJRJ e o Tribunal de Justiça do Rio Grande do Sul (TJRS) mostram uma distribuição mais equilibrada, com 52% e 50% de juízas, respectivamente.

Já na Justiça do Trabalho, a média geral é mais equânime, com 51% de juízas e 49% de juízes. No entanto, ainda há variações significativas entre os tribunais. A proporção de juízas varia de 37% no Tribunal Regional do Trabalho da 19ª Região (TRT19) a 61% no Tribunal Regional do Trabalho da 5ª Região (TRT5), enquanto a proporção de juízes varia de 39% no TRT5 a 63% no TRT19.

Seguindo para a Figura 9, que também apresenta a distribuição percentual de juízas e juízes no primeiro grau de jurisdição do Poder Judiciário brasileiro, dividido em Justiça Federal, Tribunais Superiores, Justiça Militar e Justiça Eleitoral, a análise revela uma desigualdade significativa entre homens e mulheres nesses cargos, com variações importantes entre os diferentes tribunais.

Na Justiça Federal, a média geral é de 32% de juízas e 68% de juízes. A proporção de juízas varia de 20% no Tribunal Regional Federal da 5ª Região (TRF5) a 42% no Tribunal Regional Federal da 2ª Região (TRF2), enquanto a proporção de juízes varia de 58% no TRF2 a 80% no TRF5. Os tribunais com maior desigualdade são o TRF5 (20% de juízas) e o Tribunal Regional Federal da 1ª Região (TRF1) (26% de juízas), enquanto o TRF2 (42% de juízas) e o Tribunal Regional Federal da 3ª Região (TRF3) (38% de juízas) apresentam uma distribuição mais equilibrada.

Nos Tribunais Superiores, a média geral é de 23% de juízas e 77% de juízes. A proporção de juízas varia de 20% no Superior Tribunal Militar (STM) a 27% no Superior Tribunal de Justiça (STJ), enquanto a proporção de juízes varia de 73% no STJ a 80% no STM. Ambos os tribunais, STM e STJ, apresentam disparidades significativas.

Na Justiça Militar, a média geral é de 39% de juízas e 61% de juízes. A proporção de juízas varia de 0% no Tribunal de Justiça Militar de São Paulo (TJMSP) e no Tribunal de Justiça Militar de Minas Gerais (TJMMG) a 14% no Tribunal de Justiça Militar do Rio Grande do Sul (TJMRS).

Na Justiça Eleitoral, a média geral é de 35% de juízas e 65% de juízes. A proporção de juízas varia de 0% em sete tribunais a 67% no Tribunal Regional Eleitoral de Sergipe (TRE-SE), enquanto a proporção de juízes varia de 0% no TRE-AC a 86% no TRE-PI. Os tribunais com maior desigualdade são o Tribunal Regional Eleitoral do Rio Grande do Norte (TRE-RN), Tribunal Regional Eleitoral do Ceará (TRE-CE), Tribunal Regional Eleitoral de São Paulo (TRE-SP), Tribunal Regional Eleitoral de Santa Catarina (TRE-SC), Tribunal Regional

Eleitoral de Minas Gerais (TRE-MG), Tribunal Regional Eleitoral de Rondônia (TRE-RO) e Tribunal Regional Eleitoral do Distrito Federal (TRE-DF), seguidos do Tribunal Regional Eleitoral do Mato Grosso do Sul (TRE-MS) e do Tribunal Regional Eleitoral de Roraima (TRE-RR).

Assim, da análise conjunta das duas Figuras, verifica-se que, em todas as áreas do Poder Judiciário brasileiro, há uma predominância de juízes em relação a juízas, embora a Justiça do Trabalho apresente um equilíbrio relativamente maior. No entanto, a Justiça Estadual ainda demonstra uma desigualdade mais acentuada, enquanto a Justiça Militar e os Tribunais Superiores figuram entre as áreas com as maiores disparidades. Esses dados ressaltam a necessidade de políticas para promover a igualdade de gênero no Poder Judiciário brasileiro, visando uma distribuição mais equitativa entre homens e mulheres nos cargos de juízes.

Quanto à distribuição percentual de desembargadoras/ministras e desembargadores/ministros no Poder Judiciário brasileiro, dividido em Justiça Estadual e Justiça do Trabalho, a Figura 10 também revela uma desigualdade significativa entre homens e mulheres nesses cargos, com uma predominância de homens na maioria dos tribunais.

Na Justiça Estadual, a média geral é de 21% de desembargadoras/ministras e 79% de desembargadores/ministros. A proporção de mulheres varia de 0% a 57% entre o TJRO e o TJAP. Tribunais como o Tribunal de Justiça de Pernambuco (TJPE) (2% de mulheres), TJPI (5% de mulheres), Tribunal de Justiça de Alagoas (TJAL) (6% de mulheres) e Tribunal de Justiça do Rio Grande do Norte (TJRN) (7% de mulheres) também apresentam menores proporções de mulheres, indicando uma significativa desigualdade de gênero. Por outro lado, tribunais como o Tribunal de Justiça da Bahia (TJBA) (48% de mulheres) e o Tribunal de Justiça do Pará (TJPA) (57% de mulheres) mostram uma distribuição mais equilibrada ou até mesmo uma predominância de mulheres, destacando ser possível alcançar uma maior representação feminina.

Na Justiça do Trabalho, a média geral é de 40% de desembargadoras/ministras e 60% de desembargadores/ministros. A proporção de mulheres varia de 0% no Tribunal Regional do Trabalho da 24ª Região (TRT24) a 62% no Tribunal Regional do Trabalho da 11ª Região (TRT11).. Tribunais como o TRT24 (100% de homens), Tribunal Regional do Trabalho da 13ª Região (TRT13) (11% de mulheres) e Tribunal Regional do Trabalho da 22ª Região TRT22 (25% de mulheres) apresentam as menores proporções de mulheres, evidenciando disparidades significativas.

Por sua vez, a Figura 12, dividida em Justiça Federal, Tribunais Superiores, Justiça Militar e Justiça Eleitoral, identifica uma desigualdade significativa entre homens e mulheres, com uma predominância de homens na maioria dos tribunais.

Na Justiça Federal, a média geral é de 19% de desembargadoras/ministras e 81% de desembargadores/ministros. A proporção de mulheres varia de 0% no TRF5 a 30% no Tribunal Regional Federal da 4ª Região (TRF4). Tribunais como o TRF2 (18% de mulheres) e o TRF1 (19% de mulheres) também apresentam as menores proporções de mulheres, indicando uma significativa desigualdade de gênero.

Nos Tribunais Superiores, a média geral é de 18% de desembargadoras/ministras e 82% de desembargadores/ministros. A proporção de mulheres varia de 7% no STM a 22% no Tribunal Superior do Trabalho (TST), enquanto a proporção de homens varia de 78% no TST a 93% no STM. O STM (7% de mulheres) e o STJ (19% de mulheres) apresentam as menores proporções de mulheres, evidenciando disparidades significativas.

Na Justiça Militar, a média geral é de apenas 5% de desembargadoras/ministras e 95% de desembargadores/ministros. A proporção de mulheres varia de 0% no TJMSP e TJMMG a 14% no TJMMS.

Na Justiça Eleitoral, a média geral é de 21% de desembargadoras/ministras e 80% de desembargadores/ministros. A proporção de mulheres varia de 0% em sete tribunais a 67% no TRE-SE.

Da análise conjunta das Figuras 10 e 11, tem-se uma predominância de homens nos cargos de desembargadores e ministros no Poder Judiciário brasileiro, com variações significativas entre os diferentes tribunais. A Justiça Militar apresenta considerável desigualdade, com uma média de apenas 5% de mulheres, enquanto a Justiça Estadual também demonstra uma baixa representatividade feminina, com uma média de 21% de mulheres e tribunais como o TJRO e o TJAP onde não há desembargadoras. Já a Justiça do Trabalho apresenta uma distribuição mais equilibrada, mas ainda há tribunais com disparidades significativas, como o TRT24, composto por 100% de homens.

Assim, o estudo da “Participação Feminina no Poder Judiciário de 2023” (Conselho Nacional de Justiça, 2023a) aponta dados que reforçam a necessidade de políticas para promover a igualdade de gênero no Poder Judiciário brasileiro, garantindo uma distribuição mais equitativa entre homens e mulheres nos cargos de desembargadores e ministros. Embora existam avanços pontuais, ainda há um longo caminho a percorrer para assegurar uma representação mais justa em todas as esferas do Judiciário.

Sabendo-se da disparidade de gênero interna do Poder Judiciário e buscando assegurar uma representatividade feminina justa para além da composição do próprio Judiciário, além das resoluções acima mencionadas, pelo CNJ também foram editadas as Resoluções de ns. 270 de 11 de dezembro de 2018 (Conselho Nacional de Justiça, 2018b), que dispõe sobre o uso de nome social, 284 de 05 de junho de 2019 (Conselho Nacional de Justiça, 2019), que institui o “Formulário Nacional de Avaliação de Risco” buscando prevenir e enfrentar situações de violência doméstica e familiar contra a mulher, 348 de 13 de outubro de 2020 (Conselho Nacional de Justiça, 2020c), que estabelece diretrizes, no âmbito criminal, com relação ao tratamento da população lésbica, gay, bissexual, transexual, travesti ou intersexo. Há ainda a Resolução n. 376, de 02 de março de 2021 (Conselho Nacional de Justiça, 2021b), que dispõe sobre o emprego obrigatório da flexão de gênero para nomear profissão ou demais designações na comunicação social e institucional do Poder Judiciário.

Em 2023, o CNJ aprovou mais cinco resoluções que tratam sobre paridade de gênero e enfrentamento à violência contra a mulher. A Resolução n. 492 de 17 de março de 2023 (Conselho Nacional de Justiça, 2023b), estabelece as diretrizes para adoção de Perspectiva de Gênero nos julgamentos em todo o Poder Judiciário e institui obrigatoriedade de capacitação de magistrados e magistradas, relacionada a direitos humanos, gênero, raça e etnia, em perspectiva interseccional, bem como cria o “Comitê de Acompanhamento e Capacitação sobre Julgamento com Perspectiva de Gênero no Poder Judiciário” e o “Comitê de Incentivo à Participação Institucional Feminina no Poder Judiciário”.

A Resolução n. 525, de 27 de setembro de 2023, estabelece diretrizes para promover a igualdade de gênero no Poder Judiciário. A norma determina que, nos tribunais de 2º grau que ainda não tenham alcançado a proporção de 40% a 60% de magistradas e magistrados oriundos da carreira, as promoções por merecimento devem seguir uma lógica de editais alternados, sendo uns mistos (abertos a homens e mulheres) e outros exclusivos para mulheres. Essa alternância deve continuar se atingir a paridade de gênero nesses tribunais (Conselho Nacional de Justiça, 2023c).

Por sua vez, a Resolução n. 532, de 16 de novembro de 2023 (Conselho Nacional de Justiça, 2023d), busca combater a discriminação à orientação sexual e à identidade de gênero nos processos de adoção.

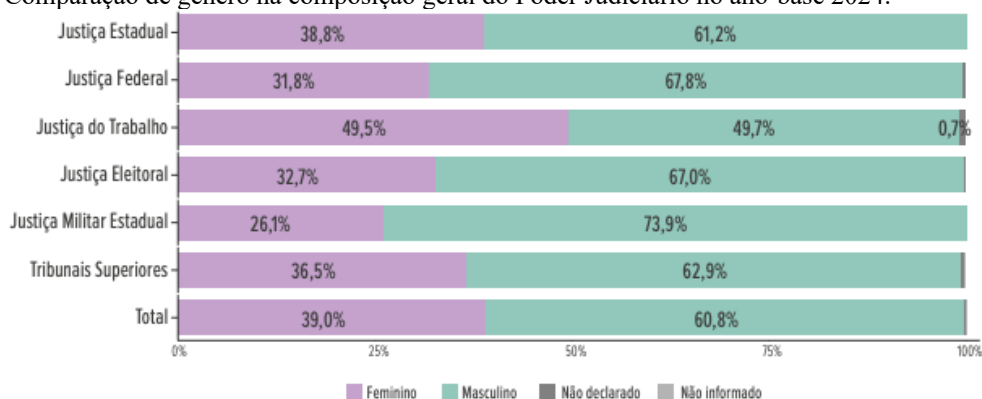
Já a Resolução n. 540, de 18 de dezembro de 2023, alterou a Resolução n. 255/2018, buscando proporcionar maior participação feminina em alguns cargos, comissões, grupos de trabalho, eventos etc. Também prevê a formação de um “Repositório Nacional de Mulheres

Juristas” e a integração, no calendário anual do CNJ, de um seminário nacional voltado ao fortalecimento e aperfeiçoamento da “Política de Incentivo à Participação Institucional Feminina” (Conselho Nacional de Justiça, 2023e). A Resolução n. 542, de 18 de dezembro de 2023, cria o “Fórum Nacional de Enfrentamento à Violência Contra a Mulher” (FONAVIM), cujo objetivo é melhorar a atuação do Poder Judiciário no combate à violência contra a mulher (Conselho Nacional de Justiça, 2023f).

No relatório “Justiça em Números” publicado em 2024 (Conselho Nacional de Justiça, 2024a), identificou-se que até o final de abril de 2024, 36,8% da magistratura brasileira é composta por mulheres, com destaque para as juízas substitutas de 1º grau, que representam 41,68% do total. Por segmento, a Justiça do Trabalho apresenta o maior percentual de mulheres magistradas (39,7%), seguida da Justiça Estadual (38,2%), Justiça Eleitoral (32,9%), Justiça Federal (31,3%), Justiça Militar Estadual (22,2%) e Tribunais Superiores (23,2%). A ocupação de cargos também reflete essa disparidade: apenas 18,8% das ministras e 23,9% das desembargadoras são mulheres, enquanto entre juízas de 1º grau a participação sobe para 39%. Entre as servidoras do Judiciário, a presença feminina é mais expressiva, com média nacional de 53,5%. Por ramo de justiça, os percentuais são: Justiça Estadual (56,9%), Justiça Federal (49%), Tribunais Superiores (48%), Justiça Militar Estadual (47,7%) e Justiça do Trabalho (42,6%).

O mesmo relatório, publicado em 2025, revela que atualmente as mulheres representam 38,80% do total de magistrados no Poder Judiciário brasileiro, enquanto os homens correspondem a 61,2%. Entre os diferentes segmentos da Justiça, apenas a Justiça do Trabalho – com 49,5% de magistradas – apresenta um índice acima da média nacional. Nos demais ramos, os percentuais são inferiores: Justiça Militar Estadual (26,1%), Justiça Federal (31,8%), Justiça Eleitoral (32,7%) e Tribunais Superiores (36,5%). Já na Justiça Estadual, o percentual é o mesmo da média geral do país (Conselho Nacional de Justiça, 2025a, p. 55-56). A Figura 12 representa os dados gerais relativos ao relatório “Justiça em Números 2025” (ano-base 2024).

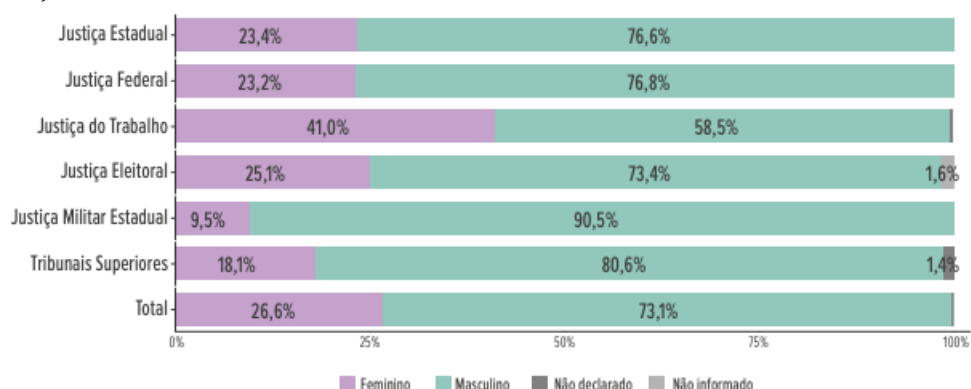
Figura 12. Comparação de gênero na composição geral do Poder Judiciário no ano-base 2024.



Fonte: Conselho Nacional de Justiça, 2025a, p. 56.

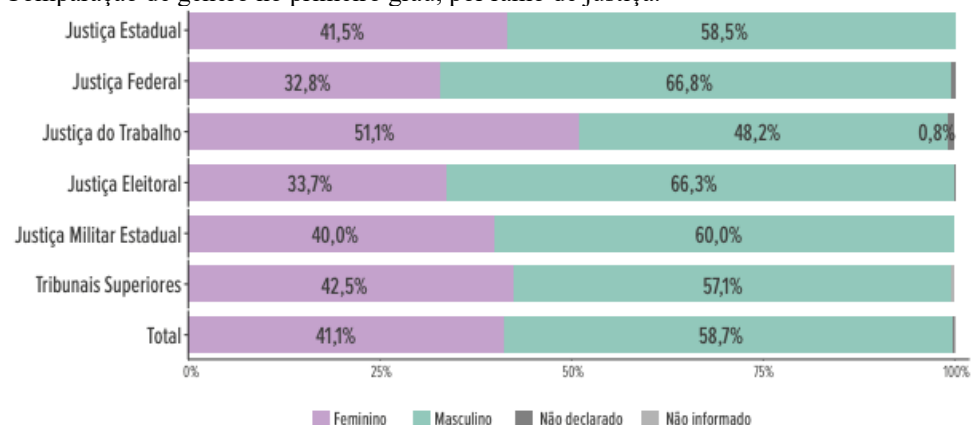
Também em 2025, a análise da participação feminina conforme o cargo ocupado evidencia uma tendência de redução do número de mulheres à medida que se ascende na carreira. As Figuras 13 e 14, retiradas do relatório, ilustram o problema (Conselho Nacional de Justiça, 2025a, p. 57).

Figura 13. Comparação de gênero de ministros nos Tribunais Superiores e de desembargadores nos demais ramos de justiça.



Fonte: Conselho Nacional de Justiça, 2025a, p. 57.

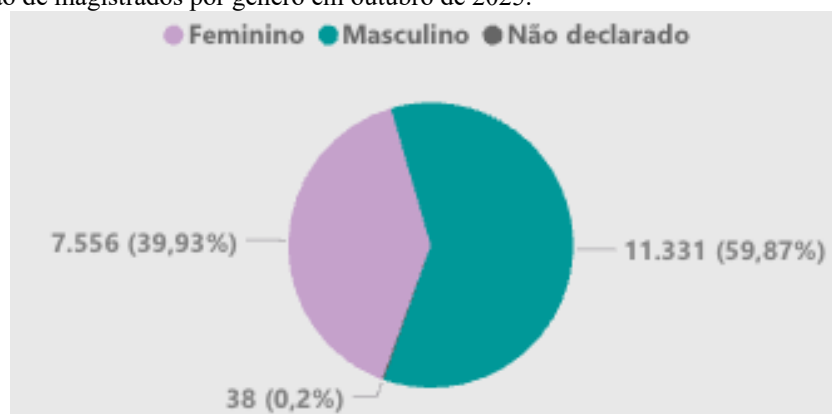
Figura 14. Comparação de gênero no primeiro grau, por ramo de justiça.



Fonte: Conselho Nacional de Justiça, 2025a, p. 57.

Ademais, em consulta ao “Módulo de pessoal e estrutura judiciária mensal do Poder Judiciário” (Sistema MPM), que se trata do sistema empregado pelos tribunais para o envio mensal de dados referentes às suas unidades judiciárias de primeiro e segundo graus, bem como às unidades de apoio direto à atividade jurisdicional, aos magistrados, servidores e demais integrantes do quadro auxiliar (Conselho Nacional de Justiça, [20–]), no mês de outubro de 2025, de um total de 19.039 magistrados, apenas 7.556 são mulheres. A Figura 15 retrata a distribuição por sexo em tal data.

Figura 15. Distribuição de magistrados por gênero em outubro de 2025.



Fonte: Módulo de pessoal e estrutura judiciária mensal do Poder Judiciário, 2025.

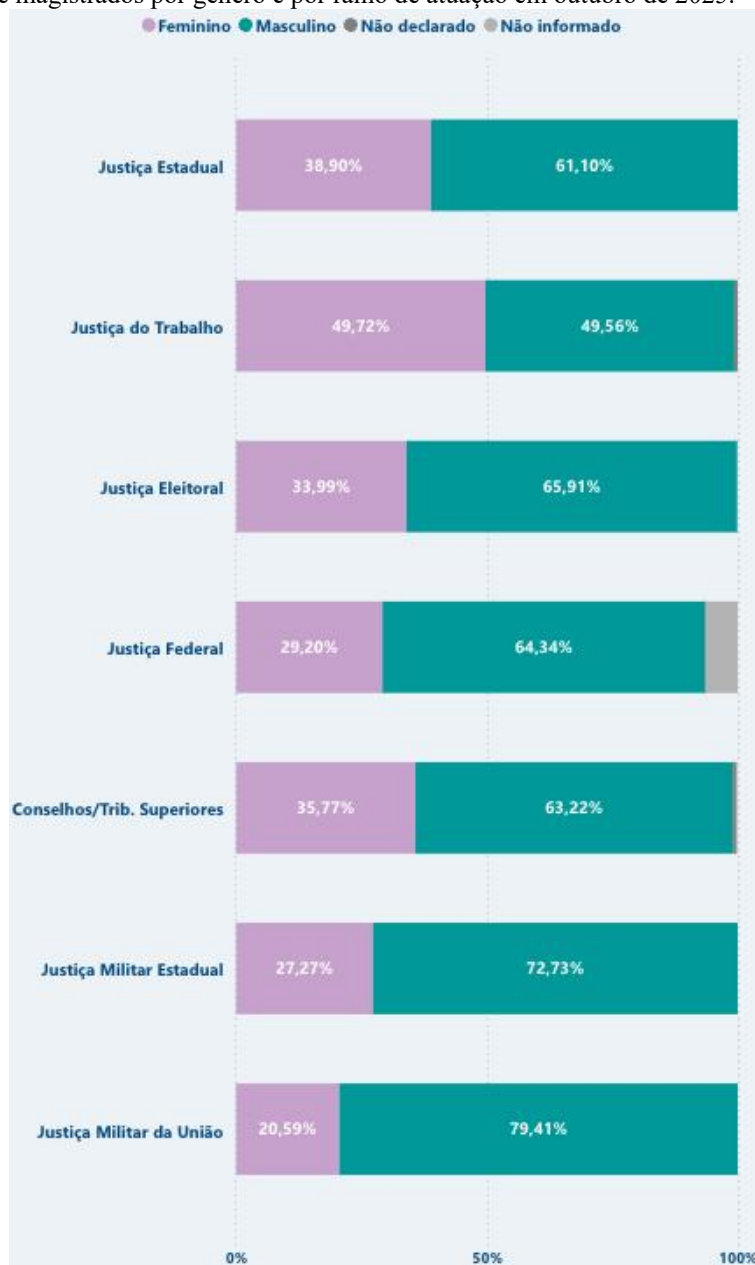
A Figura 16, por sua vez, ilustra a porcentagem de homens e mulheres em outubro de 2025 em cada um dos ramos da justiça brasileira. Ela evidencia a persistente desigualdade de gênero na composição do Poder Judiciário brasileiro. Em todos os ramos da Justiça, os homens continuam a representar a maioria da magistratura, embora haja variações significativas entre os segmentos (Módulo de pessoal e estrutura judiciária mensal do Poder Judiciário, 2025).

A Justiça do Trabalho é o único ramo que se aproxima da paridade, com 49,72% de magistradas, enquanto a Justiça Estadual também apresenta participação relativamente expressiva, com 38,90% de mulheres. Nos demais ramos, a presença feminina é substancialmente menor: 33,99% na Justiça Eleitoral, 29,20% na Justiça Federal, 35,77% nos Conselhos e Tribunais Superiores, 27,27% na Justiça Militar Estadual e apenas 20,59% na Justiça Militar da União (Módulo de pessoal e estrutura judiciária mensal do Poder Judiciário, 2025).

Esses percentuais revelam um padrão estrutural em que a participação das mulheres diminui à medida que se elevam os níveis hierárquicos e de poder institucional. Os espaços de maior prestígio e decisão permanecem majoritariamente masculinos, refletindo as barreiras históricas e culturais que ainda limitam a presença feminina nas posições mais altas da carreira

judicial. Assim, apesar de avanços pontuais em alguns segmentos, a equidade de gênero no Judiciário brasileiro no ano de 2025 continua sendo um desafio, evidenciando a necessidade de políticas institucionais voltadas à promoção da igualdade e à superação da sub-representação das mulheres na magistratura.

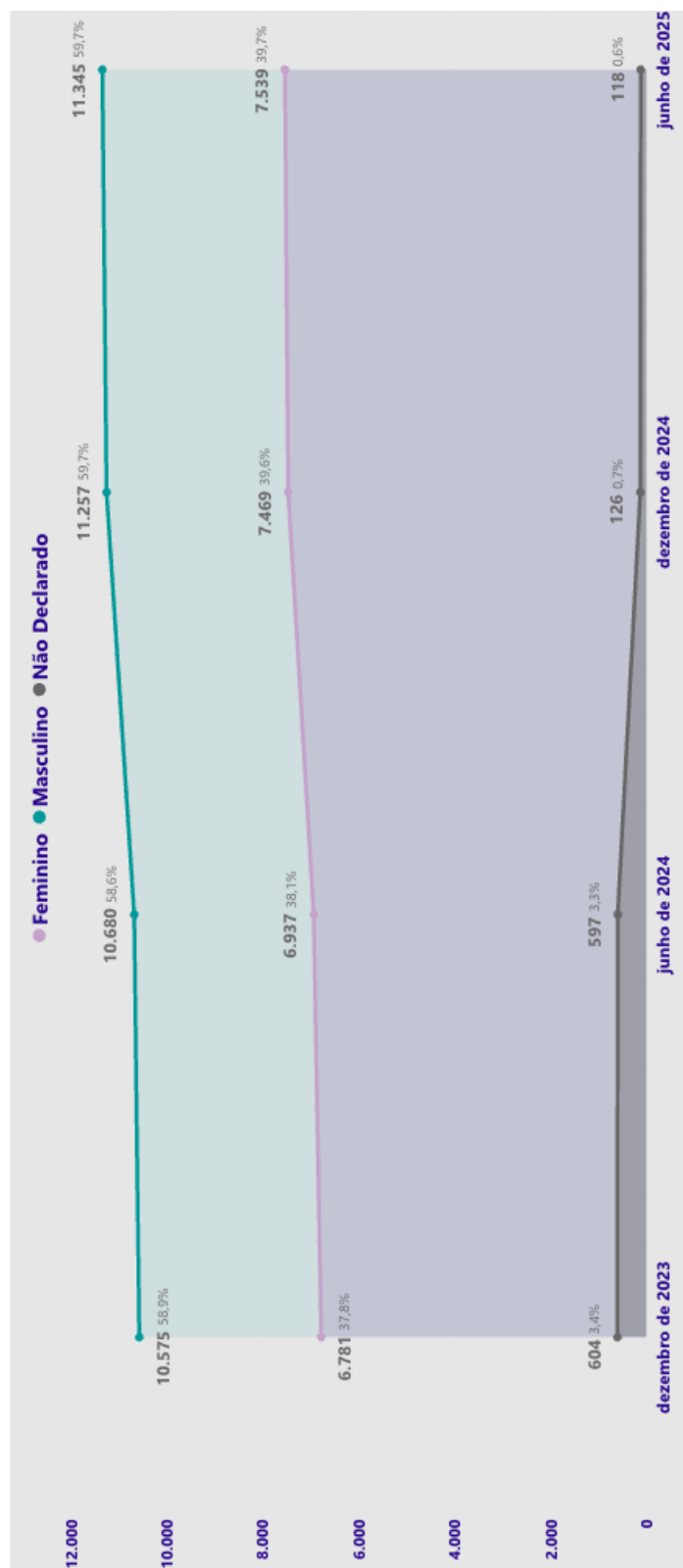
Figura 16. Distribuição de magistrados por gênero e por ramo de atuação em outubro de 2025.



Fonte: Módulo de pessoal e estrutura judiciária mensal do Poder Judiciário, 2025.

Ainda, a Figura 17 permite certificar que, ao menos desde dezembro de 2023, as mulheres estão sub-representadas no Poder Judiciário brasileiro. A maior porcentagem de mulheres, segundo a série histórica disponível pelo “Módulo de pessoal e estrutura judiciária mensal do Poder Judiciário”, é de 39,7%, em junho de 2025.

Figura 17. Série histórica de magistradas no Poder Judiciário brasileiro, entre dezembro de 2023 e junho de 2025.



Fonte: Módulo de pessoal e estrutura judiciária mensal do Poder Judiciário, 2025.

Em resumo, os dados e estudos apresentados evidenciam que o perfil sociodemográfico do Poder Judiciário brasileiro é predominantemente masculino. Apesar de avanços na participação feminina ao longo dos anos, as mulheres ainda representam uma minoria significativa, especialmente em cargos mais elevados, como desembargadoras e ministras (Lages; Cação, 2020). Em 2018, apenas 38% da magistratura era composta por mulheres, com percentuais ainda menores em posições de maior hierarquia. Além disso, a representação de grupos LGBTQIA+ e de pessoas não brancas é ainda mais limitada, reforçando a predominância de um perfil masculino, cisgênero e branco. Mesmo em 2024 e 2025, o cenário de sub-representação de mulheres no Judiciário se manteve.

A criação de políticas e resoluções, como a Resolução CNJ n. 255/2018 e outras iniciativas recentes buscam promover a equidade de gênero e a diversidade no Judiciário (seja internamente, na sua composição, seja através do seu impacto social através das decisões judiciais). No entanto, os dados demonstram que a mudança é lenta e a hegemonia masculina persiste (Conselho Nacional de Justiça, 2023a). Portanto, embora haja esforços para transformar esse cenário, o Poder Judiciário brasileiro ainda reflete um perfil sociodemográfico majoritariamente masculino, com desafios significativos a serem superados para alcançar uma representação mais equitativa e inclusiva.

A masculinização do Poder Judiciário, marcada pela predominância de homens brancos e cisgêneros, em cargos de liderança e decisão, pode ter impactos profundos e multifacetados na sociedade. Esses efeitos podem ser observados em várias áreas, como a aplicação da justiça e a percepção pública sobre a legitimidade do sistema judiciário.

Quanto ao viés de gênero na tomada de decisões judiciais, a falta de diversidade de gênero pode levar a decisões que não consideram completamente as experiências e as necessidades de mulheres e outros grupos marginalizados, especialmente em casos de violência doméstica, assédio sexual e direitos reprodutivos, onde é necessário um olhar mais sensível e atento. Além disso, magistrados homens podem interpretar as leis de forma a perpetuar estereótipos de gênero e desigualdades, reforçando estruturas patriarcais.

Cria-se também uma barreira ao acesso à justiça, na perspectiva de que mulheres e minorias de gênero podem se sentir menos representadas e compreendidas pelo sistema judiciário, gerando desconfiança e relutância em buscar justiça. Além disso, vítimas de violência de gênero podem enfrentar dificuldades adicionais ao procurar reparação, especialmente quando os magistrados não têm a formação necessária para lidar com essas questões de maneira sensível e adequada.

A masculinização do Judiciário também contribui para o reforço de desigualdades estruturais, pois a predominância masculina pode perpetuar normas sociais e políticas públicas que não priorizam questões de igualdade de gênero e direitos das mulheres.

É sabido que o Judiciário tem um papel crucial na definição de normas sociais, e um sistema mais equilibrado em termos de gênero pode contribuir para a promoção de normas mais igualitárias e inclusivas. Além disso, a equidade de gênero no Judiciário é essencial para a justiça social, pois decisões que consideram as experiências e necessidades de todos os grupos sociais são fundamentais para a construção de uma sociedade mais justa e equitativa.

Ainda, decisões judiciais podem impactar diretamente políticas públicas e um Judiciário com pouca diversidade pode ser menos inclinado a defender políticas voltadas para a equidade de gênero e a inclusão social.

Além disso, a falta de representação de gênero no Judiciário pode afetar a legitimidade percebida do sistema perante a sociedade. Quando o Judiciário não reflete a diversidade da população, ele pode ser visto como menos legítimo e justo. A presença de mulheres em posições de liderança no sistema judiciário serve como um modelo e inspiração para outras mulheres e meninas, incentivando-as a seguir carreiras jurídicas e de liderança. A ausência dessa representação, por outro lado, pode ter o efeito contrário.

Mesmo no que diz respeito ao ambiente de trabalho do próprio Judiciário, a predominância masculina pode ser menos acolhedora para mulheres e minorias de gênero, podendo resultar em questões como assédio, discriminação e falta de oportunidades de avanço. A diversidade de gênero, por outro lado, traz diferentes perspectivas e experiências que podem enriquecer a tomada de decisões e a inovação dentro do sistema, enquanto a falta de diversidade pode limitar a capacidade do Judiciário de responder eficazmente às necessidades de uma sociedade plural.

Promover a diversidade de gênero e a inclusão no Judiciário é uma questão de equidade, eficácia e justiça social. Iniciativas para aumentar a representação feminina e de outros grupos marginalizados são fundamentais para garantir que o sistema judiciário reflita e atenda às necessidades de toda a sociedade.

3.2 O mito da neutralidade do Direito e o impacto da desigualdade de gênero no Judiciário brasileiro: o viés humano-jurídico

A atuação judicial deve estar comprometida com a efetivação da justiça, tal como prevista na Constituição Federal (Brasil, 1988). No entanto, a concretização desse compromisso envolve questões complexas e multifacetadas, como a (a) neutralidade e imparcialidade do juiz, (b) a concepção de direito subjacente às decisões judiciais e (c) a dimensão ideológica do direito e da própria jurisdição (Clève, 1993, p. 301).

Existe uma ideia equivocada e recorrente sobre a imparcialidade e neutralidade dos juízes. Como seres humanos, estão sujeitos às influências do ambiente social, cultural e institucional em que vivem. Além disso, a visão do juiz como aplicador passivo da lei, isento de vontade própria, revela-se anacrônica. O juiz não é um instrumento neutro, silencioso e inerte, que apenas resolve casos aplicando a vontade previamente expressa pelo legislador. Ele desempenha um papel ativo na criação e constante reformulação da ordem jurídica. Portanto, possui vontade própria e não pode ser compreendido como um autômato subjugado por técnicas jurídicas ilusórias (Clève, 1993, p. 301).

O juiz não é neutro por natureza, mas deve se esforçar para atingir uma postura de neutralidade e imparcialidade ao tomar uma decisão. Isso exige um controle consciente dos mecanismos psicológicos, sociais e institucionais que influenciam seu processo decisório. Ele deve evitar preconceitos e automatismos que comprometam sua capacidade de analisar profundamente a complexidade dos casos concretos. Tal processo é reconhecidamente desafiador, mas é fundamental que o magistrado o pratique continuamente. A racionalidade deve ser o norte das decisões judiciais, e o juiz deve se esforçar para se libertar de processos mentais irracionais e buscar a razão, como propuseram os pensadores iluministas. A dinâmica processual e o contraditório são espaços que, assim como o método científico, visam revelar a verdade por meio do confronto racional de ideias (Clève, 1993, p. 302).

No entanto, técnicas aparentemente neutras, como subsunção, ponderação e integração, podem disfarçar a presença de valores culturais e ideológicos dominantes, reforçando assim as estruturas de poder e desigualdade (Rezende, 2018, p. 83-84). É possível que um juiz atue politicamente, ainda que inconscientemente, projetando valores não previstos na Constituição (Brasil, 1988) em suas decisões. Isso pode ocorrer por meio de interpretações superficiais ou distorcidas da lei. Assim, para o magistrado exercer a jurisdição, é essencial que ele adote uma postura intersubjetiva e não individualista, orientada pelos valores

constitucionais. Essa perspectiva deve ser acompanhada de garantias institucionais e estruturais que promovam o controle crítico do ato decisório e a responsabilização do Judiciário perante a sociedade (Rezende, 2018, p. 84-85).

Essa necessidade de vigilância crítica sobre o ato de julgar é confirmada por diversas pesquisas científicas voltadas à verificação empírica da incidência de vieses nas decisões judiciais (Gazal-Ayal; Sulitzeanu-Kenan, 2010; Devlin, 1995; Choi; Gulati, 2008). Esses estudos indicam que juízes, ainda que altamente qualificados, estão sujeitos a desvios cognitivos sistemáticos. Como afirmam Guthrie et al. (2001, p. 779), “evidências empíricas sugerem que mesmo juízes altamente qualificados inevitavelmente confiam em processos de tomada de decisão cognitiva que podem produzir erros sistemáticos no julgamento.”

Estudos demonstram que magistrados tendem a ser mecânicos na maioria das ocasiões, atuando de maneira sistemática na classificação de processos e na elaboração de suas decisões, com a finalidade de facilitar seu trabalho e dar eficiência à atividade jurisdicional (Tavares, 2021, p. 109). Fundamentos previamente elaborados costumam ser replicados no processo decisório de casos que se assemelham. Isso é natural.

No mesmo contexto, pesquisas empíricas revelam a sua sujeição às tendências cognitivas a utilizar atalhos para lidar com a pressão da incerteza e da limitação temporal inerentes ao processo judicial (Danziger et al., 2011; Boeing, 2019). É observado que, apesar da sua experiência e treinamento sólido, sua susceptibilidade a uma ilusão cognitiva durante a tomada de decisão solitária influencia seu comportamento. Um exemplo simples encontrado nos estudos, que elucida essa situação, é a inclinação do juiz que rejeita um pedido liminar e posteriormente julga improcedente o pleito. Devido a um efeito de bloqueio, constata-se que o juiz se torna menos inclinado a revisar sua decisão mesmo diante de novos dados ou após mais tempo para reflexão.

Esse bloqueio cognitivo surge devido à tendência de justificar a alocação inicial de recursos (evitando retrabalho), corroborando a correção da decisão inicial. Essa constatação deve estimular a promoção do debate como meio de desmistificar ilusões e tendências cognitivas. Nesse sentido, no contexto brasileiro, poder-se-ia ampliar a discussão para o uso de ementas de decisões judiciais e súmulas como facilitadores de julgamentos, visando exclusivamente otimizar quantitativamente o número de decisões. O emprego de súmulas e “precedentes” sem a devida análise dos casos paradigmas, baseando-se somente em ementas ou no breve texto das súmulas, sugere que tais documentos possam ter valor independentemente

do contexto que os gerou, confundindo a *ratio decidendi* (razão determinante) com algum trecho da ementa ou do voto.

Como explicam Boeing (2019, p. 60) e Rosa (2019, p. 762-763), o processo penal é um bom exemplo de como situações recorrentes (como os crimes mais comuns, furto, roubo, tráfico de drogas etc.) resultam em uma tendência à formação de padrões que geram convicções antecipadas por juízes, levando à negligência de outros detalhes, que acabam sendo considerados irrelevantes. Magistrados, em casos que se assemelham, utilizam mecanismos de conforto cognitivo, tais como “modelos de decisões”, súmulas, brocardos, a fim de buscar coerência em suas decisões.

Frank (2017) enfatiza que os juízes não são seres ideais e infalíveis. Elementos psicológicos, políticos e sociais coexistem com a racionalidade jurídica na formação da decisão judicial. Sobre os fatores que influenciam a tomada de decisão judicial, Frank (2017, p. 156, tradução nossa²⁷) explica que elementos psicológicos, políticos e sociais complementam a aplicação racional para as motivações jurídicas:

Não é absurdo manter viva a tradição ortodoxa artificial do “juiz ideal?” A alternativa racional é reconhecer que os juízes são seres humanos falíveis. Precisamos ver que preconceitos e quais condições de atenção afetam o raciocínio do juiz como o fazem o raciocínio dos homens comuns.

Preconceitos inconscientes, preferências pessoais e idiosincrasias afetam a percepção do juiz, muitas vezes de forma invisível. Reações a aspectos banais, como a voz, a aparência ou o comportamento das partes, podem influenciar de forma irracional o julgamento. É preciso reconhecer essa realidade para que se criem mecanismos de controle e revisão. Frank (2017, 2017, p. XXV, tradução nossa²⁸) aponta que na atividade decisória, há elementos não racionais ou idealistas que podem não ser facilmente identificados:

Esses preconceitos, quando são raciais, religiosos, políticos ou econômicos, às vezes podem ser percebidos por outras pessoas. Mas há alguns vieses ocultos e inconscientes de juízes ou jurados – como, por exemplo, reações positivas ou negativas a mulheres, mulheres solteiras, mulheres ruivas, morenas, homens com vozes profundas ou agudas, homens inquietos, homens que usam óculos grossos ou que têm gestos marcantes ou tiques nervosos – preconceitos dos quais ninguém pode ter consciência. Ocultos e altamente idiosincráticos, tais vieses – peculiares a cada

²⁷ No original: “Is it not absurd to keep alive the artificial, orthodox tradition of the “ideal judge?” The rational alternative is to recognize that judges are fallible human beings. We need to see that biases and prejudices and conditions of attention affect the judge's reasoning as they do the reasoning of ordinary men”.

²⁸ No original: “Those prejudices, when they are racial, religious, political, or economic, may sometimes be surmised by others. But there are some hidden, unconscious biases of trial judges or jurors – such as, for example, plus or minus reactions to women, or unmarried women, or red-haired women, or brunettes, or men with deep voices or highpitched voices, or fidgety men, or men who wear thick eyeglasses, or those who have pronounced gestures or nervous tics – biases of which no one can be aware. Concealed and highly idiosyncratic, such biases – peculiar to each individual judge or juror – cannot be formulated as uniformities or squeezed into regularized “behavior patterns.” In that respect, neither judges nor jurors are standardized”.

juiz ou jurado – não podem ser formulados como regularidades ou encaixados em 'padrões de comportamento' padronizados. Nesse sentido, nem juízes nem jurados são padronizáveis.

Fonseca (2022, p. 115) acrescenta que os magistrados frequentemente recorrem à própria memória, replicando fundamentos utilizados em casos anteriores. Isso os torna suscetíveis a vieses cognitivos, especialmente ao viés de confirmação (Nickerson, 1998), em que se selecionam apenas as informações que corroboram crenças pré-existentes, rejeitando evidências contrárias. Com isso, as decisões deixam de ser plenamente racionais e passam a refletir convicções subjetivas.

Ainda, ideias preconcebidas e crenças influenciam o processo decisório de todo ser humano. É através do viés da confirmação que, por exemplo, um indivíduo acaba selecionando, dentre todas as informações disponíveis, somente aquelas que embasam ou confirmam suas preconcepções, esquivando-se daquelas que são contrárias a isso (Nickerson, 1998). Assim, as decisões perdem seu fundamento racional, seguindo aspectos psicológicos, emocionais, sociais ou culturais (Fonseca, 2022, p. 116-117). Como explica Fonseca (2022, p. 118), o viés da confirmação pode ser entendido como a busca por evidências que suportem uma crença pessoal porque se acredita estar certa:

O viés de confirmação, por dar ênfase excessiva às crenças do tomador de decisões, faz com que ele menospreze informações importantes que levem a evidências contrárias às suas posições, prejudicando, conseqüentemente, a sua própria decisão. Assim, o viés de confirmação refere-se a um tipo de percepção seletiva que enfatiza ideias que confirmam as crenças das pessoas, enquanto desvalorizam o que quer que contradiga essas crenças.

Posner (2011), em sua obra *How Judges Think*, empreende uma análise aprofundada sobre o comportamento judicial, buscando compreender os fatores que influenciam a tomada de decisão dos magistrados. A partir de uma sistematização teórica, o autor elenca nove modelos explicativos do comportamento judicial, chegando à provocativa conclusão de que todos os juízes, em maior ou menor grau, agem politicamente. Essa afirmação não se refere necessariamente a partidarismos explícitos, mas sim à inevitável influência de valores, crenças e estratégias que moldam como os juízes interpretam e aplicam o direito.

A teoria atitudinal, por exemplo, sustenta que a forma mais eficaz de compreender as decisões judiciais está nas preferências políticas dos magistrados, que acabam sendo projetadas em seus julgados. Já a teoria estratégica destaca que os juízes, cientes dos efeitos institucionais de suas decisões, agem de maneira calculada, considerando como suas sentenças serão recebidas por outros magistrados ou instâncias superiores. A teoria sociológica, por sua vez,

parte do pressuposto de que a composição do colegiado influencia diretamente o resultado dos julgamentos, revelando como gênero, classe e origem social interferem na forma de decidir. Mulheres e homens, por exemplo, tendem a julgar certos tipos de casos de maneira distinta (Posner, 2011).

A teoria econômica pressupõe um juiz hiper-racional e autointeressado, que toma decisões maximizando sua utilidade pessoal ou institucional. A teoria psicológica entende o comportamento judicial como fruto de impulsos inconscientes e influências sociais, enquanto a teoria organizacional sustenta que a estrutura do sistema judicial condiciona a atuação dos magistrados, impondo, por exemplo, a adesão a precedentes dos tribunais superiores. A teoria pragmática, por outro lado, entende que os juízes decidem com base nas consequências práticas de suas decisões, afastando-se da ideia de um raciocínio puramente formalista. A teoria fenomenológica vincula a decisão à identidade do próprio magistrado, à sua imagem de si, e a teoria legalista compreende o juiz como mero aplicador da lei ao caso concreto, sem espaço para subjetividade (Posner, 2011).

Fatores extrajurídicos, como pressões institucionais, experiências pessoais e condições emocionais do julgador, exercem influência significativa no resultado do julgamento (Boeing, 2019; Rosa, 2019; Nunes et al., 2018; Wojciechoski; Paola, 2018). A utilização de súmulas, precedentes e ementas como base para decisões rápidas e mecânicas contribui para o distanciamento entre o caso concreto e a razão determinante (*ratio decidendi*) das decisões anteriores. Assim, a compreensão das limitações humanas no julgamento deve ser acompanhada da busca por mecanismos de formação, diversificação, controle e inovação tecnológica, com vistas à promoção de uma justiça mais responsiva, reflexiva e alinhada com os valores democráticos.

Como destaca Rezende (2018, p. 58), o Poder Judiciário sempre esteve imerso nas relações políticas que definem sua atuação cotidiana na aplicação do direito. Formado por agentes institucionais de poder (Gargarella, 1997, p. 27), a tentativa de dissociar o Judiciário da esfera política é uma visão ideológica e, em grande parte, utópica.

Em um contexto social cada vez mais marcado por interações políticas complexas, o Judiciário não é um poder neutro, mas um poder político, cujas funções não podem ser despolitizadas. O Judiciário é político porque integra o aparato estatal, o qual é um produto das relações políticas da sociedade. Sua função primordial – a aplicação do Direito – também carrega uma carga política, além de ser técnica. Assim, não é possível dissociar a atividade jurisdicional da política, uma vez que a aplicação do direito também envolve decisões que

refletem questões de poder e de relações sociais (Rezende, 2018, p. 61; Zaffaroni, 1994, p. 86; Dallari, 2010, p. 89).

Essa dimensão política do Judiciário é crucial, pois a reflexão sobre como ele exerce seu poder se desdobra em questões fundamentais: “como, por que, por quem e para quem” ele é exercido (Rezende, 2018, p. 62). Em um contexto mais amplo, o ideal de uma jurisdição objetiva, neutra e imparcial está superado. A ideia de uma aplicação do direito completamente desvinculada de valores é impossível de ser sustentada, não somente em termos filosóficos, mas também no plano prático, pois as normas jurídicas não são aplicadas de forma mecânica, como propunha o positivismo jurídico. O conceito de avaloratividade, presente no normativismo, já não encontra mais sustentação, pois o direito não pode ser compreendido apenas como uma aplicação de normas neutras e imutáveis (Rezende, 2018, p. 63-64).

Ao considerar a Teoria Pura do Direito de Kelsen (1998), é possível perceber que, embora sua obra seja frequentemente associada ao normativismo, ela não se reduz a um simples legalismo, no qual a aplicação da norma se limita à literalidade do texto legal. A teoria kelseniana, longe de ser uma postura puramente hermenêutica, possui implicações interpretativas que envolvem a compreensão das normas à luz de um processo de interpretação ativo e dinâmico.

Assim, a norma, no pensamento de Kelsen (1998), não é algo fixo e imutável, mas requer um processo de interpretação, pois a prescrição deontológica da lei é, em muitos aspectos, “materialmente vazia” (Rezende, 2018, p. 64-65). Esse vazio normativo só se preenche quando o juiz aplica a norma, e essa aplicação é um ato interpretativo, não uma simples execução de um comando pré-existente. Como explicam Cademartori e Gomes (2008, p. 101), a verdadeira função do juiz é interpretar a norma de acordo com o contexto e as circunstâncias do caso concreto.

Na teoria kelseniana, também se destaca a distinção entre interpretação autêntica e não autêntica da norma. A interpretação autêntica é aquela dada pelas autoridades jurídicas competentes, como o legislador e o juiz. O legislador cria normas gerais, enquanto o juiz aplica normas individuais. A validade dessa interpretação autêntica vem do próprio ordenamento jurídico, que confere legitimidade ao ato interpretativo dessas autoridades (Rezende, 2018, p. 65-66). Isso significa que a autoridade do juiz para interpretar e aplicar o direito decorre do fato de que ele foi legitimado previamente pelo sistema jurídico. Em outras palavras, o juiz só pode decidir o que é o direito porque, antes de sua atuação, foi designado como juiz por uma norma jurídica que lhe conferiu esse poder (Matos; Milão, 2013, p. 130). Portanto, o decisionismo

judicial, isto é, a liberdade do juiz para decidir um caso concreto, está intrinsecamente vinculado à sua legitimação normativa e à interpretação das normas pelo ordenamento jurídico.

Como argumenta Rezende (2018, p. 66), o sistema jurídico não pode normatizar de forma autêntica um caso concreto sem o processo de interpretação. Isso se deve ao fato de que o direito é, por natureza, um sistema aberto, que permite o apelo a padrões normativos externos ao próprio ordenamento jurídico. A norma jurídica, portanto, não é um conjunto de regras rígidas e fechadas, mas um conjunto de diretrizes que exigem interpretação. O direito é, portanto, um campo que permite a flexibilidade interpretativa, pois a interpretação do direito envolve elementos externos ao sistema jurídico, como as condições sociais, políticas e culturais, que influenciam a aplicação da norma.

Essa indeterminação do direito pode ocorrer de duas formas: de maneira intencional, quando a norma traz expressamente múltiplas opções de aplicação à escolha do julgador, ou de maneira não intencional, quando o texto normativo não é unívoco e permite diversas interpretações. Nesse segundo caso, a discricionariedade do juiz torna-se um fator decisivo, pois ele deve escolher entre as diferentes possíveis interpretações de uma mesma norma. Como expõe Rezende (2018, p. 66-67), a teoria kelseniana não exclui a possibilidade de o intérprete exercer uma certa liberdade na interpretação da norma, especialmente no caso de normas abertas, que exigem uma escolha interpretativa.

Além disso, a teoria de Kelsen (1998), ao ser vista também como uma forma de voluntarismo, admite que a escolha entre as várias interpretações possíveis recai sobre a vontade do juiz, que pode decidir fora da moldura normativa, desde que sua decisão seja justificada nos limites da norma (Rezende, 2018, p. 68-70; Appio, 2008, p. 23). A moldura normativa, longe de ser rígida, pode ser ampliada pelos atos de vontade daqueles que têm a autoridade para decidir, ou seja, pelos órgãos do Poder Judiciário. Esse entendimento sugere que, na prática, qualquer decisão tomada pelo juiz, dentro dos parâmetros legais, pode ser considerada válida, o que destaca a flexibilidade e a dinâmica da interpretação jurídica (Rezende, 2018, p. 69).

Essas diversas teorias, ainda que não excludentes entre si, convergem na constatação de que o ato de julgar é condicionado por uma multiplicidade de fatores e que não há neutralidade plena na decisão judicial. Juízes diferentes podem decidir de formas distintas casos idênticos, mesmo submetidos ao mesmo ordenamento jurídico.

Como se vê, uma gama de autores afirma que juízes não são neutros e isentos de ideologias, não se desvencilham de suas concepções pessoais no processo de decidir, embutindo nas decisões judiciais elementos ideológicos que lhes são próprios. Isso decorre de uma

impossibilidade antropológica de neutralidade. Não há neutralidade do Judiciário porque não há homem ou mulher neutro. Julgamentos necessariamente se associam à cosmovisão de quem os profere (Rezende, 2018, p. 78-79).

A “ideologia é um elemento essencial no processo de interpretação” (Rezende, 2018, p. 78). Como traz Rezende (2018, p. 80), o ideal é que juízes tenham uma ideologia subjetiva segundo a constitucional e que esta ideologia sirva de parâmetro ético para a efetivação responsável da Constituição (Brasil, 1988). As suas decisões deverão conferir efetividade ao texto constitucional, independentemente de sua inevitável cosmovisão particular.

Juízes possuem convicções, valores e ideologias que, embora não devam influenciar diretamente suas decisões, inevitavelmente fazem parte de sua interpretação da norma. Nesse sentido, a Constituição brasileira desempenha um papel fundamental, ao servir como um documento que reflete os acordos sociais essenciais e estabelece os limites para a interpretação das normas jurídicas (Brasil, 1988). Ela não somente orienta a aplicação do direito, mas também serve de referencial para a integração do ordenamento jurídico, funcionando como um parâmetro que delimita os contornos da atuação do Judiciário (Rezende, 2018, p. 72).

Tal como defende Streck (2010), “não pode haver uma decisão judicial que não seja fundamentada e justificada em um todo coerente de princípios que repercutam a história institucional do Direito”, existindo um “princípio fundamental do cidadão em receber a resposta correta do Judiciário segundo esse todo coerente de princípios” (Rezende, 2018, p. 73). As respostas judiciais devem concretizar a Constituição e garantir valores emancipatórios (Brasil, 1988; Rezende, 2018, p. 73).

É possível que magistrados decidam conforme a sua subjetividade, deixando de lado valores intersubjetivos comunitários, e mesmo assim sua decisão seja válida (Rezende, 2018, p. 76). Há situações em que a ideologia subjetiva e particular dos juízes pode influenciar diretamente o resultado de seu julgamento (Rezende, 2018, p. 77). A jurisdição deve ser exercida conforme os parâmetros pós-positivistas, isto é, deve ser balizada por valores sociais e constitucionais, engajada na concretização da Constituição, evitando-se ao máximo a substituição da intersubjetividade da comunidade pela subjetividade do juiz (Brasil, 1988; Rezende, 2018, p. 77-78).

Ademais, a fundamentação das decisões judiciais deve buscar obedecer às leis e à Constituição (Brasil, 1988). Mas a fundamentação não transcorre de maneira puramente técnica e despolitizada, pois, como explica Rezende (2018, p. 82), além da “subsunção ao ordenamento

e aos valores intersubjetivos, também estará presente a cosmovisão do magistrado, aparente ou não”.

Os operadores jurídicos, especialmente os juízes, devem assumir a dimensão ideológica e pessoal presente no direito. Mais do que isso, precisam explicitar as motivações e a base ideológica que fundamentam suas decisões, indicando de onde parte o vetor determinante para a solução do caso. Ignorar a existência desses elementos pessoais e ideológicos em uma decisão judicial “impede o controle e a transparência necessárias no Estado Democrático de Direito” (Rezende, 2018, p. 83). Somente com essa clareza será possível alcançar a transparência exigida pelo sistema jurídico, fundamental para a democracia. A negação da ideologia muitas vezes serve como um mecanismo para ocultar o viés ideológico das decisões, o que dificulta ou até impede o exercício do contraditório e da ampla defesa (Clève, 1993, p. 303).

A validade da atuação jurisdicional se baseia na racionalidade e na equidade da decisão. A decisão judicial precisa ser racional e, portanto, passível de controle racional. Por isso, é fundamental que o juiz, sempre que possível, busque exercer um controle sobre os fatores psicológicos que influenciam sua decisão. Além disso, a decisão deve ser justa. O critério de justiça não está em um referencial arbitrário ou em uma concepção filosófica previamente estabelecida, mas sim na própria Constituição, especialmente ao se tratar de uma constituição democrática como a brasileira (Brasil, 1988; Clève, 1993, p. 299-300).

A formação de um sistema judiciário democrático deve necessariamente considerar a dimensão política inerente à jurisdição, evitando idealizações. É fundamental reconhecer que a atuação judicial carrega uma inevitável possibilidade antidemocrática, algo que não se alinha com o sistema político brasileiro. No entanto, apenas ao aceitar essa realidade é possível avançar para uma avaliação prática. Essa abordagem deve enfrentar diretamente os desafios antidemocráticos presentes nesse Poder e superá-los de forma concreta e material (Rezende, 2018, p. 86-87).

É importante reconhecer que o processo de tomada de decisão judicial está sujeito às mesmas limitações cognitivas que afetam qualquer ser humano. Em um sistema judicial sobrecarregado, como o brasileiro (Conselho Nacional de Justiça, 2024a), é comum que magistrados, diante de limitações de tempo e recursos, adotem atalhos mentais, heurísticas e mecanismos de simplificação para decidir. Esses atalhos, muitas vezes inconscientes, podem comprometer a individualização das decisões e perpetuar padrões discriminatórios.

Contudo, a interpretação judicial, sendo um processo essencialmente dinâmico, não é um ato isolado e fechado, mas sim um componente do sistema jurídico que deve se adaptar às mudanças sociais, políticas e tecnológicas.

O reconhecimento dessas diversas perspectivas, preconceções e vieses que moldam o ato de julgar também reforça a necessidade de se repensar os critérios de composição dos órgãos da magistratura, bem como de repensar padrões decisórios diante da crescente implementação tecnológica no Poder Judiciário.

O impacto no conteúdo das decisões judiciais pode ser compreendido a partir da análise do atual perfil sociodemográfico do Poder Judiciário brasileiro, majoritariamente composto por homens brancos, cisgêneros e heterossexuais, oriundos das classes médias e altas. Tal como retratado no subcapítulo anterior, dados estatísticos evidenciam a sub-representação de mulheres, especialmente em instâncias superiores, bem como a baixa presença de pessoas negras, indígenas e de outras minorias sociais.

Conforme demonstrado pelo “Censo do Poder Judiciário de 2023”, embora as mulheres representem cerca de 40% da magistratura, a esmagadora maioria dessas juízas é branca, enquanto mulheres negras ocupam menos de 2% dos cargos da magistratura nacional (Poder Judiciário; Conselho Nacional de Justiça, 2023). Essa configuração homogênea não apenas compromete a legitimidade democrática da instituição, mas também limita o espectro interpretativo e hermenêutico no qual os julgamentos se fundamentam.

A partir dessa realidade, torna-se imprescindível reconhecer que os julgadores não operam isoladamente dos seus contextos sociais e culturais. Como discutido na presente tese, a hermenêutica jurídica é marcada por uma dimensão subjetiva inescapável. O mito da neutralidade judicial, embora reiteradamente invocado como garantia de imparcialidade, desconsidera os múltiplos fatores que condicionam a tomada de decisão – entre eles, as vivências pessoais, as construções de gênero, os valores internalizados e os repertórios sociais do(a) magistrado(a). Quando essa subjetividade opera em um ambiente institucional predominantemente masculino e estruturado sobre valores patriarcais, as consequências tendem a ser a reprodução – ainda que não intencional – de estereótipos e desigualdades, especialmente em casos que envolvem direitos das mulheres, da população LGBTQIA+ e de grupos minorizados.

Trabalhos científicos já buscaram desvendar casos em que houve decisões judiciais discriminatórias, evidenciando como o direito, ao ser aplicado, pode funcionar como mecanismo de exclusão. Muitos dos trabalhos estão atrelados à Justiça Criminal.

Dentro desse contexto, o artigo “A Soberania Patriarcal: o Sistema de Justiça Criminal no Tratamento da Violência Sexual Contra a Mulher” (Andrade, 2007) propõe uma análise crítica e feminista do sistema de justiça criminal brasileiro sob a perspectiva da criminologia crítica e da criminologia feminista. O estudo examina como o sistema penal, ao invés de proteger as mulheres da violência sexual, reproduz a lógica patriarcal, duplicando as formas de violência institucional e simbólica que afetam as vítimas (Andrade, 2007, p. 52-56).

A autora sustenta que o sistema de justiça criminal funciona como instrumento seletivo de controle social, reproduzindo desigualdades (Andrade, 2007). A partir das categorias patriarcal e capitalismo, Andrade demonstra que o sistema penal é estruturalmente androcentrista, voltado ao controle da conduta masculina e à reprodução de estereótipos femininos de passividade e vitimização (Andrade, 2007, p. 63-65). Para a autora, o controle penal e o controle familiar atuam de forma integrada: a pena pública estatal e a pena privada doméstica expressam o mesmo poder masculino e a mesma lógica de dominação (Andrade, 2007, p. 72-75).

No tratamento da violência sexual contra a mulher, o sistema penal reforça a chamada “lógica da honestidade”, pela qual apenas as “mulheres honestas” são reconhecidas como vítimas legítimas, enquanto mulheres fora do padrão moral patriarcal são desqualificadas (Andrade, 2007, p. 67-70). Essa seletividade, segundo Andrade, constitui uma forma de vitimização secundária, em que o Judiciário julga a moral da vítima em vez da violência sofrida. A autora conclui que a superação da violência sexual exige uma reforma profunda do sistema penal e das estruturas sociais, baseada em inclusão, corresponsabilização e reconhecimento da dor como categoria política e epistemológica.

Em “Discriminação judicial por identidade de gênero: um diagnóstico”, Côrtes (2019) realiza uma pesquisa empírica sobre a discriminação judicial contra pessoas trans no TJSP, especialmente em pedidos de alteração de prenome e sexo no registro civil antes da decisão do STF que reconheceu esse direito sem exigência de cirurgia (Cortês, 2019, p. 103-105).

A autora analisou 111 acórdãos entre os anos de 2010 e 2014, comparando o tratamento dado a pessoas transgêneras e cisgêneras em pedidos semelhantes (Côrtes, 2019, p. 107-108). Os resultados revelaram discriminação direta na aplicação do direito. Quando se trata de pedido de alteração do prenome: apenas 42% das decisões envolvendo pessoas trans foram favoráveis, em comparação a 52% de pessoas cisgêneras. Ainda, identifica-se que “quase na totalidade dos casos é reconhecido para pessoas cisgêneras o direito a não carregar um nome que não identifica seu gênero e leva a constantes situações vexatórias diariamente, direito esse

que não é reconhecido nem à metade das pessoas trans que procuram o TJSP” (91%). A autora também aponta que em 87% das decisões negativas, a ausência de cirurgia de transgenitalização foi usada como fundamento para negar a alteração – mesmo sem base legal na Lei de Registros Públicos (Lei n. 6.015/1973) (Brasil, 1973; Côrtes, 2019, p. 111-116).

A pesquisa mostra que o Judiciário paulista adotava critérios patologizantes e excludentes, exigindo comprovações médicas e laudos de cirurgia, em contraste com o tratamento mais flexível concedido a pessoas cis. A discriminação também se expressava na linguagem dos acórdãos: em 96% dos casos, os magistrados usavam pronomes e designações de acordo com o sexo de nascimento, não com a identidade de gênero (Côrtes, 2019, p. 122).

Por sua vez, Tafarelo (2023) investiga como os discursos judiciais em acórdãos sobre crimes de estupro reproduzem estereótipos de gênero e práticas de violência institucional. Após realizar um estudo empírico de decisões judiciais, os resultados trazidos por ela indicam que o discurso judicial ainda se ancora em valores patriarcais, sustentando uma lógica de “dupla vitimização”, na qual as mulheres são julgadas moralmente durante o processo. Tafarelo demonstra que a violência institucional se manifesta tanto pela negação de credibilidade à palavra da vítima quanto pela reprodução simbólica de papéis de gênero – como a idealização da mulher “honesta” e a suspeita sobre comportamentos considerados desviantes.

Corrêa (2024) em “A fixação do valor mínimo de reparação no estupro de vulnerável contra vítima do gênero feminino” analisa a jurisprudência do STJ sobre a indenização mínima por danos morais em casos de estupro de vulnerável, especialmente quando a vítima é uma menina.

A autora demonstra que há divergência jurisprudencial entre duas decisões paradigmáticas – o Recurso Especial n. 1.675.874/MS (Tema Repetitivo 983/STJ) (Brasil, 2018b) e o Recurso Especial n. 1.986.672/SC (Brasil, 2023b). O primeiro permite a fixação do valor mínimo de indenização em casos de violência doméstica e familiar contra a mulher, com base apenas no pedido expresso da acusação, independentemente de prova específica. Já o segundo impõe três requisitos cumulativos: pedido expresso, indicação do valor pretendido e instrução probatória específica, restringindo, portanto, o acesso à reparação (Corrêa, 2024, p. 22-27). Do estudo, tem-se que o formalismo processual e a exigência de requisitos não previstos na lei, funcionam como mecanismos de exclusão, dificultando o acesso das vítimas à justiça e reproduzindo a hierarquia de gênero (Corrêa, 2024).

Destacando também como estereótipos são empregados em decisões judiciais, em livro divulgado como resultado do projeto colaborativo intitulado “Reescrita de Decisões

Judiciais em Perspectivas Feministas no Brasil” (Severi, 2022), foram reunidos diversos projetos de reinterpretação e reescrita de sentenças e acórdãos brasileiros sob uma ótica feminista e interseccional. As decisões selecionadas para reescrita envolvem casos paradigmáticos que ilustram como o sistema de justiça ainda reproduz estereótipos de gênero e práticas discriminatórias – especialmente contra mulheres, pessoas negras e grupos em situação de vulnerabilidade.

São exemplos de algumas das decisões selecionadas pelo projeto e que demonstram a estereotipação de gênero presente no Poder Judiciário:

- a)** O acórdão do TJSP que reformou a sentença proferida na Ação Civil Pública n. 1001521.57-2017.8.26.0360 (São Paulo, 2017). No caso, em primeira instância, foi determinada laqueadura compulsória de uma mulher negra, pobre e periférica (Janaina Aparecida Quirino). O tribunal reverteu a sentença, mas manteve a linguagem patologizante e omitiu as dimensões racial e de gênero do caso (Fernandes; Bernardes, 2022).
- b)** O acórdão do TJSP, proferido em 2018, retrata o julgamento de um Habeas Corpus impetrado pela Defensoria Pública por autoaborto. No caso, o tribunal negou o pedido da Defensoria Pública e manteve a criminalização da mulher. A decisão reforçou a sacralização da maternidade e o uso de estereótipos morais, desconsiderando os direitos reprodutivos e a dignidade da gestante (Santos et al., 2022).
- c)** O acórdão do TJRS proferido na Apelação Criminal n. 0170636-96.2017.8.21.7000 (Rio Grande do Sul, 2017), sobre homicídio, tortura infantil e maus-tratos, no qual a pena da mãe foi agravada sob o argumento de “omissão dolosa” no cuidado dos filhos (Brustolin et al., 2022).
- d)** A sentença proferida em uma Comarca do Rio Grande do Sul em 2017, envolvendo pedido judicial de aborto em caso de malformação fetal (complexo OEIS), que indeferiu o pedido sob a justificativa de que o Código Penal não prevê hipótese legal para aborto em caso de inviabilidade fetal (Rio Grande do Sul, 2017; Simioni et al., 2022).

Quanto à composição de gênero do Poder Judiciário brasileiro, a pesquisa trazida em “Gênero e comportamento judicial no Supremo Tribunal Federal: os ministros confiam menos

em relatoras mulheres?”, por Gomes et al. (2018), investiga como estereótipos e desigualdades de gênero afetam as relações internas entre ministros e ministras do STF.

Os autores analisaram todas as decisões colegiadas do STF entre 2001 e 2013, examinando se o gênero do relator influencia o comportamento dos demais ministros – em especial, a frequência de divergências e pedidos de vista (Gomes et al., 2018, p. 857). Os resultados mostram que, quando o relator é mulher, há 1,2 vez mais votos divergentes em comparação aos casos relatados por homens (e 1,37 vez mais no plenário) (Gomes et al., 2018, p. 867-871). Já o gênero não afeta significativamente os pedidos de vista. Esses achados indicam que as relatoras enfrentam menor confiança de seus colegas, o que revela a persistência de vieses implícitos e hierarquias de gênero dentro da mais alta corte brasileira (Gomes et al., 2018, p. 871-872).

A partir de uma abordagem interdisciplinar que combina ciência política, economia comportamental e estudos de gênero, o artigo interpreta esses resultados à luz de pesquisas sobre estereótipos e discriminação institucional. Os resultados refletem assimetria de poder simbólico e relacional entre homens e mulheres dentro do STF, em um ambiente institucional que ainda reproduz estereótipos de gênero. O artigo transparece a realidade do Poder Judiciário, qual seja, de que, apesar da retórica da neutralidade e da imparcialidade judicial, a estrutura decisória do Supremo reflete hierarquias de gênero, o que reforça a necessidade de políticas voltadas à ampliação da representação feminina e à redução de vieses nas instâncias decisórias do Judiciário (Gomes et al., 2018).

Nesse sentido, a diversidade de gênero, raça, religião e origem regional deve ser compreendida como um dos pilares para a efetiva democratização da justiça. Como defendem Pereira e Oliveira (2018, p. 898), a pluralidade na composição dos tribunais é condição substancial para a representação democrática e para a incorporação de diferentes visões de mundo nos julgamentos. A noção de representação democrática do Judiciário serve, inclusive, como fundamento para sua conformação às demandas contemporâneas por transparência, sensibilidade social e abertura ao contraditório. Nesse contexto, a ideia de que a jurisdição constitucional desempenha função não somente contramajoritária, mas também representativa ganha destaque (Mendonça; Barroso, 2013).

A representação institucional da diversidade no Poder Judiciário amplia a legitimidade do sistema e impacta concretamente o teor das decisões. A diversidade relevante pela inclusão de novos sujeitos em espaços de poder e pela capacidade de influenciar positivamente os entendimentos adotados por outros magistrados (Pereira; Oliveira, 2018, p. 902). A baixa

diversidade nos quadros da magistratura contribui para a manutenção de padrões decisórios que podem negligenciar as experiências de vida desses grupos e frequentemente podem operar com base em compreensões normativas distorcidas sobre seus direitos e realidades.

A hegemonia de um grupo específico no exercício da jurisdição significa, na prática, uma restrição epistemológica da justiça, enquanto invisibiliza outras formas legítimas de perceber e experienciar o mundo. O que está em jogo, portanto, não é apenas a representatividade formal, mas a possibilidade de construção de uma hermenêutica jurídica plural e sensível às múltiplas desigualdades que atravessam o tecido social. Uma jurisdição verdadeiramente democrática.

Tal constatação ganha ainda mais relevância diante da crescente implementação de tecnologias baseadas em IA no Judiciário. Como discutido, os sistemas de IA, ao serem treinados com dados históricos – muitas vezes extraídos de decisões judiciais enviesadas –, tendem a reproduzir e, em alguns casos, até a intensificar os padrões discriminatórios pré-existentes. Isso significa que a ausência de diversidade e a presença de vieses humanos nos processos decisórios não são apenas um problema do passado ou do presente, mas um fator de risco concreto para o futuro da justiça, especialmente quando essa justiça for parcialmente mediada por algoritmos.

Portanto, repensar os critérios de composição do Judiciário e incorporar ativamente a diversidade é um passo necessário não só para ampliar a legitimidade da instituição, mas também para promover decisões mais justas, sensíveis e inclusivas. A presença de diferentes sujeitos nos espaços de poder judicial é capaz de enriquecer o debate jurídico, tensionar compreensões hegemônicas e criar as condições necessárias para a construção de uma justiça democrática – tanto no plano humano quanto no digital.

Sob outra perspectiva, o avanço da modernização tecnológica do Judiciário brasileiro pode oferecer oportunidade única de aprimorar a interpretação judicial através de técnicas de IA. A tecnologia pode contribuir para tornar o processo de decisão mais eficiente, reduzindo a morosidade e a burocracia, além de possibilitar uma análise mais precisa e abrangente dos casos, considerando uma quantidade maior de dados e precedentes, o que deve ser feito considerando a existência de possíveis vieses nos dados e a necessidade de seu balanceamento.

A implementação de IA na atividade jurisdicional pode ser um instrumento útil para compreender os padrões decisórios dos magistrados, permitindo identificar vieses e automatismos. Por exemplo, imagine-se a criação de instrumentos de recomendação que construam o perfil decisório dos juízes a partir de suas sentenças anteriores, com o objetivo de

revelar preconceitos ocultos e promover maior transparência. Ou então, um sistema que analise decisões judiciais de determinado Tribunal, identifique possíveis vieses e sugira diversas formas de corrigi-lo, tornando a justiça mais equânime.

Vale dizer, a modernização tecnológica não deve ser vista como um substituto para a interpretação humana e para o papel do juiz. Embora a tecnologia possa fornecer ferramentas poderosas para otimizar o processo decisório, ela não pode substituir o juízo crítico e a capacidade de interpretação do juiz, que continua sendo fundamental para garantir a justiça das decisões. A tecnologia deve, portanto, ser vista como um instrumento que auxilia na construção da decisão judicial, ampliando as possibilidades de análise e proporcionando mais agilidade ao processo, mas sempre dentro dos limites impostos pelos princípios constitucionais e pela necessidade de preservar a racionalidade e a justiça.

Em conclusão, a interpretação judicial permanece sendo uma atividade profundamente humana, que exige do juiz não apenas conhecimento técnico, mas também uma compreensão crítica dos princípios e valores constitucionais. A modernização tecnológica do Poder Judiciário brasileiro, ao melhorar a eficiência e a qualidade da prestação jurisdicional, deve ser encarada como uma oportunidade de aprimorar a interpretação judicial, garantindo que ela seja mais rápida, precisa, ética, justa e acessível.

O que se defende aqui é a imprescindibilidade da implementação de novas tecnologias acompanhada de uma reflexão ética constante, para que a racionalidade, a ética e a justiça da decisão judicial não sejam comprometidas, mas sim fortalecidas pela inovação. A harmonia entre a interpretação judicial e a modernização tecnológica permitirá, assim, um Judiciário mais justo, democrático e eficaz, alinhado com as necessidades da sociedade contemporânea.

A busca pela neutralidade não deve ser entendida como uma condição natural ou plenamente alcançável, mas como um exercício contínuo de autorreflexão, de abertura ao contraditório e de compromisso institucional com a democracia e com os direitos fundamentais.

3.3 O Protocolo para Julgamento com Perspectiva de Gênero e o resgate feminista da magistratura no Brasil

Vivemos em uma sociedade onde o déficit de representatividade feminina em cúpulas de poder é evidente (Comissão Ajufe Mulheres, 2011; Bonelli, 2011; ; IBGE – Instituto Brasileiro de Geografia e Estatística, [2013?]; Araújo, 2015; Fragale Filho et al., 2015; Melo; Tomé, 2018; Duarte et al., 2021; Naciones Unidas; CEPAL, [20–]). Como explicam Pereira e

Oliveira (2018, p. 879), os fundamentos das democracias liberais foram estabelecidos em uma época em que a participação das mulheres nos processos deliberativos estatais era praticamente inexistente. Essa realidade histórica continua a influenciar os espaços de poder atualmente, mesmo que a esfera pública seja, muitas vezes, concebida como neutra e imparcial em relação ao gênero. As discussões sobre a estrutura institucional dos órgãos estatais raramente levam em consideração o contexto histórico em que os espaços de tomada de decisão se formaram, o que acaba por perpetuar desigualdades de poder entre homens e mulheres. Corriqueiramente, desconsidera-se que a organização do Estado carrega consigo um legado histórico de divisão desigual de papéis entre os gêneros, situação que se repete também no âmbito privado.

Essa constatação é particularmente visível no Poder Judiciário brasileiro, cuja estrutura institucional permanece fortemente marcada por desigualdades de gênero. A análise do perfil sociodemográfico da magistratura brasileira, desenvolvida no subcapítulo 3.1 desta tese, evidencia que, embora o ingresso de mulheres na carreira da magistratura tenha oscilado nas últimas décadas, elas ainda são significativamente sub-representadas nos cargos mais elevados da hierarquia judicial – como nos tribunais superiores, nas presidências de tribunais e nos órgãos de cúpula do Judiciário.

A partir dos dados analisados nesta tese, é possível observar que o Poder Judiciário brasileiro reproduz essas estruturas desiguais. Trata-se de um campo profissional em que homens brancos e heterossexuais ainda constituem o padrão dominante, o que contribui para a exclusão ou marginalização de mulheres e de outros grupos sub-representados.

Como debate Almeida (2018), a presença feminina em espaços de poder está diretamente relacionada à existência de vieses implícitos e a preconceitos internalizados, que influenciam inclusive o comportamento e as escolhas das próprias mulheres. Segundo a autora, devido à presença de vieses implícitos, observa-se que os critérios utilizados para avaliar homens e mulheres em espaços de poder são frequentemente distintos. Esses preconceitos, muitas vezes internalizados, limitam a atuação feminina e dificultam o engajamento em transformações estruturais que poderiam ampliar sua participação nas esferas decisórias. Muitas mulheres evitam adotar posturas firmes ou assertivas, receosas de serem rotuladas como hostis, conflituosas ou como alguém que busca tratamento diferenciado. Além disso, mulheres enfrentam cobranças mais rigorosas – tanto externas quanto autoimpostas – e lidam com expectativas sociais relacionadas à vida familiar e à maneira como se relacionam com colegas de trabalho. Tais fatores contribuem para sua exclusão de redes de relacionamento e influência profissional, essenciais para a ascensão nas estruturas de poder. Tais preconceitos repercutem

não apenas nas relações institucionais entre magistrados e magistradas, mas também nas interações com servidoras, estagiárias e advogadas, criando um ambiente que muitas vezes é hostil à ascensão e à permanência de mulheres em posições de liderança.

A preocupação do CNJ com a disparidade de gênero no Poder Judiciário brasileiro tem motivado a implementação de políticas e instrumentos normativos voltados à promoção da equidade e à eliminação de práticas discriminatórias. Algumas dessas iniciativas já foram expostas, tal como a Resolução CNJ n. 254/2018, que institui a Política Judiciária Nacional de Enfrentamento à Violência contra as Mulheres, estabelecendo diretrizes para a prevenção e o combate à violência de gênero, bem como para a adequada solução de conflitos que envolvam mulheres em situação de violência física, psicológica, moral, patrimonial e institucional.

Complementando essa política, o CNJ lançou, em 2021, o Protocolo para Julgamento com Perspectiva de Gênero (Conselho Nacional de Justiça, 2021a), desenvolvido em parceria com a Escola Nacional de Formação e Aperfeiçoamento de Magistrados (ENFAM). A importância de tal protocolo é ímpar para o que se defende nesta tese, que visa garantir a equidade nas decisões judiciais, ainda que venham a ser parcial ou totalmente automatizadas.

O protocolo visa capacitar e orientar a magistratura na realização de julgamentos que considerem as desigualdades de gênero, promovendo uma justiça mais equitativa e sensível às questões que afetam mulheres e meninas. A adoção do protocolo busca assegurar o direito à igualdade e à não discriminação. O documento parte do reconhecimento de que o direito, quando aplicado de forma acrítica, pode ser vetor de reprodução de desigualdades, ao invés de instrumento de sua superação. Por isso, propõe que magistradas e magistrados adotem, nos processos sob sua condução, uma lente de gênero capaz de identificar e neutralizar estereótipos, desigualdades estruturais e padrões discriminatórios, muitas vezes naturalizados pela cultura institucional do Judiciário.

A proposta está em consonância com normativas internacionais de direitos humanos – como a Convenção sobre a Eliminação de Todas as Formas de Discriminação contra a Mulher (CEDAW) e decisões da Corte Interamericana de Direitos Humanos – e com o compromisso assumido pelo CNJ e pelo Supremo Tribunal Federal (STF) com os Objetivos de Desenvolvimento Sustentável da Agenda 2030 da ONU (Nações Unidas Brasil, [20-?]), especialmente o ODS 5, que trata da igualdade de gênero.

Além de oferecer conceitos que servem de base teórica para compreender como a desigualdade de gênero se manifesta e se perpetua, inclusive dentro das instituições do sistema de justiça, o protocolo também disponibiliza um roteiro metodológico detalhado para que

magistradas e magistrados incorporem a perspectiva de gênero em todas as etapas da atuação jurisdicional. Esse passo a passo propõe que, desde o primeiro contato com os autos até a elaboração da decisão, a atuação judicial seja conduzida com atenção às desigualdades materiais entre os sujeitos processuais.

O documento tem por objeto auxiliar magistradas e magistrados a analisar o contexto social dos sujeitos processuais, de modo a identificar possíveis vulnerabilidades associadas ao gênero, à raça, à classe social, à orientação sexual ou à identidade de gênero. Ele propõe, inclusive, a rejeição de estereótipos de gênero na valoração de provas, especialmente em casos que envolvam violência sexual, violência doméstica ou comportamentos considerados “inadequados” para mulheres. Ele possibilita e recomenda a avaliação crítica da neutralidade do direito, compreendendo que normas aparentemente neutras podem gerar impactos desproporcionais para grupos vulnerabilizados.

Além disso, o protocolo trata de questões específicas nos diferentes ramos do Judiciário, demonstrando como as desigualdades de gênero se expressam em áreas tão diversas quanto o direito penal, civil, trabalhista, eleitoral, previdenciário e militar. São abordados, por exemplo, temas como feminicídio, violência obstétrica, pornografia de vingança, alienação parental, desigualdade salarial, assédio moral e sexual no ambiente de trabalho, dentre outros. Para cada situação, são oferecidas orientações práticas e exemplos de aplicação da perspectiva de gênero, sempre com base em jurisprudência nacional e internacional.

De tal forma, o próprio documento destaca a sua aplicabilidade em qualquer julgamento, em qualquer área do direito, no qual desigualdades de gênero possam interferir na equidade do processo ou no resultado da decisão. Sem dúvidas, o protocolo propõe uma mudança de paradigma no exercício da jurisdição, promovendo a justiça substantiva em lugar de uma formalidade neutra que, na prática, ignora ou reforça desigualdades históricas.

A existência e a adoção institucional deste protocolo pelo CNJ demonstram que há uma crescente conscientização sobre os impactos do gênero na justiça e a necessidade de desenvolver práticas judiciais mais igualitárias.

Compreender o perfil sociodemográfico do Judiciário é passo fundamental para o enfrentamento do viés de gênero em suas múltiplas dimensões, incluindo aquela que se projeta sobre o desenvolvimento e a aplicação de tecnologias como a IA. A sub-representação feminina nos espaços de poder judicial não apenas limita a diversidade na produção de decisões judiciais humanas, mas também influencia os dados e padrões que alimentam os sistemas de IA. Tais sistemas, se treinados a partir de decisões judiciais marcadas por essas desigualdades

estruturais, arriscam perpetuar – e até acentuar – os mesmos padrões excludentes. Daí a urgência de incorporar uma perspectiva feminista crítica que reconheça e enfrente essas assimetrias, promovendo uma transformação efetiva das estruturas institucionais e a adoção de protocolos éticos e inclusivos para o uso da IA no âmbito judicial.

Defende-se, portanto, que o “Protocolo para Julgamento com Perspectiva de Gênero” (Conselho Nacional de Justiça, 2021a) deve ser minuciosamente observado na implementação de sistemas de IA no Poder Judiciário brasileiro. Tal afirmação se justifica por diversos fundamentos teóricos, institucionais e práticos que convergem para a necessidade de uma abordagem crítica e comprometida com a justiça de gênero também no desenvolvimento e uso de tecnologias.

A sua análise revela a importância de ferramentas normativas que ajudem a desconstruir os filtros androcêntricos que ainda dominam a produção e aplicação do direito e que podem ser potencialmente reproduzidos – e até ampliados – por tecnologias de IA. Ao propor critérios claros, sensíveis e contextualizados para a tomada de decisão judicial, o protocolo reafirma o papel do Poder Judiciário na promoção da justiça de gênero e fornece um modelo institucional que pode – e deve – influenciar também o desenvolvimento ético de soluções tecnológicas para o sistema de justiça.

Os sistemas de IA operam a partir de dados históricos que, no campo judicial, refletem decisões marcadas por estruturas desiguais de gênero, raça, classe e outras opressões interseccionais. Como a análise do perfil da magistratura demonstrou, trata-se de um campo dominado por homens brancos e heterossexuais, cujas decisões formam a base dos bancos de dados utilizados para treinar modelos algorítmicos. Assim, há o risco real de que esses sistemas internalizem e amplifiquem padrões excludentes, naturalizando práticas discriminatórias sob a aparência de imparcialidade e eficiência técnica.

Além disso, a ilusão de neutralidade algorítmica torna-se especialmente perigosa quando aplicada ao campo do direito. O protocolo do CNJ parte da constatação de que a aplicação acrítica da norma pode reproduzir desigualdades – crítica que ganha ainda mais relevância diante da automatização crescente de processos decisórios. A IA tende a mascarar os vieses sob uma pretensa objetividade, dificultando sua identificação e contestação. Dessa forma, é imperativo que o protocolo seja incorporado como diretriz ética e metodológica desde as fases iniciais de concepção, treinamento, implementação e validação de sistemas algorítmicos voltados à atividade jurisdicional.

Outro aspecto central é que o documento oferece uma metodologia interseccional, permitindo que desigualdades múltiplas e sobrepostas sejam reconhecidas e enfrentadas. No campo da IA, a ausência dessa lente crítica pode resultar em decisões automatizadas homogeneizantes e excludentes, que ignoram as vulnerabilidades materiais dos sujeitos processuais. A sua implementação rigorosa é condição para orientar a IA por valores de justiça substantiva, rompendo com a lógica formalista que historicamente oculta as assimetrias estruturais.

O documento expressa um compromisso institucional do Poder Judiciário com a equidade de gênero. Esse compromisso, assumido em consonância com a Constituição Federal (Brasil, 1988) e com tratados internacionais de direitos humanos, deve se estender às ferramentas tecnológicas adotadas pelo Judiciário. Se o exercício da jurisdição está vinculado aos princípios da igualdade e da não discriminação, qualquer sistema que auxilie ou substitua o julgamento humano deve estar igualmente comprometido com esses valores.

O protocolo não se limita a uma recomendação interpretativa, mas constitui um marco ético-jurídico que propõe uma mudança de paradigma na forma de julgar. Sua incorporação na IA não é apenas desejável: é necessária para garantir que a automação judicial não retroceda frente aos avanços civilizatórios construídos na luta por direitos das mulheres. A tecnologia deve servir à justiça – e não o contrário. Integrar as diretrizes do protocolo aos sistemas algorítmicos significa alinhar a inovação tecnológica com os princípios democráticos e com a responsabilidade institucional de promover justiça social.

Sob a lente de uma perspectiva feminista crítica, como propõe esta tese, é possível compreender a IA não apenas como uma ferramenta técnica, mas como um dispositivo político e cultural que pode reforçar – ou desconstruir – os filtros androcêntricos do direito. O protocolo, nesse sentido, constitui um instrumento indispensável para que a IA não se torne um novo mecanismo de exclusão, mas sim uma ferramenta de transformação institucional comprometida com a justiça de gênero.

Diante do risco concreto de reprodução de desigualdades estruturais por meio da IA, a observância rigorosa do “Protocolo para Julgamento com Perspectiva de Gênero” (Conselho Nacional de Justiça, 2021a) deve ser tratada como um requisito normativo e ético no desenvolvimento de soluções tecnológicas voltadas ao Judiciário. A tecnologia, ao contrário do que muitas vezes se propaga, não é neutra – e por isso, deve ser concebida a partir de diretrizes claras que garantam equidade, justiça material e compromisso com os direitos humanos. Incorporar o protocolo nos sistemas de IA significa assegurar que a automação da justiça não

reproduza as falhas do passado, mas contribua para a construção de um futuro mais igualitário e democrático.

CAPÍTULO 4 – A IMPLEMENTAÇÃO DA INTELIGÊNCIA ARTIFICIAL NO PODER JUDICIÁRIO BRASILEIRO

Neste capítulo, explora-se a implementação tecnológica no Poder Judiciário brasileiro e seu papel na garantia de julgamentos justos e feministas na Era Digital, que necessariamente deve observar os princípios constitucionais que asseguram o devido processo legal. Esses princípios, consagrados na Constituição Federal de 1988, garantem que todos os cidadãos tenham direito a um julgamento imparcial, à ampla defesa e ao contraditório, fundamentais para a promoção da justiça (Brasil, 1988). A integração de tecnologias no Judiciário deve, portanto, respeitar e reforçar esses direitos, assegurando que a inovação não comprometa a equidade e a transparência na aplicação da lei, mas sim contribua para um sistema mais inclusivo e justo.

Nesse cenário, à medida que a IA se torna cada vez mais integrada ao sistema judiciário, surgem tanto oportunidades quanto desafios no campo jurídico. A tecnologia promete eficiência e inovação, mas é crucial que sua implementação seja feita de maneira a respeitar princípios de justiça e equidade.

Assim, inicia-se o capítulo com a análise acerca da implementação de técnicas de IA no Poder Judiciário brasileiro. Examina-se o estado da arte desta implementação tecnológica: “o que já está sendo aplicado?”. Após a discussão quanto à etapa em que estamos atualmente, debate-se a respeito do que se pode esperar para o futuro. Essa avaliação permitirá uma compreensão clara das iniciativas em curso e das potencialidades que a IA pode oferecer para aprimorar a justiça, bem como dos riscos associados à sua adoção sem a devida cautela.

4.1 O estado da arte da implementação de sistemas inteligentes no Poder Judiciário brasileiro

A influência da IA permeia o dia a dia de todos os cidadãos, com algoritmos capazes de compreender e até mesmo influenciar as preferências individuais, desde estilo de vestimenta e gostos culinários até posicionamentos políticos, todos critérios subjetivos. Nesse contexto, a implementação tecnológica moderna, em especial, de IA, no campo do Direito é uma realidade que tem ganhado cada vez mais espaço, seja no âmbito dos tribunais que buscam o desenvolvimento de seus próprios sistemas inteligentes, seja no âmbito dos demais órgãos de justiça ou em escritórios de advocacia. A ciência e a tecnologia são vistas como instrumentos para garantir a justiça (Cui, 2020, p. 21-22).

O cenário atual de implementação tecnológica leva a interpretar que a não utilização de sistemas inteligentes é sinônimo de algo atrasado, obsoleto. No Judiciário, seja no Brasil ou mundo afora (Cui, 2020, p. 22-23; D’Almeida, [202?]; CEPEJ-AIAB, 2025), tem-se um terreno fértil para a implementação de técnicas de aprendizagem de máquina devido à riqueza de dados potencialmente tratáveis²⁹, à necessidade de agilidade na prestação da tutela jurisdicional, à disponibilidade de orçamento para o desenvolvimento de soluções inovadoras e à escassez de recursos humanos diante da elevada carga de trabalho (congestionamento de processos judiciais) (Boeing, 2019, p. 68-69).

Cada vez mais, o Poder Judiciário brasileiro busca implementar e promover o uso da IA no desenvolvimento de suas atividades rotineiras com a finalidade de otimizar a prestação de serviços (Cruz et al., 2022; Salomão, 2022). Nesse contexto, o artigo de Cruz et al. (2022) oferece um panorama detalhado sobre o estado da arte da IA no Poder Judiciário brasileiro até julho de 2020. A partir de uma pesquisa empírica e bibliográfica envolvendo os 91 tribunais nacionais, os autores demonstram que a IA tem sido amplamente adotada, sobretudo para atividades operacionais, como triagem de processos, reconhecimento de padrões e sugestões de minutas. Embora ainda incipiente no campo decisório, a IA tem colaborado significativamente na racionalização dos fluxos de trabalho e no enfrentamento do acúmulo processual, com destaque para iniciativas como o sistema Victor no STF, Sócrates no STJ e Bem-te-Vi no TST, entre outros.

Apesar dos avanços, o estudo também aponta os principais desafios da expansão da IA no Judiciário, como a necessidade de dados representativos, o enfrentamento de vieses algorítmicos e, sobretudo, a limitação dos sistemas em lidar com questões jurídico-principiológicas que exigem julgamento contextual, ponderação de valores e bom senso – atributos intrinsecamente humanos. Assim, embora as soluções inteligentes não substituam os magistrados, elas vêm se consolidando como ferramentas auxiliares potentes, capazes de ampliar a eficiência institucional.

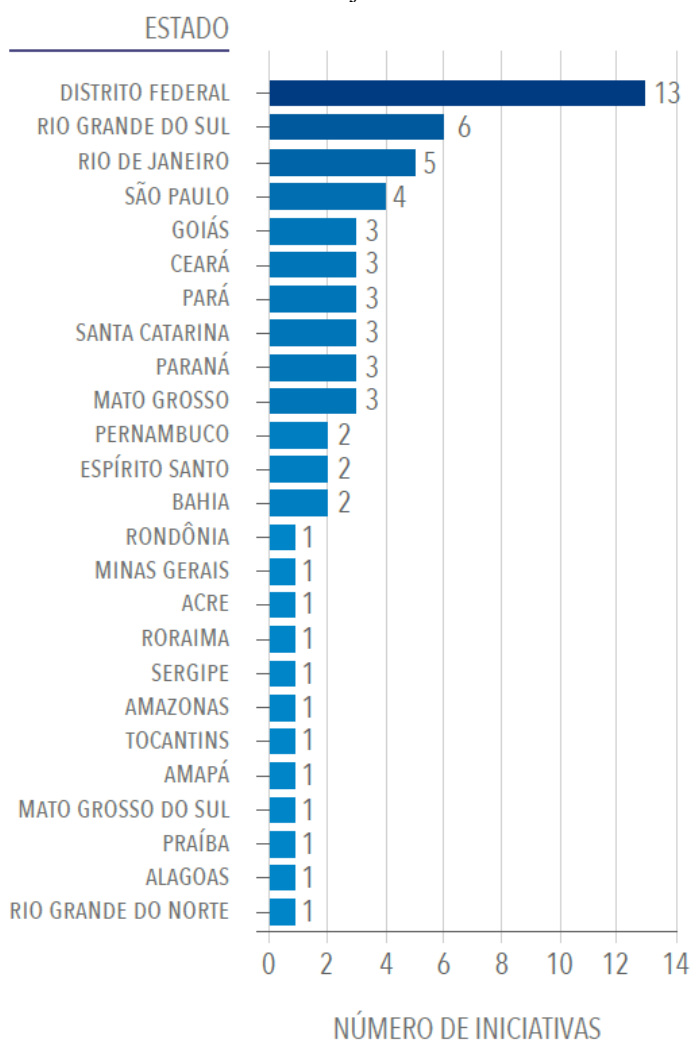
Atualmente, o Poder Judiciário brasileiro conta com uma variedade de sistemas de IA que lidam com uma ampla gama de questões, desde o atendimento inicial aos cidadãos até o gerenciamento de processos em varas e tribunais nacionais e internacionais (Salomão, 2022, p. 53-205). Como destaca Nunes (2023, p. 55), no Brasil, o Judiciário conta com iniciativas destinadas a auxiliar atividades-meio, relacionadas à administração da justiça, gerindo recursos

²⁹ Em relação à quantidade de dados, a atividade jurisdicional gera uma variedade e um volume complexo de dados, estruturados ou não, que tem origem em várias fontes, constituindo um verdadeiro *big data* jurídico (Lima e Oliveira, 2019, p. 70).

financeiros e de pessoas, iniciativas de “modelos de automação de fluxo de movimentação do processo e das atividades executivas de auxílio aos juízes”, modelos de apoio de decisão judicial que elaboram minutas de sentenças, votos e decisões, mas sem promoção de *legal reasoning* e também iniciativas que auxiliam a resolução de conflitos.

O relatório de pesquisa de mapeamento de tribunais brasileiros que usam a inteligência computacional na gestão de processos judiciais, coordenado pelo Min. Salomão (2022) identifica os Estados-membros que possuem os maiores números de iniciativas inovadoras. Na Figura 18, temos o ranking de número de iniciativas por Estado-membro até o ano de 2021 (ano em que foram colhidos os dados para o relatório), abrangendo as que seriam implantadas em 2022:

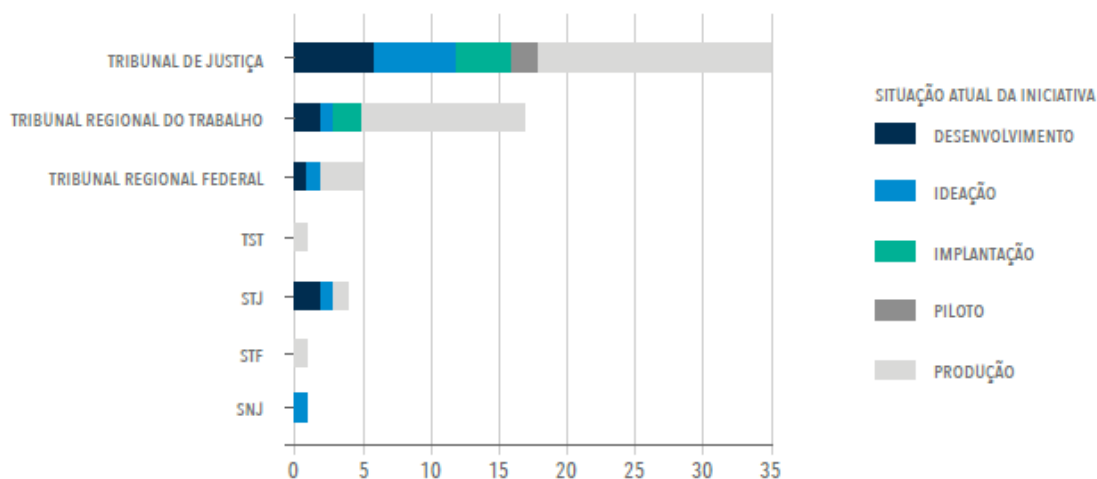
Figura 18. Números de iniciativas de uso de IA no setor jurídico.



Fonte: Salomão, 2022, p. 254.

Para fins de melhor visualização da implementação tecnológica no âmbito do Judiciário brasileiro, a Figura 19 identifica a quantidade de iniciativas por tribunal e por situação de desenvolvimento:

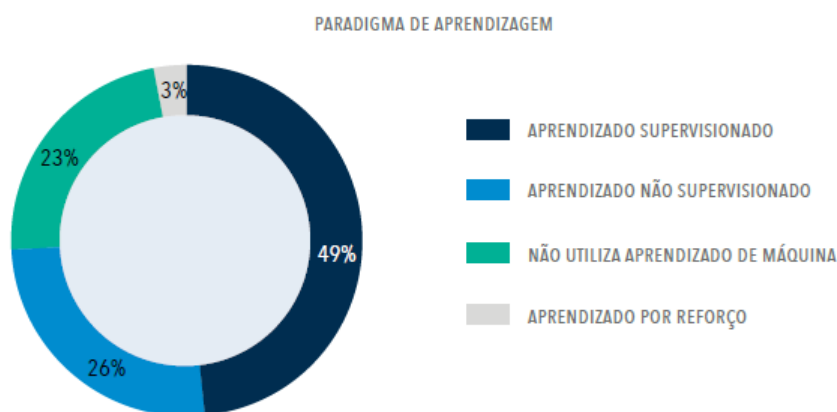
Figura 19. Quantidade de iniciativas de IA por tribunal e situação.



Fonte: Salomão, 2022, p. 255.

Também o relatório realizado por Salomão (2022) ilustra quais são as abordagens de IA mais utilizadas nos tribunais, tal como vemos na Figura 20:

Figura 20. Abordagens de AI nos tribunais - I.

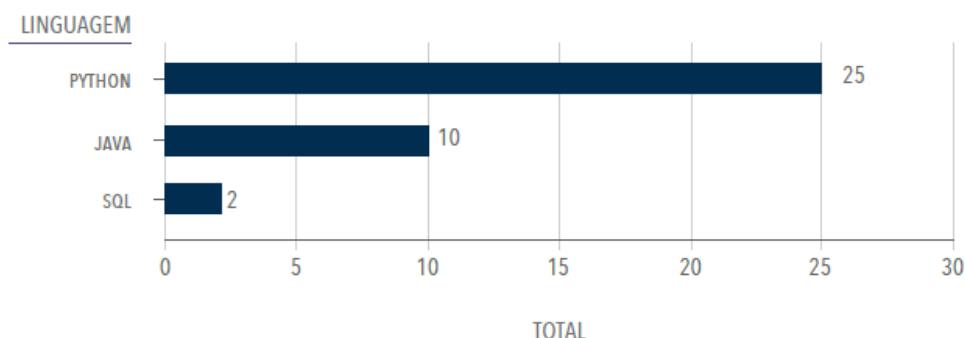


Fonte: Salomão, 2022, p. 256.

As linguagens das abordagens adotadas também são especificadas por Salomão (2022) na Figura 21, indicando a prevalência da linguagem Python. Acerca dos frameworks, identificou-se o Scikit Learn em projetos de classificação e agrupamento de documentos. O Power BI e o Qlik Sense ganharam destaque como plataformas de BI e o Elasticsearch e o Solr

nos projetos de recuperação de informação. Random Forest e XGBoost serviram para classificar documentos e o k-Means para problemas relacionados a agrupamento.

Figura 21. Abordagens de AI nos tribunais - II.



Fonte: Salomão, 2022, p. 263.

Essa integração da IA através de seus algoritmos decisórios com o Poder Judiciário representa uma inovação significativa, porém ainda carente de regulamentação e legislação, tanto em níveis internacionais quanto nacionais. Dentre as muitas questões controversas relacionadas aos mecanismos de IA, destaca-se a necessidade de conciliar segurança jurídica, proteção dos direitos fundamentais e formulação de políticas públicas, respeitando princípios como transparência, independência e inovação.

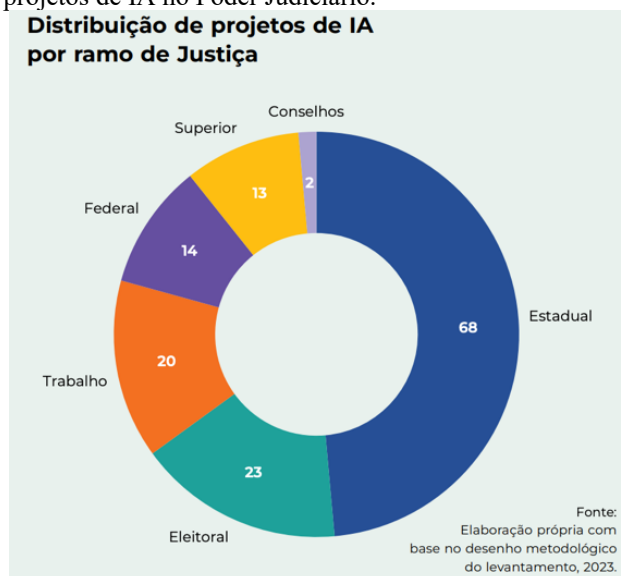
No que diz respeito às técnicas de IA aplicadas no âmbito do Poder Judiciário brasileiro, ganha destaque a aprendizagem de máquina, cujo método envolve a aplicação de algoritmos que melhoram o seu desempenho com a experiência (Flach, 2012). A aprendizagem de máquina no âmbito jurídico possui três principais aspectos: características, tarefas e modelos. As características consistem em documentos (seus termos, construções gramaticais, informações semânticas etc.), metadados e dados estruturados a eles relacionados. As tarefas, que se constituem em uma “representação abstrata de um problema que se queira resolver”, no Direito, correspondem às informações que se deseja extrair automaticamente. Os modelos, por sua vez, dizem respeito ao “resultado da aplicação de um algoritmo de aprendizagem de máquina” (Salomão, 2022, p. 29).

Os dados do relatório acima são complementados pelos da pesquisa intitulada “Uso de inteligência artificial no Poder Judiciário – 2023”, do CNJ (Conselho Nacional de Justiça, 2023f). Realizada por meio do Departamento de Pesquisas Judiciárias (DPJ), teve como objetivo principal compreender o grau de maturidade e as percepções sobre o uso da IA no

Poder Judiciário brasileiro. A iniciativa integrou o “Programa Justiça 4.0” (Conselho Nacional de Justiça, [20-?]) e buscou oferecer subsídios para a formulação de políticas públicas voltadas à transformação digital da justiça. O levantamento foi aplicado entre março e abril de 2023 por meio de um questionário eletrônico, obtendo 175 respostas de 94 órgãos judiciários, incluindo tribunais e conselhos.

A amostra coletada revelou um total de 140 projetos de IA em desenvolvimento, distribuídos de forma desigual entre os ramos da justiça: 68 projetos na Justiça Estadual (48,6%), 23 na Eleitoral (16,4%), 20 na do Trabalho (14,3%), 14 na Federal (10%), 13 nos tribunais superiores (9,3%) e 2 nos conselhos (1,4%). Em termos de estágio de desenvolvimento, 63 projetos estavam finalizados e em produção (45%), 46 em andamento (32,9%), 17 em estágio inicial (12,1%), 11 finalizados, mas ainda não implementados (7,9%) e 3 não iniciados (2,1%). A Figura 22, retirada do relatório, ilustra essa distribuição.

Figura 22. Distribuição de projetos de IA no Poder Judiciário.



Fonte: Conselho Nacional de Justiça, 2023f, p. 28.

Cada ramo de justiça apresentou motivações distintas para o uso da IA: na Justiça Estadual, o foco está na agilidade processual e no enfrentamento do volume de dados; na Justiça do Trabalho, destaca-se a otimização do tempo dos servidores e o uso de *feedbacks*; na Justiça Federal, o uso da IA visa a triagem de processos e a produção de minutas; na Justiça Eleitoral, a prioridade é melhorar o atendimento ao eleitor e a confiabilidade dos sistemas; já os tribunais superiores utilizam a IA para acelerar julgamentos e identificar temas de repercussão geral (Conselho Nacional de Justiça, 2023f, p. 47).

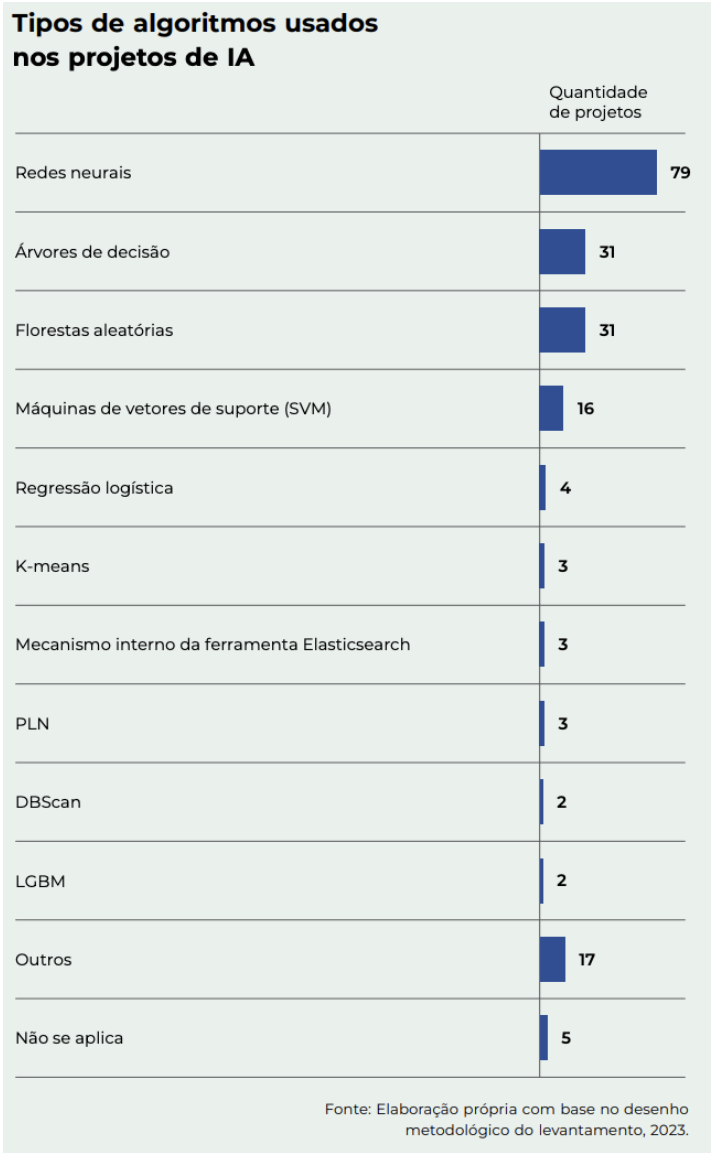
O relatório também evidenciou que, apesar do número elevado de projetos na Justiça Estadual, a densidade de projetos (média por tribunal) é maior nos tribunais superiores (2,6), seguida pelos tribunais federais (2,3) e estaduais (2,5), sugerindo maior concentração de recursos e dedicação à IA nesses segmentos (Conselho Nacional de Justiça, 2023f, p. 47). Os principais projetos em curso demonstram uma ampla aplicabilidade, com foco em eficiência, cooperação interinstitucional e desenvolvimento contínuo, mesmo diante de desafios como escassez de profissionais qualificados e infraestrutura tecnológica.

Interessante também observar que a pesquisa trouxe as principais técnicas de IA utilizadas nos projetos em andamento. As redes neurais lideram com 79 menções (56,4%), sendo preferidas devido à sua capacidade de lidar com tarefas complexas como processamento de linguagem natural, classificação de texto e reconhecimento de entidades nomeadas.. Em seguida, aparecem árvores de decisão e as *random forests*, cada uma com 31 menções (21,1%), apreciadas por sua robustez e interpretabilidade em decisões baseadas em dados estruturados. Máquinas de vetores de suporte foram citadas em 16 projetos (11,4%), e técnicas de clusterização como k-means (3 citações) e DBScan (2 citações) também foram mencionadas, evidenciando o interesse por métodos de agrupamento e detecção de padrões (Conselho Nacional de Justiça, 2023f, p. 71-72).

Além dos algoritmos, o relatório detalha os tipos de aprendizado empregados nos projetos. O aprendizado supervisionado foi o mais frequente, citado em 87 projetos (62,1%), geralmente utilizado em tarefas como classificação documental e previsão de decisões judiciais. Já o não supervisionado apareceu em 38 projetos (27,1%), aplicado na descoberta de padrões e agrupamentos em dados não rotulados, úteis para análise exploratória. O aprendizado por reforço, com 9 menções (6,4%), e o aprendizado por transferência, com 2 (1,4%), indicam iniciativas ainda incipientes, mas voltadas a tarefas mais complexas como adaptação de modelos e otimização de fluxos judiciais (Conselho Nacional de Justiça, 2023f, p. 74-75).

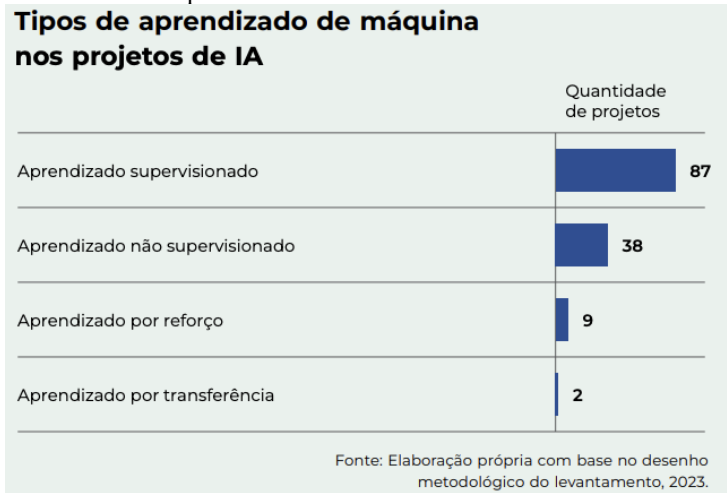
As Figuras 23 e 24, retiradas do relatório, demonstram esse panorama de forma quantitativa, revelando uma diversidade de abordagens e uma crescente sofisticação dos métodos utilizados.

Figura 23. Tipos de algoritmos usados.



Fonte: Conselho Nacional de Justiça, 2023f, p. 73.

Figura 24. Tipos de aprendizado de máquina usados.



Fonte: Conselho Nacional de Justiça, 2023f, p. 75.

Informação relevante trazida pela pesquisa foi quanto à propriedade dos dados utilizados nos sistemas de IA. A maioria dos projetos de IA se baseia em dados produzidos pelos próprios tribunais. De um total de 140 iniciativas, 114 (81,4%) utilizam dados internos, evidenciando a capacidade institucional de coletar, tratar e aplicar suas próprias informações no desenvolvimento de soluções tecnológicas. Em menor proporção, há projetos que fazem uso de dados fornecidos por outros órgãos do Judiciário (11,4%), de bases públicas disponíveis (11,4%) ou da plataforma Codex (10%). Outras fontes mencionadas incluem dados proprietários (5,7%) e, pontualmente, informações provenientes de sistemas como HuggingFace e SEEU (cada um citado em 0,7%) (Conselho Nacional de Justiça, 2023f, p. 89-90).

Quando analisadas as respostas por ramo da Justiça, nota-se uma tendência consistente de utilização de dados próprios. A Justiça Estadual se destaca, com 66,1% dos projetos baseados em informações geradas internamente, além de algum uso de dados públicos (8,8%), indicando disposição para complementar suas bases com fontes externas. Na Justiça do Trabalho, 60% das iniciativas também são alimentadas por dados internos, o que aponta para um esforço em adequar os projetos às especificidades desse ramo. Já os tribunais superiores (77%) e a Justiça Eleitoral (39,1%) seguem a mesma lógica, priorizando o uso de dados institucionais (Conselho Nacional de Justiça, 2023f, p. 90-91).

Quanto à publicização e reutilização dos sistemas, também foram colhidas respostas, identificando-se que, embora muitas soluções de IA desenvolvidas pelos tribunais sejam de uso restrito, há um número significativo de projetos que já estão acessíveis ou que possuem planos de disponibilização futura para retreino e reutilização. Do total de 140 iniciativas mapeadas, 54 projetos (38,6%) não são abertos ao público, enquanto 39 (27,9%) já permitem algum grau de reaproveitamento. Outros 23 projetos (16,4%) estão parcialmente disponíveis e 24 (17,1%) não apresentaram clareza quanto a essa questão, seja por incerteza ou pela não aplicabilidade da reutilização (Conselho Nacional de Justiça, 2023f, p. 91-93).

Na Justiça Estadual, há uma predominância de restrição: 44,1% dos projetos não estão disponíveis. Ainda assim, 25% dos projetos estão parcialmente acessíveis e 20,6% estão abertos ao público (Conselho Nacional de Justiça, 2023f, p. 93).

Na Justiça Federal, por sua vez, observa-se um cenário mais polarizado: 42,9% dos projetos são totalmente abertos e 42,9% são fechados ao público, revelando estratégias distintas entre os órgãos desse ramo. Além disso, 14,2% dos projetos federais não souberam informar sua situação. Já os tribunais superiores adotam majoritariamente uma postura aberta, com

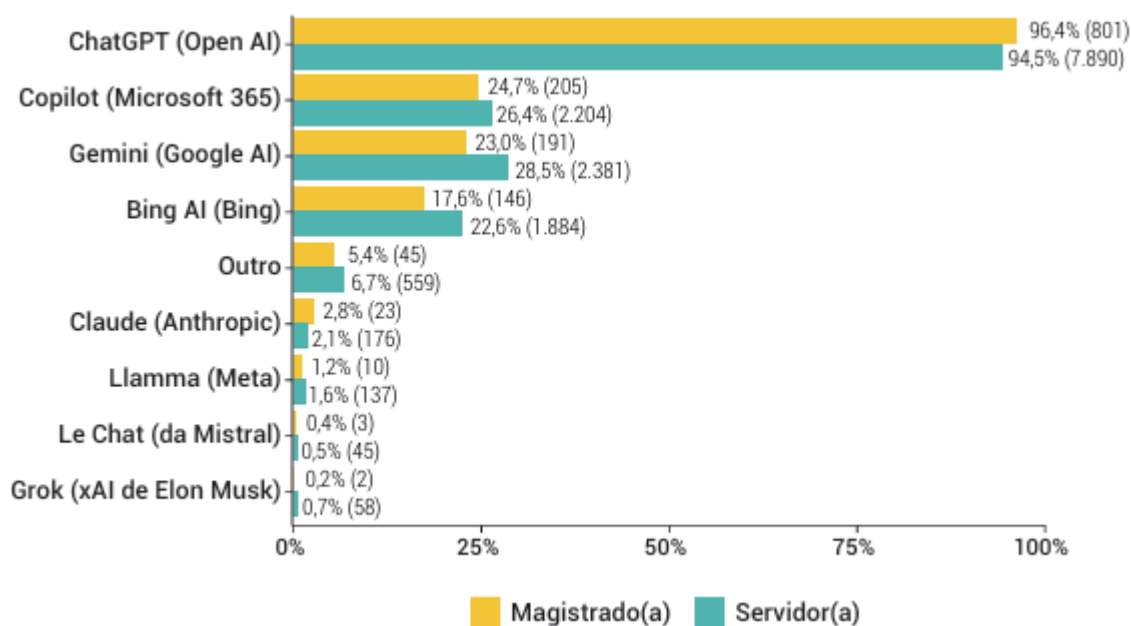
76,9% dos projetos disponíveis, enquanto a Justiça do Trabalho e a Eleitoral se mostram mais reservadas, com a maioria das iniciativas ainda restritas (Conselho Nacional de Justiça, 2023f, p. 93-94).

Outra pesquisa foi desenvolvida pelo CNJ em 2024, intitulada “O uso da Inteligência Artificial Generativa no Poder Judiciário brasileiro”. Nela, analisa-se o uso da Inteligência Artificial Generativa (IAG) por magistrados e servidores do Poder Judiciário brasileiro (Conselho Nacional de Justiça, 2024b). A pesquisa foi realizada por adesão, por meio de formulário eletrônico enviado a toda a magistratura e ao corpo técnico, para entender em que medida essas ferramentas vêm sendo utilizadas, para quais finalidades, com que frequência e quais são as percepções, dificuldades e expectativas associadas ao seu uso. Foram aplicadas perguntas objetivas e abertas, incluindo dados sociodemográficos e de experiência com IAGs (Conselho Nacional de Justiça, 2024b, p. 49-51).

Embora o formulário tenha sido enviado a todos os e-mails institucionais de juízes e servidores ativos do Poder Judiciário, somente foi respondido por 1.681 magistrados (9,1%) e 16.844 servidores (6,0%) de uma população de 18.464 magistrados e 278.755 servidores. A baixa taxa de adesão limita a representatividade dos dados e indica uma dificuldade de engajamento com o tema ou com o instrumento de coleta (Conselho Nacional de Justiça, 2024b, p. 50).

Os resultados indicam que aproximadamente metade dos respondentes já teve alguma experiência com ferramentas de IAGs, embora o uso cotidiano ainda seja pouco frequente. A maioria das pessoas que utilizam essas ferramentas o faz em versões abertas disponíveis na internet, como o ChatGPT, sendo ainda pouco expressivo o uso de ferramentas contratadas ou disponibilizadas oficialmente pelos tribunais, vide Figura 25. O uso é mais comum entre servidores e magistrados envolvidos com atividades acadêmicas, como estudantes ou docentes (Conselho Nacional de Justiça, 2024b, p. 52-60).

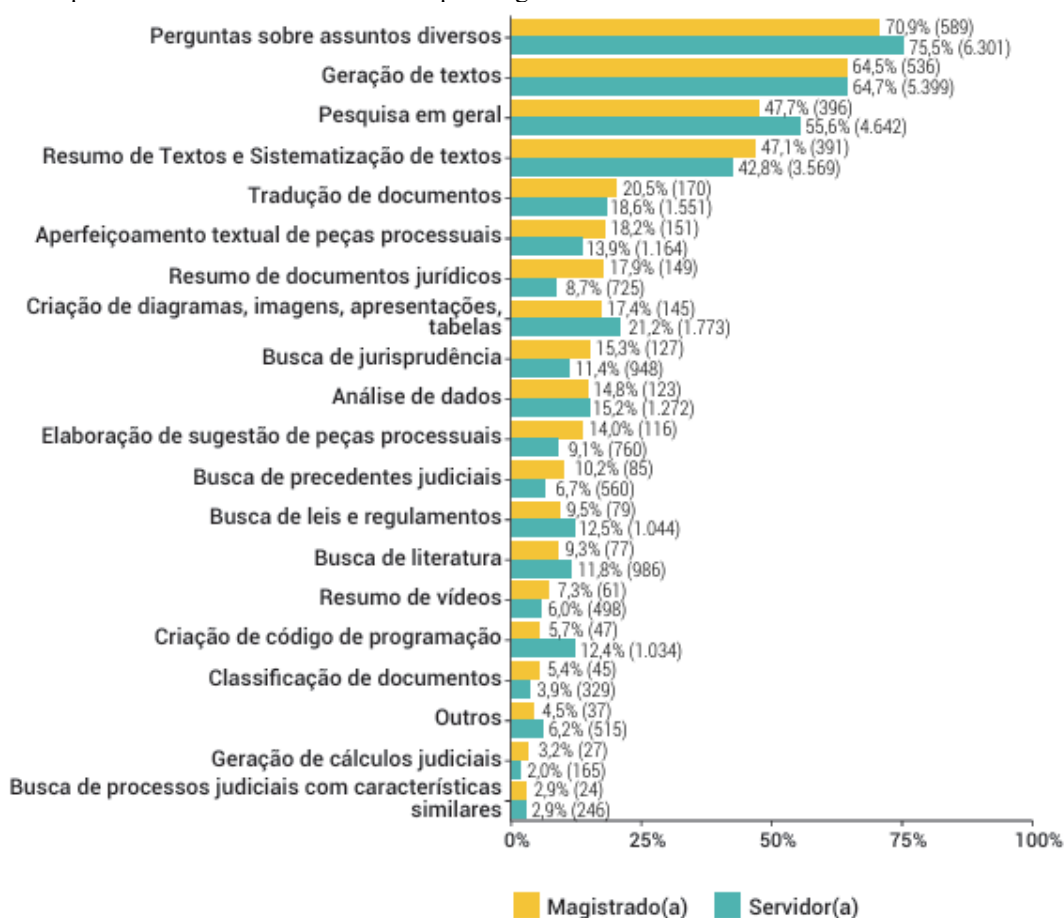
Figura 25. IAGs utilizadas por magistrados e servidores do Poder Judiciário.



Fonte: Conselho Nacional de Justiça, 2024b, p. 53.

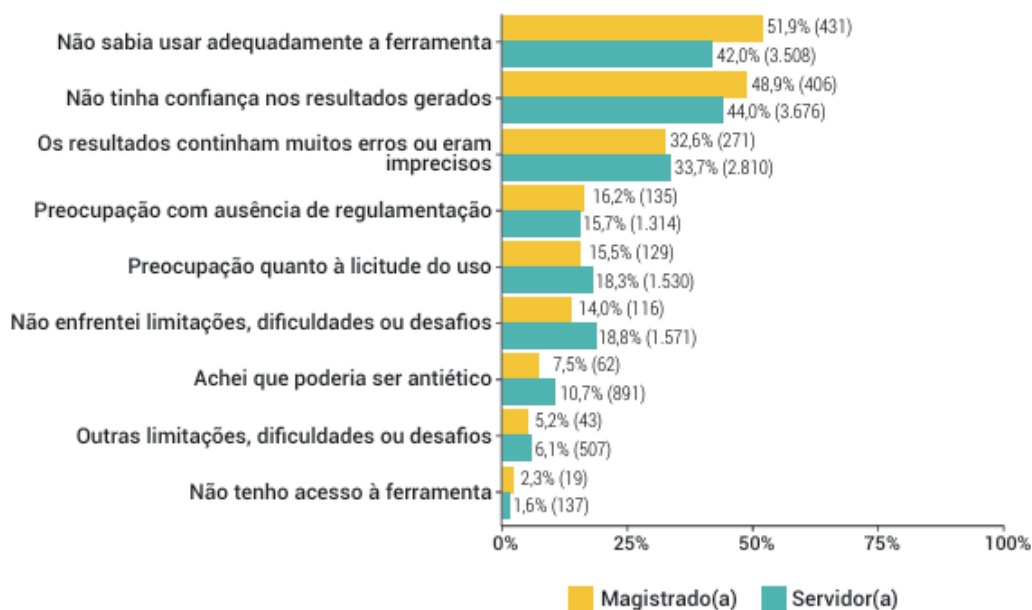
Quanto às finalidades de uso, destacam-se tarefas como geração de texto, tradução, resumo e aperfeiçoamento textual, tal como aponta a Figura 26. No entanto, também há uso para atividades mais técnicas, como buscas de jurisprudência e sugestão de peças processuais, o que levanta preocupações em razão das limitações técnicas dessas ferramentas (Conselho Nacional de Justiça, 2024b, p. 59-60). A maioria dos respondentes enxerga potencial utilidade das IAGs nas rotinas do Judiciário, mas também relata desafios, como desconhecimento sobre o uso adequado, falta de confiança nos resultados e dúvidas éticas e jurídicas, que indicam a necessidade de treinamento e diretrizes mais claras. As principais limitações relatadas estão elencadas na Figura 27, retirada do relatório.

Figura 26. Tipos de uso de ferramentas de IAG por magistrados e servidores.



Fonte: Conselho Nacional de Justiça, 2024b, p. 59.

Figura 27. Principais limitações apontadas pelos respondentes quanto ao uso de IAGs.



Fonte: Conselho Nacional de Justiça, 2024b, p. 64.

Por fim, a pesquisa revelou que cerca da metade já teve experiência com ferramentas de IA, especialmente em versões abertas disponíveis na internet, embora seu uso profissional ainda seja pouco frequente. A utilização é mais expressiva entre aqueles com atuação acadêmica, e há elevado interesse em capacitação, evidenciando uma tendência de crescimento no uso da tecnologia. Os usos mais comuns incluem pesquisa geral e busca de jurisprudência, mas persistem desafios como falta de familiaridade, erros nos conteúdos gerados e dúvidas sobre ética e licitude, que comprometem a transparência – já que a maioria não comunica o uso aos colegas ou superiores. Por fim, os respondentes destacam a importância de programas de formação, adoção oficial de ferramentas e valorização dos servidores, reconhecendo o potencial das IAGs para melhorar a eficiência e a qualidade da atuação jurisdicional (Conselho Nacional de Justiça, 2024b, p. 80-81).

Na 4ª edição do estudo intitulado “Inteligência Artificial: tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário brasileiro”, de outubro de 2025, coordenado por Salomão e Leme (2025), é feito um mapeamento dos sistemas de IA a partir do Painel de Projetos de Inteligência Artificial do CNJ (Conselho Nacional de Justiça, 2025b), identificando-se como as iniciativas se distribuem por tipo de atividade e por ramo da Justiça, desenvolvendo-se, ainda, uma análise qualitativa de casos representativos de IA generativa em operação no STF, STJ, TRT da 4ª Região etc. O estudo descreve as funcionalidades, finalidades, impactos e resultados observados em aplicações como apoio à gestão processual, mecanismos de triagem e produção de minutas.

O estudo apresenta mais do que um simples inventário de projetos de IA no Judiciário brasileiro: ele revela a arquitetura tecnológica emergente do sistema de justiça e evidencia como diferentes tribunais estão mobilizando a IA para enfrentar desafios estruturais. Embora cada iniciativa tenha objetivos específicos, a leitura conjunta permite identificar padrões claros de evolução institucional, maturidade tecnológica e prioridades funcionais (Salomão; Leme 2025).

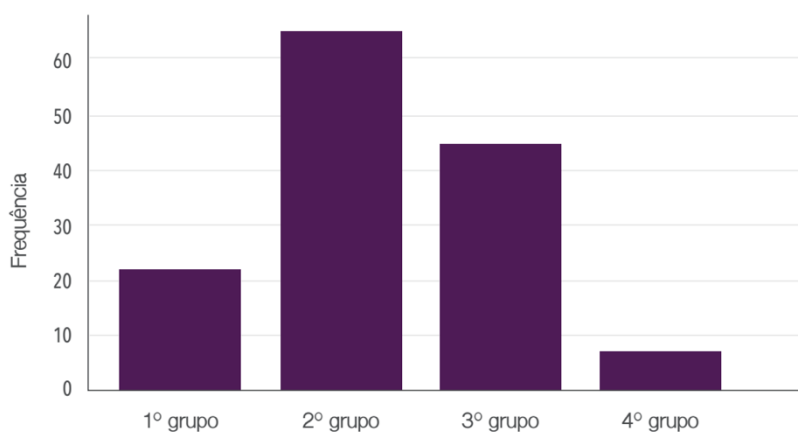
Em primeiro lugar, destaca-se a diversidade de órgãos presentes que incluem tribunais superiores, conselhos, justiças estaduais e trabalhistas. Demonstra-se que o uso de IA não está concentrado em um segmento específico, mas se espalha por todo o ecossistema do Judiciário. Essa disseminação indica que o processo de transformação digital já constitui uma estratégia nacionalizada, ainda que com níveis diferentes de desenvolvimento entre os órgãos (Salomão; Leme 2025).

A própria variação no estágio de evolução dos projetos reforça essa percepção. No estudo, observa-se a reunião desde o desenvolvimento de sistemas ainda não iniciados ou em

fase de concepção até sistemas plenamente implantados e operando em produção. Esse cenário misto sugere que, enquanto alguns tribunais já consolidaram capacidades tecnológicas internas, outros ainda estão construindo as bases para adoção futura. Ao mesmo tempo, a coexistência de projetos embrionários com soluções maduras indica um Judiciário em processo de experimentação contínua, buscando validar tecnologias, ajustar metodologias e amadurecer a governança de IA (Salomão; Leme 2025).

Outro aspecto relevante trazido por Salomão e Leme (2025, p. 54) é a distribuição dos projetos entre os quatro grupos de atividades, tal como na Figura 28. Compõem o grupo 1, ferramentas voltadas às atividades administrativas do Judiciário, como gestão interna e comunicação com o público (22 projetos de IA); o grupo 2, sistemas que automatizam fluxos processuais e tarefas de apoio, incluindo triagem, classificação, transcrição e identificação de padrões (66 projetos de IA); o grupo 3, tecnologias que auxiliam diretamente a função jurisdicional, como elaboração de minutas, análise de peças, sumarização e recomendações processuais (45 projetos de IA); e, o grupo 4, ferramentas que apoiam práticas de conciliação e ampliam métodos autocompositivos (8 projetos de IA).

Figura 28. Distribuição das ferramentas de IA até março de 2025.



Fonte: Salomão e Leme, 2025, p. 92.

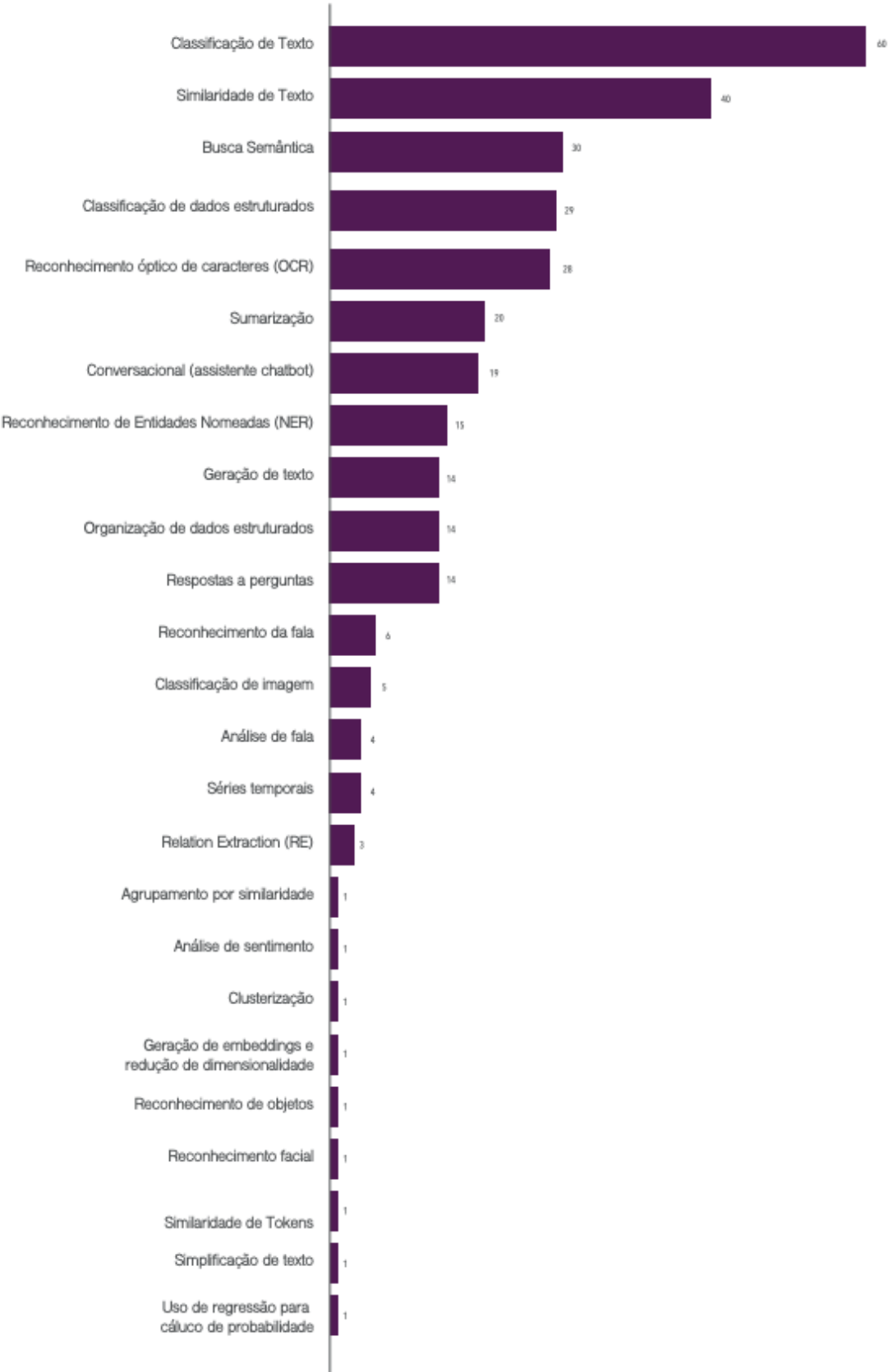
Embora existam iniciativas administrativas e algumas voltadas à conciliação, observa-se uma concentração em dois macroeixos: (1) automação dos fluxos processuais e das tarefas de apoio, e (2) suporte qualificado à atividade jurisdicional. Essa prevalência reflete o entendimento de que os maiores gargalos do Judiciário estão tanto na burocracia operacional quanto nas etapas de análise e elaboração de decisões. Ferramentas como triagem de processos, agrupamento por similaridade, extração de informações e geração de minutas aparecem como

estratégias centrais para enfrentar o volume crescente de demandas. Em outras palavras, a IA é vista não apenas como instrumento de modernização administrativa, mas como recurso para reterritorializar o trabalho jurídico, redistribuindo tarefas repetitivas e permitindo que o corpo técnico e os magistrados se concentrem nas atividades de maior complexidade decisória.

A descrição minuciosa dos projetos revela ainda que muitos tribunais buscam soluções híbridas, combinando técnicas tradicionais (como reconhecimento óptico de caracteres – OCR – ou classificação) com modelos mais avançados, incluindo uso emergente de IA generativa. A presença de iniciativas baseadas em análise preditiva, reconhecimento facial, sumarização textual ou recomendação de jurisprudência mostra uma clara tendência de sofisticação do ecossistema tecnológico, com crescente utilização de modelos complexos e acoplados a bases extensas de dados (Salomão; Leme 2025).

De acordo com Salomão e Leme (2025, p. 97), o Painel de Projetos de Inteligência Artificial do CNJ (Conselho Nacional de Justiça, 2025b), aponta que a maior quantidade de tarefas aplicadas nos projetos de IA se concentra na classificação de texto, seguida de similaridade de texto, busca semântica, classificação de dados, OCR e sumarização. Vide Figura 29 acerca das tarefas aplicadas em 2024, segundo os autores. Além disso, a maioria dos dados utilizados como fonte nos projetos são gerados nos próprios tribunais, estando diretamente relacionados às rotinas processuais dos tribunais (Salomão; Leme, 2025, p. 98-99), sendo que grande dos projetos disponibiliza o acesso completo ao código-fonte, viabilizando auditorias e correções, existindo aqueles que disponibilizam esse acesso parcialmente ou não disponibilizam (Salomão; Leme, 2025, p. 100).

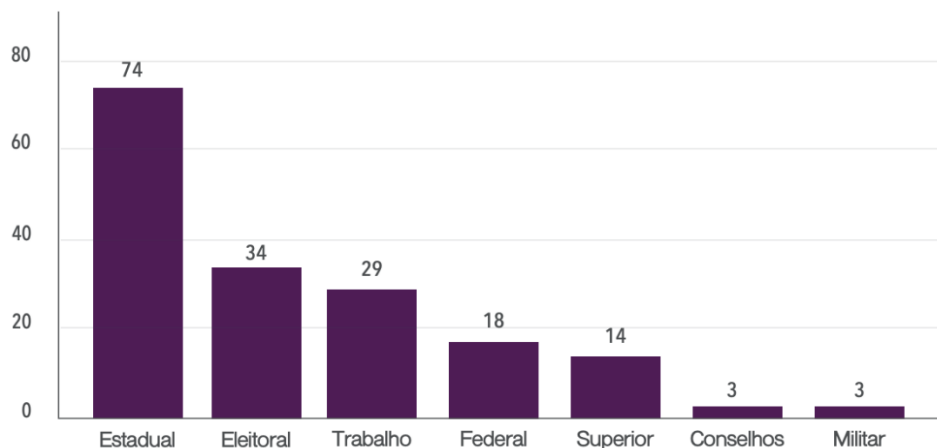
Figura 29. Tipos de tarefas aplicadas nos projetos de IA no Judiciário em 2024.



Fonte: Salomão e Leme, 2025, p. 97.

Ademais, o documento elaborado por Salomão e Leme (2025, p. 93) também aponta que a maior concentração de projetos em março de 2025 está na Justiça Estadual (74 indicações de projetos de IA), conforme dados do CNJ. Confira-se a Figura 30 retirada do estudo.

Figura 30. Projetos de IA por ramo de Justiça.



Fonte: Salomão e Leme, 2025, p. 93.

Questão preocupante é ressaltada pelos autores, que salientam que 48,6% dos projetos em desenvolvimento não possuem documentação, o que impacta a rastreabilidade e a integridade das soluções de IA em desenvolvimento (Salomão; Leme, 2025, p. 101). Além disso, quanto aos projetos de IA desenvolvidos em 2024, os maiores desafios encontrados foram a “dificuldade de encontrar profissionais qualificados para trabalhar com essa tecnologia (15,7%)”, seguida de dificuldades relacionadas às questões de privacidade e de segurança dos dados (14,9%), na “obtenção de dados em quantidade, qualidade e/ou diversidade necessária (12,3%)”, na “necessidade de adaptação de processos e rotinas já estabelecidas (12,3%)” e na “complexidade na integração de sistemas de IA com os sistemas existentes (11,9%)” (Salomão; Leme, 2025, p. 128).

Ainda quanto à aplicação da IA no Poder Judiciário, em maio de 2025, todos os tribunais brasileiros obtiveram acesso à Assistente Pessoal Operada por Inteligência Artificial (ApoIA), a primeira ferramenta de IA generativa integrada à Plataforma Digital do Poder Judiciário Brasileiro (PDPJ-Br) (Justiça Federal 2ª Região, 2025; Pereira, 2025). Desenvolvido pelo TRF2 em parceria com o CNJ, o sistema foi incorporado por meio do Conecta, iniciativa do programa Justiça 4.0 que busca mapear, compartilhar e nacionalizar soluções tecnológicas entre os tribunais (Pereira, 2025).

A ApoIA tem por objetivo atuar como um suporte operacional para magistrados e servidores em diversas atividades do cotidiano judicial, como criação de relatórios e ementas, revisão de textos jurídicos, sínteses processuais, triagem temática, visualização de acervos e identificação de litigância predatória ou ações repetitivas (Pereira, 2025). Ela transforma comandos dos usuários em produções automatizadas, por meio de um banco colaborativo de prompts, que permite o reaproveitamento e o compartilhamento de instruções entre tribunais, sempre com supervisão humana (Pereira, 2025).

Além de acelerar e padronizar a produção de conteúdos jurídicos, a integração da ApoIA à PDPJ-Br busca efetivar uma alternativa institucional à utilização de ferramentas privadas não regulamentadas, garantindo mais segurança, ética, proteção de dados e padronização nos comandos usados em todo o Judiciário (Pereira, 2025). A plataforma também conta com iniciativas educacionais: o TRF2 promoveu treinamentos sobre IA generativa e proteção de dados para gestores, alinhando o uso da ferramenta à Resolução n. 615/2025 do CNJ (Justiça Federal 2ª Região, 2025; Conselho Nacional de Justiça, 2025c).

4.2 O que podemos esperar no futuro?

De modo geral, a forma mais “avançada” de aplicação da IA no Direito visa à previsão de decisões judiciais. É preciso entender como funcionam os sistemas que estão sendo desenvolvidos e implementados e o conceito de predição analítica, de onde serão previstos padrões de conduta ou de tomada de decisões judiciais (Cárdenas e Molano, 2021).

Como visto, a aprendizagem de máquina preditiva é uma ferramenta que se baseia na análise de um grande banco de dados (*big data*) e dele se utiliza para prever decisões futuras através do estudo sistemático de decisões já exaradas (Reyes, 2019). No Direito, partindo de um conjunto de dados composto por decisões passadas, tal mecanismo pode ser utilizado para prever decisões futuras por meio de algoritmos probabilísticos e de análise de padrões. Em outras palavras, a partir de um grupo de dados se determinam padrões e relações entre diversas variáveis para prever um determinado comportamento futuro (Cárdenas; Molano, 2021, p. 5).

De toda forma, sejam as decisões que compõem sua base de dados, seja a programação dos algoritmos que delineiam as tarefas necessárias para a previsão de decisões futuras, ambas são realizadas/produzidas por seres humanos. A programação de um sistema voltado ao Direito demanda, ao menos, a colaboração dos programadores e dos analistas jurídicos, pois a qualidade e a quantidade de informações dependerão do que ambos consideram apropriado, levando em

conta a complexidade dos problemas a serem abordados. Além disso, é relevante salientar que os sistemas especialistas precisam ser atualizados constantemente com novos dados para realizar previsões em conformidade com o estado atual (Badaró et al., 2013).

Quanto às decisões judiciais, os sistemas especialistas devem ser programados com informações em total consonância com o ordenamento jurídico. Isso implica a consideração de regulamentos em diversos níveis, como leis, decretos, regulamentos, circulares, entre outros, bem como decisões dos principais tribunais e cortes, possibilitando uma análise abrangente do problema em foco. Mesmo tratados internacionais aplicáveis no contexto brasileiro devem ser levados em consideração pelo sistema inteligente.

É fundamental compreender que, independentemente da crescente implementação de IA no Poder Judiciário brasileiro, os sistemas não devem ser os únicos encarregados de tomar todas as decisões judiciais, mas sim de participar ativamente desse processo, orientando os juízes sobre o caminho ou os fatores a serem considerados na tomada de decisão. A IA deve apenas fornecer os elementos a serem ponderados na decisão com base em precedentes e na legislação em vigor. Magistrados devem, por sua própria vontade, examinar outros elementos e avaliar as evidências de forma imparcial antes de emitir sua sentença (Cárdenas; Molano, 2021, p. 9).

De fato, as máquinas podem ouvir, falar, “pensar” (calcular) e julgar. No entanto, como vemos neste trabalho, estão longe de conseguir pensar e julgar como os seres humanos. No que diz respeito às atividades judiciais, juízes, promotores, advogados, defensores e demais profissionais do Direito aplicam as normas legais com base em suas estruturas de conhecimento, experiências práticas e vivências. Assim, não há como uma IA pensar juridicamente e, eventualmente, julgar, da mesma forma que esses profissionais. A função da IA é auxiliar na gestão de casos.

A colaboração entre humanos e máquinas é uma das principais características da IA. Fomentar essa parceria para ajudar o Judiciário é o caminho adequado para uma integração profunda entre IA e justiça (Cui, 2020, p. 180). Fala-se, assim, em “corrobótica” no âmbito da administração da justiça brasileira, que consistiria na coexistência entre humanos e máquinas. Juízes somente serão “substituídos” em situações específicas, dadas as limitações do aprendizado de máquina. Torna-se importante, no entanto, compreender como os algoritmos influenciam o processo decisório, que é concretizado por juízes, bem como estar ciente do alcance da autonomia (capacidade de ditar regras) do sistema utilizado (Boeing, 2019, p. 69-70).

Como aborda Porto (2018, p. 134), a IA se instaura no Poder Judiciário sob os seguintes pretextos:

(a) auxiliar o magistrado na realização de atos de constrição (penhora *online*, *Renajud* e outros); (b) auxiliar o magistrado a identificar os casos de suspensão por decisões em recursos repetitivos, IRDR, reclamações, etc., possibilitando que o processo seja identificado e suspenso sem esforço humano maior do que aquele baseado em confirmar o que a máquina apontou; (c) auxiliar o magistrado na degravação de audiências, poupando enorme tempo; (d) auxiliar na classificação adequada dos processos, gerando dados estatísticos mais consistentes; (e) auxiliar o magistrado na elaboração do relatório dos processos, filtrando as etapas relevantes dos processos e sintetizando os mesmos; (f) auxiliar na identificação de fraudes; (g) auxiliar na identificação de litigante contumaz; (h) auxiliar na identificação de demandas de massa; (i) auxiliar na avaliação de risco (probabilidade/impacto de algo acontecer no futuro); (j) auxiliar na gestão relativa à antecipação de conflitos a partir de dados não estruturados; (k) auxiliar o magistrado na avaliação da jurisprudência aplicada ao caso; (l) possibilitar uma melhor experiência de atendimento ao usuário: sistemas conversacionais, “*chat bot*” (atendimento para ouvidoria e corregedoria); (m) identificar votos divergentes na pauta eletrônica; (n) auxiliar na gestão cartorária, identificando pontos de gargalos, processos paralisados, servidores com menor/maior carga de trabalho; (o) identificar e reunir processos para movimentação em lote; e (p) auxiliar o magistrado na elaboração de minutas de despachos, decisões e sentenças.

Este estudo aborda precisamente esse novo cenário: a realidade que inevitavelmente impactará todo o sistema do Poder Judiciário, com o crescente uso de ferramentas de IA, como algoritmos decisórios, análise preditiva, jurimetria e os chamados *judgebots* ou juízes robôs.

É fundamental respeitar as leis e características do sistema judicial e também é imprescindível entender as particularidades da IA implementada. Ao aproveitar plenamente a tecnologia moderna por meio de uma integração eficaz, podemos aprimorar conceitos e ideias, inovar sistemas e mecanismos, padronizar comportamentos judiciais e melhorar o sistema de litígios, otimizando assim o acesso à justiça.

Também se torna imprescindível que essa implementação tecnológica se dê acompanhada pela observância da necessidade de se proteger e garantir a efetivação de direitos, de modo que as inovações sejam inclusivas. Em outras palavras, o desafio está em implementar a IA atentando-se aos princípios e regras que fundamentam nosso sistema jurídico.

Se por um lado a IA promete maior acesso à justiça e otimiza as atividades do Poder Judiciário trazendo maior eficiência, por outro, pode-se estar caminhado ao encontro de um cenário de acirramento de discriminações por meio de algoritmos viesados, que estarão protegidos por uma crença de imparcialidade e cientificidade falsa, algoritmos esses que terão impacto direto em aspectos sensíveis da vida de várias pessoas (Boeing, 2019, p. 12).

Como destacam Nunes e Marques (2018), algoritmos de IA são enigmáticos e sua aplicação exige supervisão rigorosa, pois a forma como as decisões são tomadas e a subjetividade envolvida podem impactar significativamente os direitos fundamentais dos

indivíduos, gerando questões complexas no campo jurídico. O uso da IA no Direito, apesar das vantagens proporcionadas, traz consigo a imprescindibilidade de observância de princípios éticos que guiam a sua utilização, garantindo a preservação de valores essenciais ao funcionamento do Direito, como o devido processo legal e o acesso à justiça (Gabriel et al., 2025; Brunetta et al. 2025).

Para assegurar decisões mais seguras na implementação desses sistemas, é crucial promover um debate ético aprofundado sobre a melhor forma de desenvolvê-los e integrá-los, buscando reduzir problemas que envolvem, principalmente, discriminações e vieses, possibilitando sua evolução e adaptação conforme as necessidades atuais e futuras da sociedade. Ainda, ao utilizar sistemas inteligentes, é preciso levar em consideração que a IA é incapaz de substituir atualmente a apreciação e a avaliação humana na tomada de decisões judiciais. Além disso, tais sistemas não conseguem integrar todos os elementos sociais em uma decisão jurídica, o que pode comprometer a justiça e a equidade dos resultados obtidos por meio de sistemas jurídicos automatizados (Alves, 2020, p. 44-45).

Para confiar nas decisões feitas pelos sistemas inteligentes, é preciso que eles sejam interpretáveis. A interpretabilidade de um sistema inteligente significa entender como funciona por dentro para explicar por que sugere uma determinada escolha. Em alguns casos e para certas decisões judiciais, não basta apenas receber a resposta; é fundamental entender como ele chegou a essa resposta, algo que apenas é alcançável graças à interpretabilidade e à transparência:

A necessidade de interpretabilidade dos modelos decorre de uma incompletude na formalização do problema que se pretende resolver, o que significa que, para certos problemas ou tarefas, não é suficiente obter apenas a resposta do modelo. O modelo também deve fornecer uma explicação sobre como chegou à sua resposta, pois uma predição correta resolve apenas parcialmente o problema original (Equipo de Expertos en Ciencia y Tecnología, 2018, tradução nossa³⁰).

De fato, a IA pode solucionar problemas de morosidade processual e de segurança jurídica, auxiliando consideravelmente na gestão do sistema judicial brasileiro. A utilização de tais técnicas pode aumentar a confiança que os indivíduos possuem na administração da justiça, enquanto se acredita que algoritmos podem ser confiáveis, neutros e imparciais, bem como passíveis de serem explicados e transparentes.

Em que pese a IA possa trazer maior segurança jurídica, ela também pode ocasionar um engessamento do Poder Judiciário (Fonseca, 2022, p. 352). A IA surge como uma

³⁰ No original: “La necesidad de interpretabilidad de los modelos proviene de una incompletitud en la formalización del problema que se pretende resolver, lo que significa que para ciertos problemas o tareas no es suficiente obtener la respuesta del modelo. El modelo también debe proporcionarnos una explicación de cómo llegó a su respuesta, porque una predicción correcta solo resuelve parcialmente el problema original”.

ferramenta para mitigar ou responder ao que tem sido chamado de ativismo judicial, alinhando-se à corrente da autocontenção, também conhecida como minimalismo, visando alcançar a segurança jurídica desejada ao afastar viés de confirmação e princípios heurísticos. Enquanto a campanha contra o ativismo judicial o apresenta como um “vilão a ser derrotado”, é importante reconhecer seu papel relevante na discussão e manutenção dos direitos.

Por exemplo, em Fonseca (2022), entende-se que a IA e seus algoritmos decisórios representam uma ruptura no sistema judiciário, podendo potencialmente reduzir o ativismo judicial e até mesmo transformar o Judiciário como o conhecemos, fazendo dele uma máquina mais econômica e mais segura (no sentido de trazer segurança jurídica, decisões judiciais mais padronizadas e com base em precedentes).

Não se pode esquecer de que a integração de IA no Poder Judiciário brasileiro continua em fase inicial e passará por um extenso processo de debates éticos, filosóficos e sociológicos para considerar adequadamente os valores e sentimentos humanos. No futuro, é esperado que se chegue a um entendimento comum e a um uso seguro e eficaz dessas tecnologias.

Em resumo, a IA oferece numerosos benefícios à administração da justiça, otimizando o sistema Judiciário brasileiro. Atualmente, essas tecnologias não conseguem abranger todos os elementos jurídicos e mesmo assim demandam, desde já, a observância de regras éticas e de governança algorítmica. Limitar o acesso à justiça e a publicidade dos atos judiciais, mesmo com a implementação de novas tecnologias, configura uma negação do direito à informação e à participação da sociedade. A justiça deve ser transparente e acessível a todos, não um privilégio de poucos. Manter a desconfiança e a insegurança no seio do sistema judicial, mesmo com a modernização tecnológica, representa um retrocesso. A busca pela confiança e pela segurança jurídica deve ser um objetivo central em qualquer processo de reforma.

Não obstante, ao invés de priorizar a rapidez desenfreada e a economia a qualquer custo, o Poder Judiciário deve trilhar um caminho de modernização equilibrada e responsável. Isso significa investir em sistemas que garantam a qualidade das decisões, facilitem o acesso à justiça para todos e promovam a transparência dos atos judiciais.

A tecnologia, quando utilizada de forma consciente e ética, pode ser uma poderosa ferramenta para construir um Poder Judiciário mais eficiente, acessível e justo. Cabe a nós, como sociedade, exigir que essa modernização seja feita de forma responsável, priorizando a qualidade da justiça e o bem-estar social.

Nos subitens a seguir, analisam-se, de forma sucinta, os benefícios, os riscos e os desafios quanto à implementação da IA no Poder Judiciário brasileiro.

4.2.1 Os benefícios da implementação da inteligência artificial no Poder Judiciário brasileiro

A efetivação do direito fundamental de acesso à justiça, previsto no artigo 5º, inciso XXXV da Constituição Federal de 1988, permanece como um dos maiores desafios enfrentados pelo sistema jurídico brasileiro (Brasil, 1988). Apesar dos avanços legislativos e institucionais desde a década de 1990, o Judiciário ainda se mostra incapaz de oferecer uma prestação jurisdicional célere, eficiente e socialmente justa. A sobrecarga processual, os entraves econômicos e a complexidade procedimental contribuem para a perpetuação de um cenário de ineficácia e frustração social.

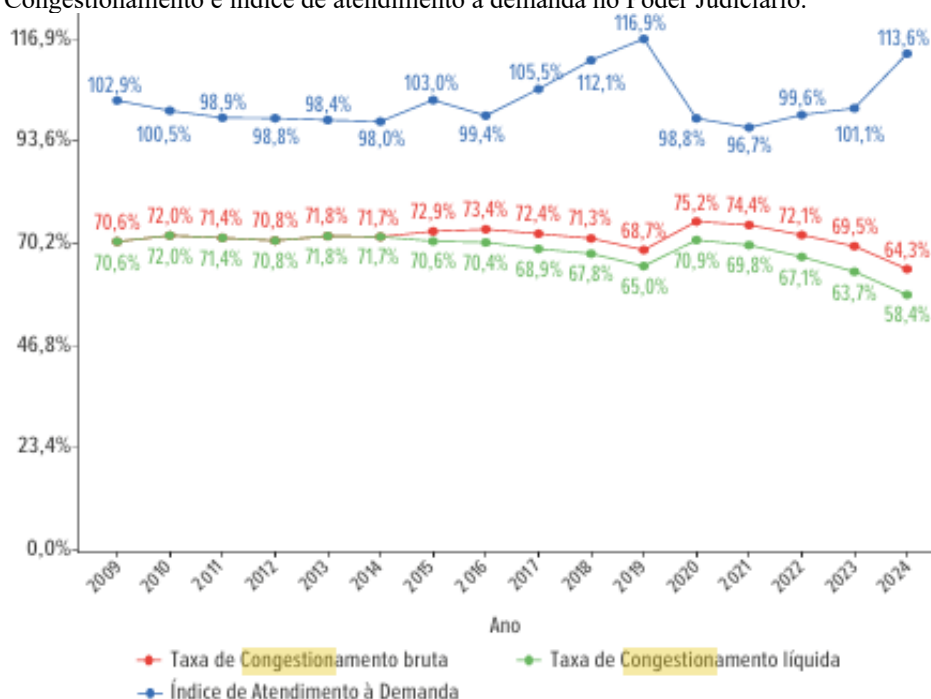
A incorporação de tecnologias disruptivas – especialmente a IA, o aprendizado de máquina e o *blockchain* – emerge como uma alternativa promissora para a reconfiguração do sistema de justiça. Pode-se dizer que entre as principais finalidades de aplicação de técnicas de IA no Poder Judiciário, estão a busca de celeridade e economia processual, uniformização de decisões e combate ao ativismo judicial. É sabido que o Poder Judiciário brasileiro enfrenta, há tempos, críticas relacionadas à morosidade processual, insegurança jurídica e falhas na prestação jurisdicional (Fonseca, 2022, p. 131).

A fim de estabelecer um diálogo entre o Direito e a IA, Oliveira (2019, p. 13-15) explica como e em que medida se dá a automação no Poder Judiciário brasileiro, esclarecendo ser ela uma estratégia adequada para a solução de diversos problemas encontrados no sistema jurídico brasileiro, como, por exemplo, o congestionamento de processos e o ativismo judicial.

A taxa de congestionamento do Poder Judiciário possui vinculação com o crescimento do número de ações ajuizadas, com a diminuição do número de juízes e servidores, com o excessivo número de demandas repetitivas, a falta de cultura conciliatória e de instrumentos efetivos de resolução extrajudicial de conflitos (Oliveira, 2021). A situação crítica do volume de processos no Judiciário é visualizável no relatório “Justiça em números 2024” do CNJ (Conselho Nacional de Justiça, 2024a) e no relatório “Justiça em números 2025” do CNJ (Conselho Nacional de Justiça, 2025^a).

Da Figura 31, retirada do relatório Justiça em números 2025, tem-se que, em 2024, a taxa de congestionamento bruta do Poder Judiciário brasileiro foi de 64,3%. A taxa de congestionamento líquida, que desconsidera os processos suspensos, sobrestados ou em arquivo provisório, atingiu 58,4% em 2024, apresentando uma redução de 4,3 ponto percentual em relação a 2023 (Conselho Nacional de Justiça, 2025a, p. 160).

Figura 31. Congestionamento e índice de atendimento à demanda no Poder Judiciário.



Fonte: Conselho Nacional de Justiça, 2025a, p. 259.

Apesar dessa melhora marginal, o relatório indica que há um volume expressivo de processos pendentes: ao final de 2024, o Judiciário contava com 80,6 milhões de processos em tramitação (Conselho Nacional de Justiça, 2025a, p. 623). Neste contexto, a IA, em particular, apresenta-se como ferramenta capaz de lidar com o chamado “big data jurídico”, automatizar tarefas repetitivas, identificar padrões decisórios e otimizar a gestão processual. A experiência brasileira, exemplificada pelo projeto Victor do STF, demonstra que a aplicação de IA pode reduzir custos, acelerar a tramitação de recursos e ampliar a eficiência institucional.

O estudo de Salomão e Leme (2025, p. 128), que debate os dados do “Painel de Projetos de IA no Poder Judiciário” de 2024, do CNJ, traz consigo um panorama dos benefícios buscados ou já alcançados com a IA no Judiciário no ano de 2024, com quatro resultados se destacando: maior eficiência e rapidez no tratamento de documentos e informações (15,7%); diminuição do tempo de tramitação processual (13,6%); uso mais racional de recursos, reduzindo custos operacionais (13,4%); e automação de tarefas rotineiras e burocráticas (13,2%).

Mais do que uma inovação técnica, a IA representa uma nova fase no movimento de renovação do acesso à justiça. Se as três ondas identificadas por Cappelletti e Garth (1998) – assistência jurídica gratuita, proteção de interesses difusos coletivos e simplificação procedimental – buscaram ampliar as portas de entrada do sistema judicial, a quarta onda

tecnológica visa reestruturar também as portas de saída, garantindo que os conflitos sociais sejam resolvidos em tempo razoável e com justiça substancial.

Pesquisas recentes corroboram o potencial transformador da IA. Salomão (2020) aponta que a IA contribui não apenas para a celeridade, mas também para o fortalecimento da segurança jurídica. Ela surge como uma resposta promissora à morosidade judicial, cujas causas estão associadas ao excesso de demandas, à falta de estrutura, à desigualdade de acesso e à escassez de pessoal. A IA pode beneficiar o sistema ao automatizar tarefas repetitivas, como triagem, agrupamento de processos semelhantes e sugestão de minutas, reduzindo a sobrecarga dos juízos e aumentando a eficiência no trâmite das ações.

Em “Robôs substituem juízes? O estado da arte da inteligência artificial no Judiciário brasileiro”, Cruz et al. (2022) analisam o uso da IA no sistema de justiça brasileiro, destacando os principais benefícios já observados em sua aplicação prática. Segundo os autores, a IA tem desempenhado papel importante na modernização da gestão processual, especialmente por meio da automatização de tarefas operacionais como triagem, classificação documental, identificação de padrões e extração de informações relevantes (Cruz et al., 2022, p. 34). Ferramentas como o Victor (STF), Sócrates (STJ) e Bem-te-vi (TST) demonstram ganhos expressivos de tempo e produtividade – o sistema Victor, por exemplo, realiza em cinco segundos uma atividade que levaria 44 minutos para ser feita manualmente por um servidor (Cruz et al., 2022, p. 16).

Outro benefício importante é a contribuição da IA para a uniformização e previsibilidade das decisões judiciais, ao identificar precedentes, agrupar casos semelhantes e facilitar o cumprimento do sistema de precedentes qualificados (Cruz et al., 2022, p. 19-21). Isso fortalece a segurança jurídica e auxilia na redução da litigiosidade. Além disso, com mais de 78 milhões de processos em estoque em 2018, os autores defendem que a IA é essencial para enfrentar a sobrecarga crônica do Judiciário, promovendo maior produtividade e economia de recursos (Cruz et al., 2022, p. 32-33). Iniciativas como o projeto Sinapses, desenvolvido pelo CNJ em parceria com o TJRO, também são destacadas como soluções inovadoras que incentivam a interoperabilidade e o compartilhamento de modelos entre tribunais (Cruz et al., 2022, p. 15).

Em Dykstra et al. (2022), esclarece-se como a integração entre IA e tecnologias digitais no âmbito do Programa Justiça 4.0 representa um avanço estratégico para o fortalecimento da governança do Poder Judiciário brasileiro. Essa combinação possibilita uma gestão mais transparente e orientada por dados, na qual decisões administrativas e operacionais

passam a ser apoiadas por análises precisas e por indicadores confiáveis, permitindo identificar gargalos, antecipar demandas e alocar recursos de forma mais racional.

Além de aprimorar a governança, a utilização de IA contribui diretamente para o aumento da produtividade, na medida em que automatiza tarefas repetitivas e operacionais, libera servidores e magistrados para atividades que exigem maior capacidade de análise e julgamento e acelera significativamente o trâmite processual. Esse ganho de eficiência se traduz em menor tempo de espera para as partes e na redução do acúmulo de processos, um dos problemas históricos do sistema judicial brasileiro (Dykstra et al., 2022, p. 286).

Outros estudos também apontam que a IA pode auxiliar na preservação do sistema de precedentes e na identificação de violações a direitos fundamentais. Em Ribeiro (2024), por exemplo, verifica-se que a IA tem potencial para promover uma transformação profunda no funcionamento da justiça brasileira. O autor apresenta uma análise dos principais benefícios operacionais, estruturais e sociais da IA, ressaltando seu papel central na promoção de uma justiça mais eficiente, acessível e transparente. Dentre eles, destaca-se a automação de tarefas repetitivas – como triagem de processos, análise de documentos e geração de relatórios – que libera juízes e servidores para atuarem em tarefas mais complexas (Ribeiro, 2024, p. 4-5). A IA também amplia o acesso à justiça, especialmente por meio de assistentes virtuais e chatbots que orientam cidadãos sobre seus direitos, agendamentos e procedimentos, com destaque para populações em regiões remotas ou em situação de vulnerabilidade (Ribeiro, 2024, p. 6-7). Além disso, a análise preditiva de decisões judiciais contribui para maior previsibilidade, segurança jurídica e consistência nas sentenças, fortalecendo a confiança no sistema (Ribeiro, 2024, p. 6).

A IA tem se mostrado especialmente eficaz no enfrentamento de desafios estruturais, como a morosidade, a sobrecarga de processos e a escassez de recursos humanos, problemas historicamente enfrentados pelo sistema judicial brasileiro. Entre os principais benefícios apontados pelos autores, ganha ênfase a otimização da gestão processual e administrativa, com a automação de tarefas rotineiras – como classificação de processos, identificação de prazos e triagem de demandas – que permite aos magistrados e servidores concentrarem-se em atividades mais complexas e decisórias. Além disso, a IA contribui significativamente para a celeridade processual, ao facilitar o agrupamento de processos por similaridade, sugerir precedentes e apoiar a elaboração de minutas, como já ocorre em tribunais superiores com sistemas como Victor (STF) e Athos (STJ) (Siqueira et al., 2023).

Outro aspecto importante abordado no artigo é o aumento da produtividade e a redução de custos no Judiciário, especialmente em um contexto de crescente volume de ações judiciais

e limitação orçamentária. A automação racional de tarefas repetitivas permite melhor alocação de recursos públicos e maior eficiência institucional.

Em Barbosa et al. (2023), destaca-se que a principal contribuição da IA para o Judiciário está na agilidade no trâmite processual, na redução do congestionamento, na uniformização de decisões e na eficiência da gestão de informações jurídicas. O texto descreve ferramentas como o sistema JuLIA, do TJPI, e a plataforma Vitória, em desenvolvimento no STF, como exemplos nacionais de uso exitoso da IA na automação de tarefas, análise de dados, intimações automatizadas e agrupamento de processos por similaridade temática. Essas soluções visam, sobretudo, aumentar a celeridade dos processos e liberar os profissionais para tarefas mais complexas, contribuindo para uma prestação jurisdicional mais eficaz.

Observa-se que a IA oferece diversos benefícios concretos ao Poder Judiciário brasileiro, que vão além da automação de tarefas rotineiras. A IA é capaz de identificar padrões em grandes volumes de dados, como jurisprudências e decisões anteriores, contribuindo para a uniformização das decisões judiciais e o aumento da segurança jurídica. Essa tecnologia também facilita o cumprimento do princípio constitucional da razoável duração do processo, promovendo maior celeridade processual por meio da automação de atividades como análise de petições, triagem de processos e designação de audiências de conciliação.

Portanto, pode-se dizer que, sob determinado ponto de vista, a IA não apenas aprimora a eficiência operacional do Judiciário, mas também amplia o acesso à justiça e favorece a construção de um sistema mais transparente e responsivo.

4.2.2 Os riscos e desafios da implementação da inteligência artificial no Poder Judiciário brasileiro

A incorporação da IA ao Poder Judiciário brasileiro tem sido amplamente debatida como uma solução promissora para enfrentar os desafios históricos de morosidade, sobrecarga processual e ineficiência estrutural. No entanto, como destacam Cruz e Souza (2025) e Brunetta et al. (2025), ao lado dos potenciais benefícios, o uso de tecnologias automatizadas no contexto judicial impõe riscos substanciais à proteção dos direitos fundamentais, à equidade processual e à legitimidade das decisões judiciais.

Um dos principais pontos de atenção abordados por Cruz e Souza (2025) refere-se à desumanização do processo decisório, especialmente em situações nas quais a IA opera sem o devido controle humano. Embora ferramentas como o sistema Victor e Sócrates promovam

agilidade ao trâmite processual, existe o risco de que a automação de etapas críticas do julgamento leve à exclusão de elementos subjetivos e humanitários indispensáveis à construção de uma decisão justa.

Como exemplificam as autoras (Cruz; Souza, 2025, p. 13-14), o emblemático caso da “menina do galinheiro” – uma criança em situação extrema de vulnerabilidade cujo contexto foi ignorado em nome de uma aplicação meramente técnica da norma – demonstra que empatia e sensibilidade humana continuam sendo aspectos insubstituíveis na concretização da justiça substantiva. Embora a IA seja bastante eficiente e capaz de processar uma enorme quantidade de dados com celeridade e precisão, ainda “carece da sensibilidade necessária para captar nuances emocionais, culturais e sociais inerentes às disputas judiciais” (Cruz; Souza, 2025, p. 13).

O ato de julgar está profundamente vinculado à dimensão humana e aos seus aspectos sociais, elementos que nenhum sensor ou tecnologia é capaz de captar, pois constituem traços inerentes à própria natureza do ser humano (Lee, 2019). A aptidão para experimentar emoções – como dor, alegria, amor ou ira – é exclusiva da consciência humana (Andrighi; Bianchi, 2020). Ainda que a IA possa desempenhar um papel relevante em demandas jurídicas mais simples, ela não possui a capacidade de lidar plenamente com as complexidades que exigem raciocínio pragmático, sensibilidade emocional e discernimento ético – atributos que continuam sendo exclusivos da atuação humana.

A manutenção do elemento humano no funcionamento do sistema jurídico é uma maneira de garantir que as decisões judiciais transcendam a mera aplicação mecânica da lei, alcançando um patamar de justiça que seja também ético, sensível e atento às complexidades da condição humana. A IA, com sua habilidade para processar grandes volumes de dados e oferecer análises objetivas e tecnicamente precisas, pode desempenhar um papel valioso como ferramenta de apoio. No entanto, somente o discernimento humano é capaz de interpretar as sutilezas emocionais, culturais e sociais que permeiam cada caso, considerando aspectos que escapam à lógica algorítmica. Essa dimensão humana permite equilibrar rigor jurídico com empatia, prevenindo que o direito se afaste de sua finalidade maior: promover uma justiça que seja, ao mesmo tempo, correta na forma e sensível no conteúdo.

Outra preocupação debatida na literatura se chama “decisionismo tecnológico”. Essa situação se reporta ao risco de empobrecimento das decisões judiciais quando há uma parametrização excessiva, resultando em sentenças padronizadas que desconsideram as especificidades de cada caso. Em outras palavras, quando um algoritmo enquadra situações

singulares numa categoria genérica – sem que as partes interessadas tenham ciência disso –, compromete-se o pleno exercício do direito à ampla defesa e ao contraditório, além de dificultar a realização de auditorias sobre as decisões proferidas pelos meios tradicionais (Junquillo, 2022). Questão semelhante é abordada por Brunetta et al. (2025) e Santos (2025), destacando que a “mecanização excessiva das decisões judiciais” pode conduzir a julgamentos padronizados, sem contextualização e a análise subjetiva, deixando-se de lado a sensibilidade e a ponderação de valores.

Além disso, o uso indiscriminado de IA pode comprometer a privacidade dos jurisdicionados. Sistemas automatizados dependem da análise massiva de dados – muitos dos quais são sensíveis – o que levanta preocupações relacionadas à proteção das informações pessoais e aos direitos da personalidade. Quando o cidadão não tem clareza sobre quais dados são usados ou como eles influenciam o julgamento, há risco concreto de violação do devido processo legal e retrocesso em garantias civis historicamente consolidadas (Alves; Andreda, 2021; Cruz; Souza, 2025, p. 12). Vazamentos, uso indevido ou acesso não autorizado a informações sensíveis configuram riscos elevados tanto para indivíduos quanto para instituições.

Um risco significativo está associado à opacidade algorítmica, ou ao fenômeno da “caixa-preta”, no qual nem mesmo os operadores compreendem os critérios utilizados pelos algoritmos para tomar decisões (Alves; Andreda, 2021). A falta de clareza na lógica interna desses sistemas frustra a expectativa humana por respostas objetivas a indagações como “por quê?” e “como você sabe?” (Deeks, 2019).

O uso de sistemas de IA opacos pelo Poder Público levanta preocupações relevantes, sobretudo por seus possíveis reflexos sobre direitos fundamentais, como a liberdade, a propriedade e a privacidade (Bonat et al., 2023). Como destacou o ministro Gilmar Mendes, “vivemos na era das escolhas de Sofia automatizadas” (Brasil, 2020).

Interessante é a comparação feita em Holanda (2025, p. 47-48), para quem uma decisão baseada em IA sem fundamentação clara equivale a uma decisão judicial nula, conforme o Código de Processo Civil (art. 1.103, § 3º, inciso IV) (Brasil, 2015), e que, portanto, advogados, partes e tribunais devem compreender o funcionamento dos algoritmos antes de empregá-los em suas atividades. Sistemas opacos que não atendam aos requisitos legais e éticos de transparência e fundamentação das decisões judiciais podem prejudicar o exercício do contraditório e da ampla defesa (Nunes Pecego, 2024, p. 23). Isso contraria frontalmente o princípio da motivação das decisões judiciais e prejudica a possibilidade de contestação ou

revisão, elementos centrais para um processo justo. Nesse contexto, princípios como justiça, imparcialidade e transparência são imprescindíveis para um ambiente judicial que faz uso responsável e ético da IA.

O problema envolvendo o fenômeno da “caixa-preta” é especialmente analisado em Alves e Andrade (2021). Tal artigo analisa que a implementação de sistemas de IA em áreas sensíveis, como o Poder Judiciário, apresenta riscos e desafios significativos, sobretudo quando se utilizam modelos complexos e opacos baseados em *deep learning* (Alves; Andrade, 2021, p. 351). Um dos riscos identificados pelos autores é a falibilidade técnica: algoritmos podem produzir resultados incorretos devido a dados de entrada incompletos ou mal calibrados, erros na interpretação desses dados ou treinamentos inadequados (Alves; Andrade, 2021, p. 353). Mesmo quando o resultado é “correto”, ele pode decorrer de raciocínios equivocados, como no chamado “efeito Clever Hans”, o que compromete a confiabilidade e a robustez da decisão (Samek; Müller, 2019; Alves; Andrade, 2021, p. 354-355).

A opacidade algorítmica agrava o problema, pois dificulta a compreensão do processo decisório e inviabiliza auditorias adequadas. Enquanto sistemas tradicionais permitem rastrear as regras aplicadas, modelos complexos de aprendizagem de máquina e *deep learning* processam milhões de variáveis e interações, tornando sua lógica interna indecifrável mesmo para especialistas (Alves; Andrade, 2021, p. 359-361). Essa falta de transparência compromete a possibilidade de auditoria, dificulta a responsabilização e enfraquece a confiança dos usuários nos sistemas (Mozetić, 2025, p. 39). A ausência de transparência ou a obscuridade das decisões automatizadas é motivo de questionamento, afetando não apenas os cidadãos, mas também os órgãos responsáveis por sua regulação e supervisão (Doneda et al., 2018).

No contexto judicial, isso é particularmente grave, pois ferramentas como o Victor (STF) e o Athos (STJ) processam milhares de casos em altíssima velocidade, de modo que eventuais erros ou vieses podem se propagar em larga escala, comprometendo direitos fundamentais (Alves; Andrade, 2021, p. 360). Um modelo transparente não exige a divulgação do código-fonte, mas deve permitir entender como e por que o algoritmo chegou ao resultado, inclusive para avaliar responsabilidades em caso de erro (Holanda, 2025, p. 47).

Vale ressaltar que a relação entre confiança e IA é explorada no *artigo Why Should I Trust You? Explaining the Predictions of Any Classifier*, onde os autores propõem métodos para aumentar a confiança humana em modelos de aprendizagem de máquina por meio de explicações claras e interpretáveis. Argumenta-se que confiar em uma previsão individual (*trusting a prediction*) e confiar no modelo como um todo (*trusting the model*) dependem de

entender o raciocínio do modelo, o que não é possível tratando-o como uma *black box* (Ribeiro et al. 2016).

Outro ponto crítico é a ocorrência de vieses algorítmicos, quando sistemas reproduzem preconceitos de raça, gênero ou classe social já presentes nos dados de treinamento, tal como já apontado nesta tese (veja subcapítulo “2.3.2 A inteligência artificial e o mito da neutralidade: vieses nos dados e nos algoritmos”). A presente tese serve de alerta para a possibilidade de reprodução e amplificação de vieses estruturais – de gênero, raça ou classe – caso os dados de treinamento estejam contaminados por preconceitos históricos e a equipe técnica responsável não tenha diversidade suficiente para reconhecê-los.

Sistemas de IA podem reproduzir ou amplificar preconceitos presentes nos dados de treinamento, resultando em discriminação de gênero, raça, classe social ou outros marcadores sociais. Esses vieses nem sempre são evidentes, mas seus impactos podem ser significativos e afetar diretamente a equidade das decisões automatizadas (Mozetić, 2025, p. 27-28).

Além disso, o desafio técnico ligado à qualidade dos dados utilizados para treinamento, à capacidade de generalização dos modelos e à confiabilidade de seus resultados é ponto determinante para a implementação ética da IA. Bases de dados incompletas, enviesadas ou desatualizadas podem comprometer a acurácia e gerar distorções, o que é particularmente grave em setores sensíveis como o jurídico (Mozetić, 2025, p. 27-28).

Também é necessário chamar a atenção para um paradoxo envolvendo a implementação de IA no Poder Público em geral: embora a IA possa ampliar o acesso à justiça por meio de plataformas digitais e interfaces automatizadas, ela também pode aprofundar a exclusão digital, especialmente entre cidadãos com baixo nível de letramento digital ou sem acesso adequado à internet (Britto et al., 2023). Ferramentas como os *Online Dispute Resolution Systems* (ODR), se não forem adaptadas à realidade brasileira, arriscam substituir barreiras geográficas por barreiras tecnológicas.

Outro risco existente está na falta de regulamentação clara sobre a responsabilidade por decisões automatizadas. Em caso de erro, não está definido quem responderia – se o juiz, o desenvolvedor, o tribunal ou o sistema em si. Esse é um debate que envolve tanto a análise de falhas técnicas quanto a avaliação de usos inadequados. Nesse ponto, Cruz e Souza (2025, p. 12) citam debates internacionais, como a proposta de Barfield (2018), que sugere atribuir personalidade jurídica a sistemas inteligentes – uma solução que evidencia os dilemas éticos e jurídicos ainda não resolvidos sobre o uso da IA no Judiciário.

A implementação da IA também encontra barreiras organizacionais e culturais. A sua adoção requer mudanças estruturais, capacitação de equipes e adequação de processos internos. Resistências institucionais, aliadas à falta de compreensão sobre o funcionamento e as limitações da tecnologia, podem dificultar sua integração efetiva (Mozetić, 2025, p. 34).

Ainda quanto à análise dos riscos envolvendo a IA no Poder Judiciário, em 2024 foi elaborado o relatório de pesquisa “Mapeando riscos da IA no Poder Judiciário brasileiro” (Caballero et al., 2024). O relatório analisou 140 projetos de IA cadastrados na Plataforma Sinapses do CNJ em 2023, com base na minuta de alteração da Resolução CNJ n. 332/2020, que adota uma abordagem regulatória baseada em níveis de risco (Conselho Nacional de Justiça, 2020a; Caballero et al., 2024, p. 10-14).

O relatório levou em consideração o texto da minuta apresentada pelo Grupo de Trabalho sobre IA no Poder Judiciário do CNJ, grupo este que foi instituído pela Portaria Presidência n. 338, de 30 de novembro de 2023, para atualizar a Resolução n. 332/2020 (Conselho Nacional de Justiça, 2020b). A classificação foi feita a partir das descrições dos projetos registradas na própria plataforma, utilizando as hipóteses previstas nos artigos 7º-B (risco excessivo), 7º-C (alto risco) e 7º-D (baixo risco) da minuta (Caballero et al., 2024, p. 14-16).

Na classificação adotada, como baixo risco (art. 7º-D) enquadram-se sistemas destinados a tarefas administrativas e de apoio, como execução de atos processuais ordinatórios, sumarização e transcrição de documentos, relatórios gerenciais sem substituição da avaliação humana, anonimização de dados e atividades similares. Confira-se:

Art. 7º-D. Consideram-se de baixo risco as soluções que utilizem técnicas de inteligência artificial desenvolvidas e utilizadas para as seguintes finalidades e contextos:

I – execução de atos processuais ordinatórios ou de tarefas de apoio à administração judiciária, mediante a extração de informações de sistemas e de documentos, com a finalidade de classificação e agrupamento de dados e de processos, de enriquecimento de cadastros, de certificação e transcrição de atos processuais, de sumarização de documentos, entre outras finalidades de gestão processual e operacional;

II – detecção de padrões decisórios ou de desvios de padrões decisórios, observado o caráter complementar da técnica de inteligência artificial, desde que não haja substituição de avaliação humana sobre processos e que seja destinado para uso interno do tribunal e de uniformização da jurisprudência;

III – fornecimento aos magistrados de subsídios para a tomada de decisão mediante relatórios gerenciais e análises que adotem técnica jurimétrica, com a integração de fontes de informação relevantes ou a detecção de padrões decisórios, desde que não haja substituição da avaliação humana e desde que a solução não realize valorações de cunho moral sobre provas ou sobre perfis e condutas de pessoas;

IV – produção de textos de apoio para facilitar a confecção de atos judiciais, desde que a supervisão e a versão final do documento sejam realizadas pelo magistrado, bem como as decisões acerca das preliminares e questões de mérito;

V – aprimoramento ou a formatação de uma atividade humana anteriormente concluída, desde que não se altere materialmente o seu resultado, ou a realização de uma tarefa preparatória para uma tarefa considerada como de alto risco
 VI – realização de análises estatísticas para fins de política judiciária;
 VII – transcrição de áudio e vídeo para o auxílio das atividades do magistrado;
 VIII – anonimização de documento ou de sua exibição, especialmente para a proteção de dados pessoais. (NR) (Caballero et al., 2024, p. 12-13).

No alto risco (art. 7º-C) estão as soluções que envolvem formulação de juízos conclusivos sobre a aplicação da norma jurídica, avaliação de provas, apoio a perícias e previsão de resultados processuais. Veja-se:

Art. 7º-C. Consideram-se de alto risco as soluções que utilizem técnicas de inteligência artificial desenvolvidas e utilizadas para as seguintes finalidades e contextos: (NR)

I – a identificação de perfis e de padrões comportamentais de pessoas naturais ou de grupos de pessoas naturais, em situações que não se encontram enquadradas como de risco excessivo

II – a avaliação da credibilidade e a valoração de elementos de provas em seus mais diversos tipos (documentais, testemunhais, periciais e outras), em ações criminais ou não;

III – a averiguação, a valoração e a interpretação de fatos voltados para a aplicação da norma jurídica, especialmente quando envolva a restrição de liberdades individuais, em ações criminais ou não;

IV – a formulação de juízos conclusivos sobre a aplicação da norma jurídica ou precedentes a um conjunto determinado de fatos concretos, inclusive para a quantificação ou a qualificação de danos suportados por pessoas ou grupos, em ações criminais ou não;

V – a identificação e a autenticação biométrica para o monitoramento de comportamento de pessoas naturais, excluindo-se a autenticação biométrica cujo único objetivo seja a confirmação da identidade de uma pessoa natural específica.’ (NR) (Caballero et al., 2024, p. 12).

Já o risco excessivo (art. 7º-B) abrange ferramentas que avaliem traços de personalidade ou comportamentos para prever crimes ou reincidência, que classifiquem pessoas com base em atributos comportamentais ou que realizem reconhecimento de emoções, estando vedadas ao Poder Judiciário (Caballero et al. 2024, p. 11-13). Observe-se:

Art. 7º-B. São vedados ao Poder Judiciário, por acarretarem risco excessivo à segurança da informação, aos direitos fundamentais dos cidadãos e à independência dos magistrados, o desenvolvimento e a utilização de soluções:

I – que não possibilitem a revisão humana dos dados utilizados e dos resultados propostos ao longo de seu ciclo de treinamento, de desenvolvimento e de uso, ou que gerem vinculação absoluta do usuário ao resultado proposto;

II – que avaliem traços da personalidade, características ou comportamentos de pessoas naturais ou de grupos de pessoas naturais, para fins de avaliar o cometimento de crimes ou a probabilidade de reiteração delitiva;

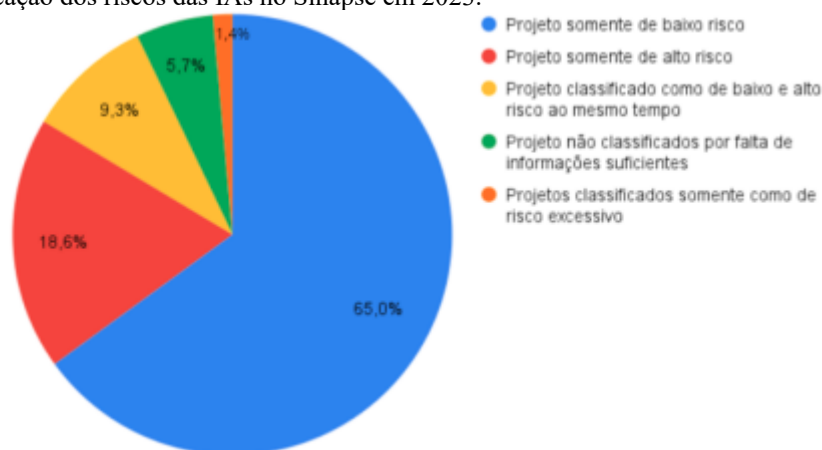
III – que classifiquem ou ranqueiem pessoas naturais, com base no seu comportamento social ou em atributos da sua personalidade, para a avaliação da plausibilidade dos direitos por elas vindicados.

IV – que envolvam a identificação e a autenticação de padrões biométricos para o reconhecimento de emoções.’ (NR) (Caballero et al., 2024, p. 11-12).

A metodologia consistiu na leitura das descrições e no enquadramento por consenso da equipe de pesquisa. Critérios específicos foram definidos: *chatbots* e sistemas de triagem processual foram classificados como baixo risco; ferramentas para elaboração de minutas de atos jurisdicionais, apoio a perícias ou previsão de conciliação foram consideradas de alto risco; e dois casos específicos – “Sistema de Combate à Desinformação nas Eleições” (Tribunal Regional Eleitoral de Pernambuco – TRE-PE) e “Predição de risco de revitimização em violência doméstica” (Tribunal de Justiça do Ceará – TJCE) – foram enquadrados como risco excessivo (Caballero et al., 2024, p. 15-16).

Dos 140 projetos identificados, oito não continham informações suficientes para classificação (Caballero et al., 2024, p. 18). Entre os demais (132 projetos), 91 foram classificados como de baixo risco – por exemplo, o projeto “Transcrição de vídeos de audiências” (Tribunal Regional do Trabalho da 17ª Região – TRT17), que gera transcrições a partir de gravações (Caballero et al., 2024, p. 18-19). Outros 26 foram classificados como de alto risco – como o “Intelligentia” (TRF2), que identifica temas aplicáveis a demandas repetitivas e formula sugestões de enquadramento jurídico (Caballero et al., 2024, p. 19). Apenas 2 foram considerados de risco excessivo, ambos já mencionados (Caballero et al., 2024, p. 19-20). Alguns projetos possuíam múltiplas funcionalidades; nesses casos, o nível final foi definido pelo maior risco identificado (Caballero et al., 2024, p. 23). A Figura 32, retirada do relatório, ilustra como ficou a classificação de riscos das IAs inseridas no Sinapses.

Figura 32. Classificação dos riscos das IAs no Sinapse em 2023.



Fonte: Caballero et al., 2024, p. 20.

Embora o relatório identifique que a maioria dos projetos de IA em 2023 eram de baixo risco, um dos apontamentos feitos foi de que os projetos na plataforma Sinapses precisam de mais detalhamentos, identificando-se um problema de transparência quanto às funcionalidades

e aos dados usados por algumas IAs (Caballero et al., 2024, p. 23-24). Assim, é imperioso o acompanhamento de cada um dos projetos em trâmite a fim de que se avaliem os riscos envolvidos na implementação de tais sistemas no Poder Judiciário.

Diante do exposto, fica clara a necessidade de marcos regulatórios claros e políticas de governança robustas que orientem o desenvolvimento, a implementação e a supervisão da IA no Poder Judiciário brasileiro, assegurando que ela seja utilizada de acordo com princípios de legalidade, equidade, transparência e segurança. Em síntese, a adoção da IA não se limita à aquisição de tecnologia; trata-se de um processo que requer uma abordagem multidimensional, articulando aspectos técnicos, humanos, jurídicos e éticos, com ênfase na supervisão contínua e na regulamentação específica para mitigar riscos e potencializar benefícios.

CAPÍTULO 5 – GARANTIAS ÉTICAS PARA O ENFRENTAMENTO DOS VIESES DE GÊNERO EM SISTEMAS INTELIGENTES APLICADOS NO JUDICIÁRIO

A incorporação de sistemas de IA no Poder Judiciário brasileiro apresenta-se como um fenômeno inevitável no contexto da transformação digital da administração pública e da busca por eficiência processual. No entanto, como demonstrado nos capítulos anteriores, o uso dessas tecnologias no campo jurídico não se limita a um avanço técnico: trata-se de uma mudança estrutural que envolve dimensões institucionais, epistemológicas e éticas. A IA, ao operar sobre dados e regras de decisão, pode tanto contribuir para a modernização da justiça quanto reproduzir e amplificar desigualdades pré-existentes, em especial aquelas relacionadas ao gênero.

Os capítulos precedentes evidenciaram que os algoritmos e sistemas inteligentes são construções humanas, moldadas por escolhas metodológicas, culturais e políticas. Assim, o problema dos vieses algorítmicos é, antes de tudo, uma questão social que se manifesta tecnicamente. Nesse contexto, a ética se torna um elemento estruturante do debate sobre a legitimidade e a responsabilidade do uso da IA no sistema judicial, exigindo o estabelecimento de protocolos, diretrizes e mecanismos de governança que garantam decisões automatizadas transparentes, auditáveis e alinhadas aos direitos humanos e aos princípios constitucionais.

A implementação da IA enfrenta um conjunto complexo de desafios técnicos, jurídicos, éticos e organizacionais, que precisam ser adequadamente tratados para que seu uso seja seguro, eficaz e socialmente legítimo. Essa é uma questão que exige vigilância crítica quanto às suas implicações sociais.

Como exposto por Monteiro (2021, p. 217), algoritmos realizam tarefas de classificação, inferência e, em alguns casos, de tomada de decisão, mas frequentemente operam em formato *black box* – ou seja, sem que o processo de raciocínio seja plenamente explicável ou compreensível para humanos. Muitos desses sistemas são compostos por camadas complexas de operações automáticas, resultando em modelos de difícil interpretação, o que compromete a transparência e a possibilidade de questionamento de suas decisões.

Sistemas de IA são capazes de executar tarefas que normalmente exigiriam reflexão humana, e embora simulem certas formas de raciocínio e resposta, sua lógica interna permanece opaca. Enquanto nas relações jurídicas entre pessoas há mecanismos consolidados de responsabilização, baseados em doutrina e jurisprudência, nas decisões automatizadas ainda

não existe um sistema plenamente desenvolvido de atribuição de responsabilidade ou revisão crítica dos resultados gerados por algoritmos (Monteiro, 2021).

Para elucidar os resultados de tais sistemas, torna-se necessário desenvolver metodologias explicativas que forneçam *insights* sobre o processamento interno dos modelos e sobre as relações entre variáveis e critérios de desempenho. Contudo, como observa a literatura especializada, essas metodologias representam apenas aproximações parciais do que o modelo realmente realiza, reforçando a urgência de mecanismos de explicabilidade e governança ética que sustentem o uso da IA em contextos sensíveis como o judicial.

É importante reconhecer que um sistema de IA é tão confiável quanto a programação de seus algoritmos e a qualidade de seus dados. Em plataformas voltadas ao sistema judicial, essa relação torna-se ainda mais crítica: a confiabilidade técnica e a legitimidade social do sistema dependem diretamente da observância de princípios éticos desde sua concepção. Quando algoritmos são mal elaborados, ou quando as bases de dados utilizadas apresentam vieses estruturais, o resultado pode ser a produção de decisões substancialmente injustas, tecnicamente consistentes, mas socialmente discriminatórias.

Desse modo, se não forem observadas regras éticas e critérios de qualidade na elaboração e no treinamento dos modelos, a IA pode entregar resultados tendenciosos, reforçando preconceitos de gênero, raça ou classe social – preconceitos que ela herda de seus programadores e das bases de dados que refletem desigualdades históricas. A justiça automatizada, portanto, não está imune aos mesmos condicionantes sociais que moldam o comportamento humano: sem governança ética, ela apenas transfere o *locus* do viés, sem eliminá-lo.

Uma IA voltada ao sistema de justiça deve ser confiável sob duas perspectivas complementares: a tecnológica, relacionada à forma como o sistema é projetado, testado e utilizado; e a normativa, que diz respeito às regras, leis e princípios que regem sua aplicação. Essa dupla dimensão de confiabilidade exige padrões éticos de *design*, desenvolvimento, implementação e uso, orientados para a promoção do bem-estar humano e o fortalecimento da democracia e da igualdade (High-Level Expert Group on Artificial Intelligence, 2018, p. 1-2).

Assim, resolver problemas de IA envolvendo vieses requer mais do que ajustes técnicos: implica repensar o modo como se constroem e governam sistemas inteligentes, garantindo que princípios éticos e valores sociais estejam incorporados em todas as fases – desde a formulação do problema e a curadoria de dados até a validação e o monitoramento dos resultados.

A atual conjuntura de desenvolvimento e aplicação da IA no Poder Judiciário brasileiro gera preocupações quando se tem em mente a necessidade de garantir transparência e ética. Isso é facilmente observável com o “Painel de Pesquisa sobre Inteligência Artificial 2024” (Conselho Nacional de Justiça, 2025b). Somente 16% dos respondentes à questão sobre adoção de medidas assecuratórias para garantir a transparência e a ética em projetos de IA em desenvolvimento e uso afirmaram estabelecer diretrizes e políticas claras sobre o seu uso, 14,4% asseveraram efetuar treinamentos e capacitação de servidores e magistrados sobre o uso ético da IA, 4,9% indicaram implementar mecanismos de explicabilidade dos resultados obtidos pelos sistemas de IA, 3,0% realizam auditorias periódicas nos algoritmos e modelos de IA e 0,5% criaram um comitê multidisciplinar para o estabelecimento de diretrizes, políticas claras sobre o uso de IA, capacitação e treinamentos. Além disso, 22,8% afirmaram que sequer adotam medidas a fim de garantir transparência e ética nesta seara.

Nesse sentido, com o intuito de promover o desenvolvimento e o uso ético (e feminista) da IA no Judiciário brasileiro, este capítulo sistematiza as garantias éticas necessárias ao enfrentamento dos vieses de gênero em sistemas inteligentes aplicados à justiça, organizadas em três subcategorias analíticas interdependentes: técnicas, regulatórias e políticas. As garantias técnicas (subcapítulo 5.1) referem-se à estruturação dos dados, à modelagem algorítmica e aos mecanismos de equidade, transparência, explicabilidade e auditoria, buscando assegurar confiabilidade e rastreabilidade nas decisões automatizadas. As garantias regulatórias (subcapítulo 5.2) abrangem os marcos normativos e institucionais voltados à governança da IA, definindo parâmetros de responsabilidade, transparência, controle social e garantia a direitos humanos e princípios constitucionais como o do devido processo legal. Já as garantias políticas (subcapítulo 5.3) dizem respeito à democratização do desenvolvimento e da aplicação da IA, à diversidade na formulação das políticas públicas e à promoção da conscientização ética no uso da IA e do letramento digital. Essas três dimensões não constituem esferas autônomas, mas camadas interdependentes de um mesmo sistema ético, em que a coerência técnica depende de fundamentos regulatórios sólidos e de uma orientação política inclusiva.

A proposição formulada nesta tese busca dialogar diretamente com o CNJ, oferecendo subsídios conceituais e metodológicos para o aprimoramento da governança da IA no Poder Judiciário. Trata-se de uma proposta de caráter afirmativo, ao compreender a IA não apenas como ferramenta de automação e eficiência, mas como instrumento potencial de mitigação da sub-representação feminina nas instâncias decisórias do sistema judicial. Ao propor diretrizes éticas e políticas integradas, a tese reconhece que a consolidação de uma governança

algorítmica justa e inclusiva exige o enfrentamento das assimetrias de gênero reveladas pelos dados sociodemográficos analisados, de modo a reconfigurar a própria cultura institucional do Judiciário sob a perspectiva da igualdade substantiva.

A partir dessa base, as seções seguintes aprofundam a reflexão sobre os fundamentos éticos da IA no contexto judicial e sobre as iniciativas internacionais e nacionais voltadas à construção de uma IA confiável, justa e inclusiva, compatível com os princípios constitucionais e com os compromissos de igualdade de gênero assumidos pelo Estado brasileiro.

5.1 Garantias técnicas para uma inteligência artificial ética e feminista no Poder Judiciário brasileiro

As garantias técnicas se referem à estruturação dos dados, à modelagem algorítmica e aos mecanismos de explicabilidade e auditoria, constituindo o núcleo operativo da confiabilidade dos sistemas inteligentes aplicados à justiça. Volta-se aqui a atenção para as garantias éticas necessárias à implementação de sistemas de IA com uma perspectiva de gênero no Poder Judiciário, delineando-se um conjunto de medidas técnicas voltadas à eliminação de vieses de gênero nos processos decisórios automatizados.

Essas garantias abrangem dimensões fundamentais como *fairness*, transparência, explicabilidade e robustez dos sistemas, reconhecendo que a mitigação dos vieses algorítmicos depende tanto da integridade técnica das soluções quanto de seu alinhamento a princípios éticos e de equidade. A *fairness* (que pode ser traduzida para justiça ou equidade algorítmica) diz respeito à capacidade dos sistemas de tratar diferentes grupos sociais de maneira justa, evitando que atributos como gênero, raça ou classe influenciem desproporcionalmente os resultados. A transparência e a explicabilidade, por sua vez, relacionam-se à possibilidade de compreender, documentar e reconstruir o processo de decisão automatizada, tornando-o acessível e auditável tanto por especialistas quanto pelos usuários do sistema de justiça.

Em primeiro lugar, essas garantias implicam atenção à qualidade, integridade e representatividade das bases de dados que alimentam os modelos, de modo a prevenir distorções resultantes de amostras incompletas, enviesadas ou marcadas por estereótipos de gênero. A curadoria de dados deve obedecer a critérios de diversidade, anonimização e governança responsável, assegurando que os conjuntos de treinamento reflitam a pluralidade social e não reproduzam desigualdades históricas contidas em registros administrativos, jurisprudências ou textos legais. Nesse sentido, a governança de dados no Poder Judiciário emerge como um eixo

estruturante dessas garantias, pois define padrões de coleta, armazenamento e uso de informações que influenciam diretamente o comportamento dos sistemas inteligentes.

No que se refere à modelagem algorítmica, as garantias técnicas demandam o uso de metodologias de desenvolvimento que incorporem, desde a concepção, princípios de transparência, verificabilidade e controle de vies. Isso inclui a documentação rigorosa de todo o ciclo de vida do modelo – desde o treinamento até a implantação –, o registro de parâmetros e versões e a realização de auditorias algorítmicas periódicas capazes de identificar e corrigir distorções discriminatórias.

Os mecanismos de explicabilidade e rastreabilidade completam o conjunto dessas garantias, ao permitir que o caminho lógico percorrido por um sistema para chegar a determinada decisão possa ser compreendido e reconstruído. A explicabilidade técnica – por meio de abordagens de *explainable AI* (XAI), *model cards* ou análise de importância de variáveis – deve ser acompanhada por uma explicabilidade institucional, voltada a traduzir os resultados algorítmicos em linguagem acessível aos operadores do Direito e à sociedade. Tal requisito está diretamente vinculado à necessidade de *design* acessível e à responsabilização pelos resultados produzidos por sistemas automatizados.

Assim, as garantias técnicas não se restringem à dimensão operacional da IA, mas representam um pilar ético essencial para a confiabilidade, a rastreabilidade e a legitimidade das decisões automatizadas no contexto judicial. Quando orientadas por uma perspectiva de gênero, essas garantias funcionam como instrumentos de prevenção ativa de vieses discriminatórios e como base para a construção de sistemas de IA mais justos, transparentes e inclusivos, coerentes com os princípios constitucionais de igualdade e dignidade da pessoa humana.

5.1.1 Métricas de *fairness* (justiça ou equidade algorítmica)

A *fairness*, justiça ou equidade algorítmica, ocupa posição central nas discussões sobre ética e governança da IA. Representa o compromisso de que as decisões automatizadas sejam produzidas de modo não discriminatório, garantindo tratamento equânime entre indivíduos e grupos, independentemente de atributos sensíveis como gênero, raça, idade, classe social ou orientação sexual (Fairlearn, [20–]; Forbes Technology Council, 2024; Fit Instituto de Tecnologia, 2025).

O grau de justiça de um algoritmo é um dos três elementos que integram o debate internacional sobre ética em IA, ao lado de *accountability* (responsabilização) e *transparency* (transparência). Um algoritmo é considerado justo quando seus resultados são independentes de determinadas variáveis sensíveis, como gênero, raça ou orientação sexual (Salomão, 2022). Assim, a *fairness* busca identificar se o modelo é tendencioso nas categorizações de classes minoritárias dentro dos dados analisados, funcionando como um instrumento de diagnóstico e mitigação de vieses sistêmicos (Salomão, 2022, p. 32). Autores como Freitas (2025, p. 105), asseveram a importância de se pensar em uma IA com base em equidade, critérios éticos e legal a fim de afastar preconceitos e enviesamentos.

Sob a ótica técnico-jurídica, a *fairness* articula duas dimensões complementares: o plano ético-normativo da igualdade e o plano técnico-operacional dos algoritmos. Em outras palavras, constitui o ponto de convergência entre o dever jurídico de isonomia e as práticas de engenharia e governança de dados. No contexto judicial, essa relação é decisiva, pois as decisões derivadas de sistemas automatizados impactam diretamente direitos fundamentais e podem reproduzir padrões de desigualdade estruturais – notadamente os vieses de gênero historicamente presentes no sistema de justiça (Biondi, 2020; Monteiro, 2021).

A adoção de técnicas e métricas de *fairness* desempenha papel substancial na mitigação de vieses de gênero em decisões judiciais automatizadas, uma vez que permite mensurar, de modo objetivo, se modelos de IA tratam homens e mulheres de forma equitativa em contextos decisórios equivalentes. Ao monitorar diferenças nas taxas de acertos, erros e resultados favoráveis entre grupos sensíveis, tais técnicas possibilitam identificar assimetrias estruturais que, de outro modo, permaneceriam ocultas nos dados históricos e nos padrões algorítmicos. No ambiente do Poder Judiciário, essa prática assume caráter ético e institucional, pois assegura que o uso da IA não reforce estereótipos de gênero, mas contribua para o fortalecimento do princípio da isonomia e da imparcialidade nas decisões.

Além disso, a aplicação sistemática de técnicas de *fairness* permite que os tribunais estabeleçam políticas de auditoria e validação contínua dos modelos, integrando parâmetros de equidade de gênero às rotinas de governança tecnológica. Essa integração torna possível o desenvolvimento de sistemas de IA mais justos e explicáveis, em consonância com os direitos fundamentais e com as diretrizes da Resolução CNJ n. 615/2025, que enfatiza a transparência, a diversidade e a responsabilidade social no uso da tecnologia judicial (Conselho Nacional de Justiça, 2025c). Desse modo, a *fairness* opera não apenas como instrumento técnico de correção

de vieses, mas como mecanismo de promoção ativa da igualdade de gênero e da confiança pública nas decisões automatizadas.

À luz da literatura recente, a justiça algorítmica compreende diferentes noções e famílias de métricas (Hardt et al., 2016; Semenova, 2019; Garg et al, 2020; Dhabliya et al., 2024). Segundo autores como Mehrabi et al. (2021) e Tang et al. (2023), existem noções associativas, que avaliam correlações estatísticas entre atributos sensíveis e os resultados do modelo (como paridade demográfica, igualdade de *odds* e paridade de previsão); e noções causais, que exigem modelagem explícita das relações de causa e efeito, incluindo abordagens como *fairness* contrafactual e *fairness* de caminho específico (*path-specific*). Também se distinguem a *fairness* de grupo (comparação entre grupos sensíveis) e a *fairness* individual (tratamento semelhante para indivíduos similares), além de classificarem as estratégias de mitigação de viés segundo o estágio do pipeline: pré-processamento, in-processamento e pós-processamento.

Do ponto de vista técnico, a *fairness* pode ser mensurada e monitorada por meio de métricas específicas que permitem identificar, comparar e corrigir desigualdades no desempenho de um modelo entre diferentes grupos sensíveis. Ferramentas como o *Fairlearn* e relatórios de governança de IA enfatizam que essas métricas devem ser utilizadas de forma contextualizada e complementar, pois nenhuma isoladamente é capaz de eliminar integralmente o viés (Fairlearn, [20–]; Forbes Technology Council, 2024).

As principais métricas empregadas na literatura e aplicáveis ao Poder Judiciário são paridade demográfica (*demographic parity* ou *statistical parity*), igualdade de oportunidades (*equal opportunity*), probabilidades igualadas (*equalized odds*) e paridade de previsão (*predictive parity*).

Formalmente, considerando que:

$\hat{Y}$ é a previsão do modelo A é o atributo sensível (por exemplo, gênero) Y é o resultado real 0 e 1 são valores diferentes para o atributo sensível $\forall y \in \{0,1\}$ representa a classe positiva e negativa
--

a) Paridade demográfica (*Demographic Parity* ou *Statistical Parity*)

A paridade demográfica exige que a probabilidade de um resultado favorável do modelo seja independente da pertença a um grupo sensível, como o gênero (Semenova, 2019, p. 12; Mehrabi et al., 2021, p. 12; Forbes Technology Council, 2024). Formalmente:

$$P(\hat{Y} = 1 \mid A = 0) = P(\hat{Y} = 1 \mid A = 1)$$

Essa expressão significa que a probabilidade do modelo produzir um resultado positivo ($\hat{Y} = 1$) deve ser a mesma para todos os grupos sensíveis (por exemplo, $A = 0$ = homens; $A = 1$ = mulheres).

Em outras palavras, a paridade demográfica exige que a proporção de decisões favoráveis (como concessão de um benefício, aprovação de crédito ou deferimento de liberdade provisória) não dependa da categoria a que o indivíduo pertence. Ela busca igualar os resultados positivos entre grupos, independentemente de suas características sensíveis.

Quanto à implicação judicial, por exemplo, em um sistema de recomendação de medidas cautelares, a probabilidade de receber uma decisão favorável (como liberdade provisória) deve ser a mesma para homens e mulheres, independentemente das diferenças observadas nos dados históricos. Se o modelo prevê decisões favoráveis para 70% dos homens e apenas 50% das mulheres, ele viola a paridade demográfica, pois o resultado está associado ao gênero.

Essa métrica, portanto, mede equidade de resultados – isto é, se o sistema distribui “oportunidades” ou “benefícios” de forma proporcional entre os grupos.

b) Igualdade de oportunidades (*Equal Opportunity*)

Garante que a taxa de verdadeiros positivos (TPR) seja equivalente entre grupos sensíveis, isto é, que indivíduos de diferentes grupos que realmente merecem um resultado favorável tenham igual probabilidade de obtê-lo (Semenova, 2019, p. 12; Mehrabi et al., 2021, p. 12; Dhabliya et al., 2024, p. 7; Fairlearn, [20–]):

$$P(\hat{Y} = 1 \mid Y = 1, A = 0) = P(\hat{Y} = 1 \mid Y = 1, A = 1)$$

Entre todos os indivíduos que realmente pertencem à classe positiva ($Y=1$) – ou seja, que merecem um resultado favorável – a probabilidade de o modelo prever corretamente esse

resultado ($\hat{Y} = 1$) deve ser a mesma para todos os grupos sensíveis (representados por $A = 0$ e $A = 1$). Corresponde à igualdade de (TPR) entre grupos.

Em outras palavras, ela é uma métrica de *fairness* baseada em acertos (verdadeiros positivos). Ela busca garantir que os indivíduos de diferentes grupos sociais tenham as mesmas oportunidades de obter um resultado positivo quando realmente o merecem. Isso significa que o modelo não pode acertar mais frequentemente para um grupo do que para outro, quando ambos têm direito ao mesmo resultado real.

Um exemplo de implicação judicial seria em um sistema de apoio à concessão de benefícios ou de liberdade provisória, em que a igualdade de oportunidades assegura que mulheres não tenham menor chance de decisão favorável mesmo em condições idênticas às dos homens, promovendo efetividade ao princípio da isonomia.

c) Probabilidades igualadas (*Equalized Odds*)

Exige simultaneamente que as taxas de verdadeiros positivos (TPR) e falsos positivos (FPR) sejam equivalentes entre grupos, garantindo equilíbrio na precisão e nos erros do modelo (Hardt et al., 2016; Semenova, 2019, p. 12; Mehrabi et al., 2021, p. 12; Forbes Technology Council, 2024). Formalmente:

$$P(\hat{Y} = 1 \mid Y = y, A = 0) = P(\hat{Y} = 1 \mid Y = y, A = 1), \forall y \in \{0,1\}$$

A métrica *equalized odds* exige que o modelo trate igualmente os grupos sensíveis em termos de seus acertos e erros. Isso implica duas condições simultâneas: As pessoas que realmente pertencem à classe positiva (p.e.: têm direito a um benefício judicial) devem ter a mesma chance de serem corretamente identificadas como positivas, independentemente de pertencerem ao grupo $A=0$ (p.e.: homens) ou $A=1$ (p.e.: mulheres). As pessoas que não pertencem à classe positiva (p.e.: não têm direito ao benefício) devem ter a mesma chance de serem classificadas erroneamente como positivas em ambos os grupos.

O objetivo é garantir equidade de desempenho do modelo entre grupos, o que vai além de simplesmente igualar resultados finais (como faz a *demographic parity*). Ela busca que a qualidade da decisão (acurácia e erro) seja a mesma para todos os grupos – evitando que um grupo seja sistematicamente favorecido ou prejudicado.

No sistema judicial, por exemplo, essa métrica asseguraria que o sistema não cometa mais erros injustificados contra um grupo específico. Ela evitaria que mulheres fossem super-representadas em previsões negativas indevidas, fortalecendo a legitimidade do uso da IA.

d) Paridade de previsão (*Predictive Parity*)

Garante que a precisão das previsões positivas seja igual entre grupos sensíveis (como gênero, raça, idade etc.), ou seja, que a probabilidade de um resultado ser correto, dado que foi previsto como positivo, seja a mesma (DiCiccio et al., 2023). Formalmente:

$$P(Y = 1 \mid \hat{Y} = 1, A = 0) = P(Y = 1 \mid \hat{Y} = 1, A = 1)$$

Isso significa que, entre todos os casos que o modelo previu como positivos, a proporção de acertos (verdadeiros positivos) deve ser a mesma para todos os grupos. Em termos práticos: se o modelo prevê que uma pessoa tem alto risco de reincidência criminal ou merece um benefício judicial, essa previsão deve ser igualmente correta para homens e mulheres.

No Poder Judiciário, por exemplo, a métrica de paridade de previsão poderia, em sistemas de previsão de reincidência ou de análise de risco, assegurar que as decisões automatizadas sejam igualmente confiáveis para homens e mulheres, evitando discriminações indiretas na execução penal.

Essa diversidade de formulações implica que não existe uma definição universal de justiça em IA (Mehrabi et al., 2021). De fato, como observam Garg et al. (2020) e Tang et al. (2023), certas métricas podem ser mutuamente incompatíveis – por exemplo, satisfazer paridade demográfica pode violar igualdade de oportunidades. Assim, a escolha de qual métrica adotar envolve *trade-offs* técnicos e normativos, devendo levar em conta os objetivos da aplicação, os marcos legais e o impacto social de longo prazo.

Semelhante ao que foi demonstrado nesta tese, Mehrabi et al. (2021, p. 4-6), asseveram que os vieses presentes em sistemas de IA derivam de múltiplas fontes, que se manifestam em diferentes etapas do ciclo de vida dos modelos. Entre as principais, os autores identificam vieses históricos, de amostragem, de representação, de medição de variáveis, de modelagem e de avaliação.

Essas categorias demonstram que a injustiça algorítmica não é um fenômeno isolado em um único estágio, mas uma consequência cumulativa de decisões técnicas, sociais e organizacionais que moldam todo o sistema. Mehrabi et al. (2021) defendem, portanto, que a

busca por *fairness* deve abranger todo o ciclo de desenvolvimento da IA, envolvendo etapas de detecção, mitigação e avaliação de viés, com participação interdisciplinar de especialistas, juristas e sociedade civil.

No entanto, é preciso ter em mente que a *fairness* não é um conceito neutro nem plenamente mensurável. Diversos autores argumentam que a justiça algorítmica não pode ser reduzida a uma definição matemática, mas deve ser entendida como uma questão política e social (Foulds et al., 2019; Hutchinson e Mitchell, 2019). Ainda que as métricas de *fairness* e os processos de mitigação técnica de viés – como auditorias algorítmicas – sejam fundamentais, eles não bastam para eliminar a discriminação de gênero na IA. A escolha de qual métrica aplicar é também uma decisão normativa e política, e não apenas técnica.

Estudos apontam que a mitigação técnica de vieses é necessária, mas insuficiente, pois a IA está inserida em contextos institucionais que produzem e reproduzem desigualdades. Assim, a justiça algorítmica deve ser entendida não como um estado final de neutralidade, mas como um campo de disputa ética e social, no qual as definições de “justiça” e “imparcialidade” refletem valores, estruturas de poder e contextos culturais (Mehrabi et al. 2021). A aplicação de métricas deve ser contextualizada e acompanhada por análise dos efeitos de longo prazo, dos impactos interseccionais e das relações causais envolvidas (Tang et al., 2023).

As métricas quantitativas de *fairness* – ainda que úteis – representam apenas uma dimensão da equidade algorítmica, pois se concentram em medir o desequilíbrio observável entre grupos, mas não abordam as causas subjacentes do viés. Tal como aponta o estudo desenvolvido por Mehrabi et al. (2021, p. 11-13) não existe uma definição universal de justiça. De tal modo, diferentes métricas podem ser mutuamente incompatíveis: satisfazer uma forma de *fairness* (por exemplo, *demographic parity*) pode implicar violar outra (como *equal opportunity*).

Como observa Wajcman et al. (2020), o próprio conceito de justiça é historicamente situado e socialmente construído: o que é considerado “justo” varia de acordo com os valores, experiências e posições de poder de quem define o sistema. Equipes homogêneas de desenvolvimento tendem a reproduzir suas próprias perspectivas, criando algoritmos que refletem seus contextos culturais e ignoram as experiências de grupos marginalizados, como mulheres e minorias de gênero (Young et al., 2021). Assim, a inclusão de métricas de *fairness* não exclui a necessidade de diversidade no grupo de programadores, por exemplo. Nesse sentido afirma Hempel (2017), citando Melinda Gates: “Se não incluirmos mulheres e pessoas negras na mesa – tecnólogos de verdade, fazendo o trabalho de fato – vamos acabar viesando

os sistemas. Tentar corrigir isso daqui a uma ou duas décadas será muito mais difícil, se não praticamente impossível.”

Embora as métricas de *fairness* sejam ferramentas úteis para identificar e quantificar disparidades, elas frequentemente refletem apenas os valores e prioridades de um grupo restrito – geralmente técnicos e pesquisadores – e não as complexas realidades sociais envolvidas. Além disso, quando aplicadas genericamente, sem considerar o contexto, podem até perpetuar desigualdades. Uma métrica que busca resultados iguais entre grupos, sem levar em conta condições históricas e estruturais desiguais, pode, paradoxalmente, reproduzir as mesmas injustiças que pretende corrigir (Foulds et al., 2019; Hutchinson; Mitchell, 2019).

Ademais, embora métricas como essas sejam valiosas para avaliar a distribuição dos resultados entre grupos, elas não substituem a necessidade de explicabilidade e contextualização ética. As métricas tratam dos sintomas do viés, mas não de suas origens. Por isso, as métricas devem ser abordadas como um processo contínuo e sistêmico, que combine soluções algorítmicas, revisões institucionais e políticas públicas de governança tecnológica.

Muitas métricas se concentram em equalizar resultados, mas não enfrentam as causas estruturais do viés, como a falta de diversidade nas equipes de desenvolvimento ou a baixa representatividade feminina nas bases de dados. Quando essas causas não são abordadas, as métricas podem apenas mascarar o problema, sem eliminá-lo.

Portanto, a *fairness* é inerentemente subjetiva e dependente dos valores dos agentes que projetam e supervisionam os sistemas de IA. Para que ela produza efeitos substantivos, é necessário adotar uma abordagem holística e interseccional, que envolva:

- a) Diversificação das equipes de desenvolvimento, incorporando perspectivas de gênero e raça;
- b) Melhoria da qualidade e representatividade dos dados, com curadoria ética e revisão periódica;
- c) Consideração explícita das implicações sociais e políticas de cada definição de justiça aplicada à tecnologia.

Nesse sentido, é imprescindível também o uso de dados mais detalhados e análises empíricas sobre a presença feminina na IA, de modo a compreender os mecanismos pelos quais o viés é gerado e reproduzido. O objetivo não é apenas detectar o viés, mas evitar a criação de IAs estruturalmente enviesadas desde sua concepção.

A adoção do princípio da *fairness* no âmbito do Poder Judiciário deve se dar em conjunto com outras técnicas para a garantia de uma IA ética. Por outro lado, pode-se afirmar

que aplicar métricas e práticas de equidade algorítmica em sistemas judiciais significa reconhecer que a neutralidade tecnológica é um mito e que toda decisão automatizada reflete escolhas humanas, políticas e culturais.

Assim, tal como aponta a perspectiva proposta por Mehrabi et al. (2021), a *fairness* é um processo interdisciplinar e dinâmico, que exige tanto mecanismos técnicos de mensuração e mitigação quanto estruturas institucionais de responsabilização e diversidade. Somente dessa forma é possível que os sistemas de IA aplicados ao Judiciário contribuam efetivamente para reduzir as desigualdades de gênero, e não para automatizá-las.

A equidade algorítmica, nesse contexto, deve servir como instrumento de mitigação da sub-representação feminina e das assimetrias de gênero no Judiciário, articulando-se às diretrizes de governança da IA do CNJ e ao Programa Justiça 4.0 (Conselho Nacional de Justiça, [20-?]). Assim, a *fairness* torna-se não apenas um parâmetro técnico, mas um princípio orientador da ética pública e da inclusão institucional, promovendo uma IA que não apenas funcione, mas que funcione de modo justo, transparente e humano.

5.1.2 Transparência dos sistemas inteligentes

A transparência constitui um dos pilares fundamentais da governança ética da IA, especialmente em domínios públicos sensíveis como o Poder Judiciário, em que a legitimidade institucional depende da compreensão, auditabilidade e controle social das decisões. No contexto da Justiça, a IA não pode ser tratada como uma *black box*, mas como um instrumento sujeito a escrutínio público, revisão técnica e fundamentação jurídica.

A transparência de um modelo está “relacionada à habilidade de entender o modelo gerado de forma simples e direta” (Salomão, 2022, p. 32).

Na legislação brasileira, o princípio da transparência é elencado como requisito de observância no tratamento de dados pessoais pelo art. 6º, VI da Lei Geral de Proteção de Dados (LGPD). No art. 9º do referido diploma, estabelece-se implicitamente como direito do titular de dados a transparência quanto às informações sobre o tratamento de seus dados (Brasil, 2018a).

No contexto do Poder Judiciário, a Resolução n. 332/2020 do CNJ (Conselho Nacional de Justiça, 2020b), que estabelecia princípios e diretrizes para o uso da IA no sistema judicial brasileiro, definia a transparência como requisito essencial da ética algorítmica e da proteção dos direitos fundamentais. O preâmbulo do ato normativo explicita que a IA deve atender a

“critérios éticos de transparência, previsibilidade, possibilidade de auditoria e garantia de imparcialidade e justiça substancial”, reconhecendo que o uso de sistemas automatizados pode afetar diretamente o devido processo legal, a isonomia e a motivação das decisões judiciais.

O art. 8º da Resolução detalhava o conceito de transparência, desdobrando-o em seis dimensões complementares, que materializam o princípio da publicidade e a obrigação de prestação de contas do Poder Judiciário (Conselho Nacional de Justiça, 2020a):

- a) divulgação responsável, que impõe a comunicação pública sobre o uso de modelos de IA, respeitando a sensibilidade e a confidencialidade dos dados judiciais;
- b) indicação de objetivos e resultados pretendidos, exigindo clareza sobre a finalidade da aplicação e sobre os parâmetros de sucesso do modelo;
- c) documentação de riscos e medidas de segurança, que reforça a governança técnica e a necessidade de gestão contínua dos impactos éticos e informacionais;
- d) identificação do motivo em caso de dano, permitindo a rastreabilidade das decisões automatizadas e a responsabilização em casos de erro ou prejuízo;
- e) apresentação de mecanismos de auditoria e certificação de boas práticas, criando condições para verificação externa e controle institucional; e
- f) fornecimento de explicação satisfatória e auditável por autoridade humana, o que alinha o princípio de transparência à exigência constitucional de motivação das decisões judiciais (art. 93, IX, da Constituição Federal) (Brasil, 1988).

Esses dispositivos convergem com a literatura internacional sobre explicabilidade (*explainability*) e responsabilidade algorítmica (*accountability*), ao defender que as decisões mediadas por IA devem ser compreensíveis e justificáveis por seres humanos – não apenas em termos técnicos, mas também sob a ótica normativa e ética. A transparência deve ir além da mera abertura de código: trata-se de tornar o funcionamento dos algoritmos socialmente inteligível, permitindo o questionamento e a correção de eventuais injustiças.

Além disso, os arts. 24 e 25 da Resolução n. 332/2020 reforçavam a preferência pelo uso de *software* de código aberto, associando a transparência ao princípio da interoperabilidade,

da colaboração e da cooperação institucional (Conselho Nacional de Justiça, 2020b). Soluções abertas facilitam a integração entre tribunais, aumentam a segurança pela auditabilidade do código e ampliam a confiança pública, permitindo a participação de universidades, órgãos públicos e sociedade civil na fiscalização e aprimoramento das ferramentas de IA.

O art. 25 ainda determinava que qualquer sistema de IA do Judiciário deve assegurar “total transparência na prestação de contas”, garantindo que seus impactos sejam mensuráveis, auditáveis e positivos para os usuários e para a sociedade. Trata-se, portanto, de uma noção de transparência que não se limita à técnica, mas que incorpora o dever institucional de *accountability* e de justiça substantiva.

Em síntese, a Resolução nº 332/2020 do CNJ estabelecia que a transparência é a condição de legitimidade da IA judicial: ela permite que as partes compreendam os fundamentos de uma decisão automatizada, que os tribunais prestem contas de suas ferramentas, e que o próprio sistema judicial mantenha o controle humano sobre as decisões assistidas por algoritmos (Conselho Nacional de Justiça, 2020b). Essa concepção reforça o ideal de uma IA explicável, auditável e ética, capaz de promover eficiência sem sacrificar o direito à informação, à igualdade e à motivação judicial.

Sobre o assunto, vale ressaltar que a Resolução CNJ n. 615/2025 representa um avanço decisivo na consolidação da transparência como eixo estruturante da governança da IA no Poder Judiciário. Ela revoga e aprimora a Resolução n. 332/2020, contemplando os desafios trazidos pela IA generativa e pelos grandes modelos de linguagem (LLMs), bem como reafirmando que o uso de algoritmos na esfera judicial deve ocorrer de forma segura, ética, auditável e plenamente compreensível para magistrados, servidores e jurisdicionados (Conselho Nacional de Justiça, 2025c)

Sob essa regulamentação, a transparência não é tratada apenas como princípio abstrato, mas como um dever técnico, administrativo e democrático, expresso em diversas disposições. Logo no art. 1º, §3º, da Resolução CNJ n. 615/2025, define-se que os padrões de transparência judicial devem observar indicadores claros e relatórios públicos, apresentados em linguagem acessível, garantindo que os cidadãos saibam quando e como a IA está sendo empregada em processos judiciais (Conselho Nacional de Justiça, 2025c).

O art. 2º, inciso XII, reforça que a transparência abrange também os relatórios de auditoria e de avaliação de impacto algorítmico, enquanto o art. 3º, inciso II, a insere, ao lado da eficiência, explicabilidade, contestabilidade e auditabilidade, como um princípio central da ética digital e da justiça algorítmica (Conselho Nacional de Justiça, 2025b).

Nos arts. 12 a 14, a transparência é vinculada às medidas de governança interna: cada tribunal deve publicar relatórios sobre o funcionamento dos sistemas, os dados utilizados e as formas de supervisão humana, além de adotar mecanismos que permitam rastrear decisões, registrar *logs* de operação e disponibilizar documentação compreensível sobre o funcionamento dos modelos de IA. O texto também estabelece que a explicabilidade é parte integrante da transparência (art. 13, VII), assegurando que as decisões automatizadas sejam passíveis de entendimento e revisão por pessoas humanas – uma exigência diretamente associada à motivação das decisões judiciais prevista no art. 93, IX, da Constituição Federal (Brasil, 1988; Conselho Nacional de Justiça, 2025b).

A Resolução introduz, ainda, uma camada de transparência pública e institucional por meio dos arts. 22 a 25, que criam um sistema de registro e publicidade das soluções de IA adotadas no Judiciário. Todos os modelos devem ser cadastrados na plataforma Sinapses, com informações sobre finalidades, grau de risco, desenvolvedores e resultados pretendidos. O art. 25 determina que o CNJ publique periodicamente um catálogo acessível à sociedade, em linguagem simples, contendo a descrição das aplicações, o nível de risco e as justificativas para eventual suspensão ou descontinuidade. Essas medidas aproximam o Judiciário dos padrões internacionais de transparência algorítmica e governança aberta, ao possibilitar controle social, fiscalização cidadã e diálogo interinstitucional (Conselho Nacional de Justiça, 2025b).

Além da transparência informacional, a norma prevê a transparência procedimental e de responsabilização. O art. 39 impõe “total transparência na prestação de contas”, exigindo divulgação de responsáveis, custos, resultados, riscos e canais de comunicação acessíveis ao público. Essa obrigação traduz o conceito de *accountability* algorítmica, conectando a gestão da IA à ética da administração pública e à garantia de impacto social positivo (Conselho Nacional de Justiça, 2025b).

Já o art. 33 estende o princípio à qualidade de direito à transparência em favor de usuários externos, impondo a comunicação clara sobre o uso de IA nos serviços judiciais (Conselho Nacional de Justiça, 2025b).

Por fim, o referido documento entende a transparência como instrumento de legitimidade democrática. Ela articula esse princípio com a proteção de dados, a explicabilidade, a diversidade das equipes de desenvolvimento (art. 35) e o controle público por meio do Comitê Nacional de Inteligência Artificial do Judiciário (arts. 15-18), que deverá elaborar relatórios anuais e disponibilizá-los ao público (Conselho Nacional de Justiça, 2025b). Essa arquitetura normativa consagra uma visão madura e inclusiva de transparência

algorítmica, que combina abertura técnica, prestação de contas, diversidade e inteligibilidade social, reconhecendo que a legitimidade da IA judicial depende da sua capacidade de representar e respeitar a pluralidade social, inclusive de gênero.

Nesse sentido, a transparência também desempenha papel essencial na mitigação de vieses de gênero em sistemas inteligentes aplicados à justiça. Ao permitir a visibilização das bases de dados, dos critérios de treinamento e das métricas de desempenho, ela torna possível identificar sub-representações femininas, padrões discriminatórios e distorções estatísticas que, em contextos opacos, permaneceriam invisíveis. Além disso, a abertura dos processos de desenvolvimento e o incentivo à diversidade nas equipes promovem a inclusão de perspectivas plurais, reduzindo o risco de que os algoritmos reproduzam as mesmas assimetrias históricas presentes nas decisões humanas.

Em termos institucionais, a transparência cria condições de auditoria e contestabilidade das decisões automatizadas, fortalecendo o controle público e permitindo que eventuais discriminações de gênero sejam detectadas, explicadas e corrigidas. Isso converte o princípio da transparência em um mecanismo ativo de igualdade substantiva, que assegura não apenas o conhecimento técnico sobre a IA, mas também a participação social e epistêmica das mulheres na construção, supervisão e revisão dessas tecnologias.

É crucial garantir transparência e divulgação responsável em relação aos sistemas de IA, de modo que as pessoas possam compreender os resultados baseados em IA e contestá-los, se necessário. Nesse ponto, é imperioso que as informações fornecidas sejam amplas e de fácil compreensão, possibilitando a conscientização da população interessada sobre a sua interação com o sistema de IA (Salomão, 2022, p. 49).

O objetivo da Resolução é inequívoco: garantir que a automação de tarefas e o apoio algorítmico às decisões judiciais ocorram sem opacidade, discriminação ou exclusão, assegurando que cada resultado produzido por um sistema inteligente possa ser compreendido, auditado e, quando necessário, contestado – preservando, assim, o núcleo ético do devido processo legal e a autonomia da magistratura. A transparência dos dados e dos modelos de IA é, portanto, condição indispensável para que as decisões tomadas com apoio de algoritmos sejam fundamentadas, legítimas, confiáveis e livres de vieses de gênero, afastando a possibilidade de se criarem “caixas-pretas” decisórias dentro do sistema judicial.

Todavia, é importante reconhecer que, na prática, muitos sistemas de IA ainda carecem de transparência. A complexidade dos modelos algorítmicos, aliada à falta de *accountability* de entidades públicas e privadas, frequentemente dificulta que o público compreenda os impactos

dessas tecnologias sobre os direitos fundamentais e sobre a própria estrutura de poder das instituições.

Para enfrentar esse cenário, recomenda-se que os relatórios públicos sobre o uso de IA sejam regulares, claros e acessíveis, indicando onde o sistema é utilizado, quais dados o alimentam, para que fins é empregado e quais são seus resultados e possíveis riscos. Também devem ser publicizadas as medidas de mitigação de viés e de segurança, além de se incentivar o uso de dados e códigos abertos e a realização de auditorias independentes – práticas que ampliam a confiança pública e reduzem a opacidade institucional (Accessnow.Org, 2018, p. 33).

Por fim, o emprego de sistemas de IA transparentes, explicáveis e eticamente orientados deve incluir mecanismos de notificação e reparação às pessoas potencialmente afetadas, sobretudo quando houver risco de discriminação algorítmica. Assim, ao garantir transparência e comunicação adequada, o Poder Judiciário fortalece a confiança social, assegura a compreensibilidade das decisões automatizadas e reafirma seu compromisso com uma IA a serviço da justiça, da igualdade de gênero e da democracia.

5.1.3 Explicabilidade dos sistemas inteligentes

“Talvez o motivo mais difundido para se explicar algoritmos seja o seu amplo e crescente impacto social” (Watson; Floridi, 2020, p. 9213, tradução nossa³¹). À medida que sistemas de IA passam a mediar decisões que afetam diretamente a vida das pessoas – como o acesso à justiça, à saúde, ao crédito ou a políticas públicas –, a necessidade de torná-los explicáveis deixa de ser apenas uma questão técnica e passa a constituir um imperativo ético e democrático. Explicar algoritmos significa abrir a “caixa-preta” de seus processos internos, permitindo compreender quais fatores influenciaram determinado resultado, quais dados foram utilizados e de que modo o sistema chegou à sua conclusão.

A explicabilidade, portanto, não se limita a traduzir cálculos complexos, mas a garantir que os processos decisórios mediados por IA sejam inteligíveis, rastreáveis e passíveis de contestação. No contexto jurídico, essa exigência adquire dimensão ainda mais relevante: a compreensão dos fundamentos de uma decisão – seja humana, seja automatizada – é condição essencial do devido processo legal e da própria legitimidade do Poder Judiciário. Assim,

³¹ No original: “Perhaps the most popular reason to explain algorithms is their large and growing social impact”.

explicar algoritmos é assegurar que o exercício da IA permaneça subordinado à razão pública, aos direitos fundamentais e à responsabilidade institucional.

Sistemas inteligentes devem ser explicáveis ao público em geral, especialmente aos seus usuários. A IA deve ser explicável no sentido de ser possível demonstrar quais são os processos técnicos empregados no sistema e qual é o raciocínio aplicado às decisões por ele tomadas, possibilitando às partes envolvidas compreender e contestar o seu uso. Seres humanos precisam ser cientificados a respeito das capacidades e limitações dos sistemas que utilizam ou aos quais estão submetidos.

Segundo Monteiro (2021, p. 56), uma das principais limitações e desafios ao direito à explicação seria: “como garantir que explicações sejam compreensíveis e úteis aos seus destinatários, de modo a assegurar que estes possam delas extrair proveito para contestar as decisões que afetem significativamente sua esfera de direitos?”.

De acordo com o autor, o direito à explicação vai além da mera compreensão da lógica, forma e critérios empregados sobre os dados pessoais utilizados por sistemas inteligentes, possibilitando também a prevenção de impactos, com a finalidade de evitar a violação de direitos individuais e coletivos (Monteiro, 2021, p. 43):

O direito à explicação pode ser entendido como o direito a receber informações suficientes e inteligíveis que permitam ao titular dos dados, e à sociedade, entenderem, compreenderem, a lógica, a forma, os critérios utilizados para tratar dados pessoais e, a partir dessa explicação, prever, antever os seus impactos, com o fim de evitar práticas discriminatórias, ilegítimas e indesejadas, que podem ter impacto no plano individual e coletivo, para que possam, ainda, ser desafiadas e revisadas por meio do exercício de direitos e do devido processo.

O requisito da explicabilidade está atrelado à necessidade de que sistemas de IA sejam confiáveis. Freitas (2025, p. 162-178) traz recorte importante sobre como a explicabilidade em IA não é trivial, mas uma tarefa que exige conhecimentos técnicos desde a fase de projeto de um programa de computador até os requisitos de execução, de base de dados e de resultados.

Na ciência da computação, esse princípio tem originado um campo de pesquisa próprio denominado XAI, que busca desenvolver métodos, modelos e interfaces capazes de tornar os resultados da IA mais compreensíveis para humanos (Gilpin et al., 2018; Gunning; Aha, 2019; Miller, 2019). A área emergiu com força a partir da segunda metade da década de 2010, impulsionada pela crescente adoção de redes neurais profundas, cuja natureza opaca deu origem ao chamado “problema da caixa-preta” (Doshi-Velez; Kim, 2017; Lipton, 2018). Pesquisas em XAI têm se concentrado em métodos pós-hoc – como o Local Interpretable Model-agnostic Explanations (LIME) e o SHapley Additive exPlanations (SHAP) – que geram explicações

locais sobre previsões específicas, e em métodos intrínsecos, que estruturam modelos transparentes desde sua concepção (Ribeiro et al., 2016; Lundberg; Lee, 2017).

Esses estudos enfatizam que explicabilidade não se limita à descrição matemática de um modelo, mas envolve a produção de justificativas cognitivamente acessíveis, que permitam ao ser humano compreender e confiar no sistema (Gilpin et al., 2018; Miller, 2019). A pesquisa em XAI tem se tornado interdisciplinar, conectando fundamentos da computação, psicologia cognitiva, ética e direito. No contexto judicial, tais abordagens têm sido particularmente relevantes para a construção de sistemas de decisão assistida que possam ser auditáveis, contestáveis e juridicamente defensáveis.

A confiabilidade, por sua vez, é condição de legitimidade e eficácia social: um sistema de IA somente pode ser considerado justo quando suas decisões são compreensíveis e rastreáveis por pessoas humanas. Além disso, a explicabilidade pode melhorar as propriedades relacionadas à equidade de determinado sistema de IA em variados níveis e intensidades. Ao permitir a compreensão de como o modelo chegou a determinado resultado discriminatório, ou se características importantes foram levadas em conta para torná-lo mais justo, a explicabilidade fomenta e reforça a confiabilidade de um modelo, aumentando a sua transparência e a sua “justeza” (Doshi-Velez; Kim, 2017; Abdollahi; Nasraoui, 2018; Dodge et al., 2019).

A explicabilidade também pode melhorar a responsabilidade, fornecendo justificativas para decisões (in)justas baseadas em IA. Esses aspectos são relevantes no contexto de “IA responsável”, que considera questões morais, éticas e valores sociais.

Além de fundamento ético e jurídico, a explicabilidade possui um papel estratégico e substantivo na mitigação de vieses de gênero no Poder Judiciário. Ao tornar visíveis as variáveis consideradas, as ponderações estatísticas e as relações de correlação aprendidas pelo modelo, a explicabilidade permite identificar se o sistema está reproduzindo padrões discriminatórios presentes nos dados históricos – como o tratamento desigual entre homens e mulheres em decisões de custódia, progressão de regime ou promoção funcional. Essa visibilidade viabiliza auditorias direcionadas, ajustes de modelagem e curadoria de dados sensíveis, prevenindo que vieses implícitos se consolidem como decisões automatizadas. Em outras palavras, sistemas explicáveis fortalecem o dever de motivação judicial e o controle da discriminação indireta, contribuindo para uma justiça mais equitativa e inclusiva.

Em especial, no enfrentamento do viés de gênero, a explicabilidade atua como uma ferramenta de justiça substantiva, ao tornar perceptíveis os mecanismos de inferência que poderiam perpetuar desigualdades históricas entre homens e mulheres. Quando um sistema de

IA é capaz de explicar quais variáveis influenciaram determinada predição, é possível identificar se o gênero, direta ou indiretamente, afetou o resultado. Essa transparência permite não apenas detectar o viés, mas também revisá-lo, ajustando parâmetros, bases de dados ou regras de decisão. Assim, a explicabilidade não se limita a traduzir o funcionamento técnico do modelo, mas cria condições efetivas para a revisão institucional e a *accountability* democrática de decisões automatizadas (Doshi-Velez; Kim, 2017; Watson; Floridi, 2020; Monteiro, 2021).

Nessa mesma direção, Binns (2018, p. 5-8) argumenta que a justiça algorítmica exige não apenas a correção de resultados desiguais, mas também o reconhecimento das formas simbólicas de desigualdade reproduzidas pelos sistemas de aprendizado de máquina. Ao discutir a noção de *representational fairness*, o autor observa que muitos dos vieses mais persistentes em IA – como os de gênero, presentes em bases linguísticas e modelos de predição – estão ligados à forma como grupos sociais são representados ou invisibilizados nas estruturas de dados.

A explicabilidade, nesse sentido, é fundamental porque torna transparentes essas representações e revela as escolhas valorativas embutidas no processo algorítmico, permitindo que os sistemas judiciais compreendam o impacto social de suas inferências automatizadas (Binns, 2018). Assim, sistemas explicáveis criam condições para o escrutínio público, o questionamento normativo e a revisão institucional, funcionando como instrumentos de *accountability* e de correção democrática das desigualdades de gênero. Em última instância, a explicabilidade não apenas garante que a IA reproduza a realidade de forma compreensível, mas também que contribua ativamente para transformá-la em direção à igualdade de gênero.

Ainda, a explicabilidade está igualmente relacionada ao direito à informação, na medida em que os indivíduos possuem o direito de saber e compreender os critérios que o sistema inteligente utiliza para a tomada de decisão – quais são os dados empregados, as escolhas possíveis, as variáveis consideradas e os critérios que determinaram a conclusão automatizada. O direito à informação, como corolário do devido processo legal, opera como mecanismo institucional de garantia do direito à explicação em face de decisões automatizadas, assegurando que nenhum cidadão seja afetado por uma decisão opaca ou ininteligível.

Assegurar que o titular compreenda efetivamente o funcionamento do sistema algorítmico, sua lógica subjacente e os impactos das decisões automatizadas em seus direitos vai além do simples fornecimento de informações. O direito à explicação requer que o responsável apresente informações conectadas, claras e inteligíveis, capazes de permitir uma

compreensão efetiva do sistema automatizado – incluindo os fundamentos das decisões, os dados de entrada e os critérios que conduziram à automação dos resultados.

Os autores Watson e Floridi (2020, p. 9220-9221) destacam que o direito à explicação envolve a interpretação do algoritmo em seu conjunto. As explicações podem se referir tanto à predição individual quanto à lógica geral do modelo matemático adotado. Explicar não é apenas revelar o código, mas tornar compreensível o raciocínio algorítmico: quais inferências foram produzidas, por qual razão determinadas variáveis tiveram maior peso e como o sistema chegou a um resultado específico. De acordo com os autores, as explicações devem ser precisas, simples e relevantes, oportunizando o refazimento dos procedimentos algorítmicos, a compreensão do caminho traçado para a obtenção da solução e permitindo o questionamento pelos indivíduos afetados (Watson; Floridi, 2020, p. 9226-9229).

É crucial aumentar a transparência e a capacidade de auditoria dos sistemas, tanto desenvolvendo as habilidades necessárias para observar, compreender e auditar seu funcionamento, quanto investindo consideravelmente em pesquisas sobre “explicabilidade” no campo da IA. Além disso, é imperativo adaptar a proteção dos direitos e liberdades para lidar com o potencial abuso associado ao uso de sistemas de aprendizado de máquina (Villani et al., 2018, p. 140-142).

No contexto normativo do Judiciário brasileiro, a Resolução n. 332/2020 do CNJ consagrou a explicabilidade como elemento essencial do uso ético da IA judicial (Conselho Nacional de Justiça, 2020b). O art. 19 do ato estabelecia que “os sistemas computacionais que utilizem modelos de Inteligência Artificial como ferramenta auxiliar para a elaboração de decisão judicial observarão, como critério preponderante para definir a técnica utilizada, a explicação dos passos que conduziram ao resultado”. Essa diretriz vinculava a escolha da técnica de IA ao dever de fundamentação das decisões judiciais (art. 93, IX, da Constituição Federal) (Brasil, 1988), reforçando que nenhuma decisão automatizada pode ser legítima sem que o raciocínio que a originou seja compreensível e auditável por magistrados e partes.

A Resolução n. 615/2025 do CNJ avança nesse mesmo sentido, ao reconhecer a explicabilidade como princípio estruturante da ética e da governança digital (art. 3º, II). Ela determina que todos os sistemas baseados em IA empregados no Judiciário sejam acompanhados de documentação técnica detalhada, relatórios de avaliação de risco e registros de operação (arts. 12 e 13), garantindo a rastreabilidade das decisões e a revisão humana obrigatória. O ato também exige que as explicações sejam disponibilizadas em linguagem acessível, permitindo que jurisdicionados compreendam de modo claro os critérios e dados que

embasaram uma decisão automatizada. Trata-se de uma inovação significativa, pois desloca a explicabilidade do plano técnico para o plano democrático – vinculando-a ao controle público e à legitimidade institucional do sistema de justiça (Conselho Nacional de Justiça, 2025c).

Além de fundamento ético e jurídico, a explicabilidade possui um papel estratégico na mitigação de vieses de gênero no Poder Judiciário. Ao tornar visíveis as variáveis consideradas, as ponderações estatísticas e as relações de correlação aprendidas pelo modelo, a explicabilidade permite identificar se o sistema está reproduzindo padrões discriminatórios presentes nos dados históricos – como, por exemplo, o tratamento desigual entre homens e mulheres em decisões de custódia, progressão de regime ou promoção funcional. Essa visibilidade viabiliza auditorias direcionadas, ajustes de modelagem e curadoria de dados sensíveis, prevenindo que vieses implícitos se consolidem como decisões automatizadas. Em outras palavras, sistemas explicáveis fortalecem o dever de motivação judicial e o controle da discriminação indireta, contribuindo para uma justiça mais equitativa e inclusiva.

Em que pese a regulamentação acima citada esteja centrada na proteção do indivíduo, ela não está alinhada com a lógica introduzida por esses sistemas – isto é, a análise de grandes volumes de informações para identificar tendências e comportamentos ocultos – e seus impactos sobre grupos de indivíduos. Para abordar essa lacuna, é essencial estabelecer direitos coletivos sobre os dados. Ao mesmo tempo, é imperativo garantir que as organizações que implementam e utilizam esses sistemas permaneçam sujeitas à lei por qualquer dano que possam causar (Villani et al., 2018, p. 140-142).

Ainda, observa-se que a relação entre explicabilidade e justiça nem sempre é positiva. Os autores esclarecem que a explicabilidade pode impactar a percepção de justiça. Nesse contexto, alguns ressaltam os riscos associados a uma maior explicabilidade e dependência dela. Aivodji et al. (2019) colocam como exemplo de risco a *fairwashing* e a racionalização e possível justificação de certos tipos de discriminação. Por sua vez, Ananny e Crawford (2018) apontaram que a capacidade de tornar um modelo transparente, explicável, visível não equivale necessariamente à capacidade de “controlá-lo” ou “compreendê-lo” em sua completude e, portanto, de mitigar o viés. A transparência, por si só, não assegura controle, compreensão integral nem mitigação de vieses. Um sistema pode ser tecnicamente “aberto” e ainda assim permanecer opaco em termos sociais, políticos ou cognitivos – especialmente quando envolve decisões complexas, como as judiciais.

Nessa relação entre justiça e explicabilidade, Barocas et al. (2020), ao abordar as tensões entre explicabilidade, complexidade algorítmica e poder de decisão, destacam as

tensões entre o desenvolvimento de modelos mais simples, com o intuito de garantir transparência, e o fato de que tais modelos podem ser menos eficazes em atender a algumas exigências de justiça.

Em síntese, a explicabilidade constitui um dos pilares da ética algorítmica contemporânea, especialmente no contexto do Poder Judiciário. Sua função transcende a mera transparência técnica: ela viabiliza a responsabilização institucional, a revisão pública das decisões automatizadas e o fortalecimento do devido processo legal. Quando aplicada de forma robusta, a explicabilidade permite não apenas compreender como um sistema chegou a determinado resultado, mas também avaliar a legitimidade das premissas que o sustentam. Essa dimensão cognitiva e normativa é indispensável para que a IA não se converta em um instrumento de opacidade ou de reforço de desigualdades estruturais, mas em um meio de promoção da justiça, da igualdade de gênero e da confiança pública nas instituições judiciais.

Além disso, ao iluminar os mecanismos internos dos modelos e revelar as escolhas valorativas embutidas em seus processos de inferência, a explicabilidade cria um espaço de diálogo entre o técnico e o jurídico, aproximando a engenharia algorítmica dos princípios constitucionais de transparência, motivação e controle social. Ainda que existam limites inerentes – como o risco de *fairwashing* ou de simplificação indevida dos modelos –, o avanço de práticas explicáveis e auditáveis é condição necessária para uma governança algorítmica democrática. Assim, o verdadeiro desafio não é apenas explicar algoritmos, mas assegurar que essas explicações produzam compreensão efetiva, permitam contestação informada e reforcem o compromisso ético do Judiciário com a justiça substantiva e com a equidade de gênero.

5.1.4 Diversidade e não discriminação na construção e no emprego dos sistemas inteligentes

A diversidade, tanto dos dados utilizados no treinamento de modelos de IA quanto das equipes responsáveis por sua concepção e desenvolvimento, constitui um requisito elementar para a construção de sistemas justos, auditáveis e socialmente legítimos. A homogeneidade – seja dos programadores, seja das fontes de informação que alimentam os algoritmos – tende a reproduzir e amplificar desigualdades pré-existentes, especialmente aquelas relacionadas a gênero, raça e classe social.

A discriminação, segundo o High-Level Expert Group on Artificial Intelligence (2018, p. 16), refere-se à variação dos resultados da IA entre indivíduos ou grupos, fundamentada na exploração de diferenças em suas características, que podem ser consideradas intencionais ou

não (como etnia, gênero, orientação sexual ou idade), e que podem afetar negativamente esses indivíduos ou grupos. Tal discriminação pode ter origem em vieses técnicos, escolhas de *design* ou má governança de dados e tende a se manifestar de forma indireta e não intencional, perpetuando desigualdades estruturais e marginalização.

No contexto do Poder Judiciário, essa questão assume uma dimensão ainda mais crítica. Os sistemas de IA aplicados à justiça – como os utilizados para triagem processual, apoio à decisão e análise preditiva – são alimentados majoritariamente por bases de dados judiciais históricas, produzidas em um ambiente institucional marcado por acentuada desigualdade de gênero e sub-representação feminina. Como apontam relatórios do CNJ, as mulheres representam menos de 40% da magistratura brasileira e uma proporção ainda menor em cargos de direção e cúpula, o que revela um contexto institucional predominantemente masculino e branco. Essa falta de diversidade repercute nos dados que registram práticas, decisões e linguagens judiciais, resultando em bases enviesadas que refletem apenas uma perspectiva parcial da realidade jurídica e social.

Quando tais dados são utilizados para treinar sistemas inteligentes sem curadoria ética e representativa, os modelos aprendem e reproduzem padrões de decisão e linguagem historicamente masculinizados, tendendo a replicar vieses institucionais na predição de resultados, na classificação de demandas e na priorização de casos. Por exemplo, um sistema de IA treinado com sentenças majoritariamente proferidas por magistrados homens pode internalizar vieses inconscientes que afetam negativamente temas relacionados à violência de gênero, direitos reprodutivos, assédio ou discriminação no trabalho. Assim, a ausência de diversidade nos dados e nas equipes de desenvolvimento compromete a imparcialidade algorítmica e reforça a exclusão simbólica das mulheres no sistema de justiça.

Os sistemas de IA, tanto durante o treinamento quanto na operação, podem ser afetados pela incorporação inadvertida de tendências históricas, lacunas de informação e estruturas de governança inadequadas. A perpetuação desses vieses pode resultar em discriminação não intencional e indireta contra grupos específicos ou indivíduos, contribuindo para a amplificação do preconceito e da marginalização. Por essa razão, a busca pela diversidade e pela não discriminação deve orientar todas as etapas do ciclo de vida de um sistema de IA – desde o *design*, a seleção e o tratamento de dados até o monitoramento pós-implantação.

A diversidade deve ser compreendida, portanto, não apenas como um valor social, mas como um princípio técnico e ético de governança algorítmica. Destaque-se que o High-Level Expert Group on Artificial Intelligence (2018, p. 16–17) defende que a diversidade deve

compor o *design* da IA como forma de proteger e garantir a pluralidade de valores, a autodeterminação e a autonomia individual. Nesse sentido, a heterogeneidade das equipes de desenvolvimento – incorporando profissionais de diferentes gêneros, raças, formações disciplinares e origens regionais – é condição indispensável para a criação de sistemas mais justos e confiáveis. A diversidade cognitiva e social favorece a identificação precoce de falhas, a contestação de pressupostos implícitos e o desenvolvimento de mecanismos de mitigação de vieses antes que eles se consolidem no modelo.

Alguns especialistas recomendam a diversidade mesmo entre os programadores de sistemas de IA, pois isso facilita a identificação e a abordagem de possíveis falhas de treinamento dos algoritmos (Fonseca, 2022, p. 297). Estudos recentes mostram que equipes homogêneas tendem a reproduzir uma visão limitada de neutralidade técnica, o que impede a percepção de como o sistema pode impactar desproporcionalmente mulheres, pessoas negras e outras minorias (Binns, 2018; Buolamwini; Gebu, 2018; Noble, 2018; Obermeyer et al., 2019; Ahmad et al., 2022; Bhatnagar; Gajjar, 2024). Por isso, a diversificação das equipes técnicas é também uma estratégia de *accountability*, pois amplia a capacidade crítica sobre as escolhas éticas e políticas incorporadas ao *design* dos sistemas.

A Resolução n. 615/2025 do CNJ reforça esse entendimento ao reconhecer a diversidade como princípio estruturante da governança ética da IA judicial (art. 35), determinando que o desenvolvimento e a aplicação de sistemas inteligentes considerem a composição plural das equipes e o monitoramento contínuo de vieses nos dados e resultados (Conselho Nacional de Justiça, 2025c). Essa diretriz está em consonância com o princípio da não discriminação previsto na Constituição Federal (art. 3º, IV) (Brasil, 1988) e com os compromissos internacionais assumidos pelo Estado brasileiro, como a Convenção sobre a Eliminação de Todas as Formas de Discriminação contra a Mulher (CEDAW).

A diversidade, portanto, deve ser um elemento intrínseco ao *design* dos sistemas de IA. Quando identificados, preconceitos devem ser prontamente eliminados do sistema, e os processos de auditoria e revisão devem assegurar que o aprendizado algorítmico não reproduza distorções históricas.

Garantir diversidade nos dados e nas equipes técnicas é, assim, um imperativo ético, jurídico e democrático, fundamental para a promoção da igualdade substantiva e para a legitimação do uso da IA no Poder Judiciário. Somente por meio de uma IA plural e inclusiva será possível promover a justiça sem discriminação e construir sistemas que reflitam a

complexidade social brasileira, e não apenas a visão restrita de seus grupos historicamente dominantes.

5.1.5 *Design acessível dos sistemas inteligentes*

A construção de sistemas de IA voltados ao Poder Judiciário deve observar o princípio da acessibilidade como requisito técnico e ético indispensável à promoção da justiça social (Leiner et al., 2025). Um *design* acessível é aquele que permite o usufruto pleno e igualitário dos sistemas de IA por todas as pessoas, independentemente de idade, deficiência, nível educacional, condição socioeconômica ou localização geográfica, garantindo que as tecnologias desenvolvidas sirvam à totalidade da sociedade e não apenas a grupos social ou economicamente privilegiados (High-Level Expert Group on Artificial Intelligence, 2018, p. 15).

Esse princípio está intimamente ligado ao direito fundamental de acesso à justiça, positivado em uma ampla gama de instrumentos internacionais e constitucionais. No cenário global, ele encontra respaldo na Declaração Universal dos Direitos Humanos (Organização das Nações Unidas, 1948), no Pacto Internacional de Direitos Civis e Políticos (Organização das Nações Unidas, 1966) e na Convenção Interamericana de Direitos Humanos (Pacto de São José da Costa Rica, 1992). No âmbito interno, a Constituição Federal de 1988, em seu artigo 5º, inciso XXXV (Brasil, 1988), consagra que “a lei não excluirá da apreciação do Poder Judiciário lesão ou ameaça a direito”, assegurando a todos o acesso a uma ordem jurídica justa.

Conforme destaca Piovesan (2017), o acesso à justiça constitui condição *sine qua non* para a concretização de todos os demais direitos humanos, individuais ou coletivos. É o direito sem o qual nenhum outro se efetiva, pois garante a própria possibilidade de reivindicação e proteção das liberdades e garantias fundamentais. Nessa mesma linha, Sadek (2005) observa que qualquer barreira ao acesso à justiça implica restrições à cidadania, perpetuando desigualdades e limitando a inclusão democrática.

De acordo com Cappelletti e Garth (1988), a expressão “acesso à justiça” possui duplo significado: por um lado, refere-se ao direito de todo indivíduo reivindicar seus direitos perante o Estado; por outro, à necessidade de que o sistema produza resultados socialmente justos. Para alcançar esses objetivos, os autores propuseram três “ondas renovatórias” do acesso à justiça: a) a garantia de assistência judiciária aos vulneráveis; b) a representação de direitos difusos e

coletivos; e c) o aperfeiçoamento de mecanismos e procedimentos para prevenir e resolver disputas de forma célere, inclusiva e eficaz.

Essas transformações refletem a passagem de um acesso formal à justiça – entendido como mero direito abstrato – para um acesso material e substancial, que assegura a efetividade do direito mediante a remoção das barreiras concretas que impedem sua realização (Cappelletti; Garth, 1988). A consolidação desse paradigma implica reconhecer que a efetividade do direito de acesso à justiça no século XXI depende da incorporação de ferramentas tecnológicas acessíveis, seguras e inclusivas, capazes de democratizar o uso da IA no sistema judicial.

No entanto, é necessário cautela para que a tecnologia, em vez de promover a inclusão, não agrave as desigualdades existentes. Sob pena de universalizar de modo abstrato o direito humano ao acesso à justiça, as políticas públicas devem considerar a existência de grupos que não dispõem das condições materiais, cognitivas ou tecnológicas para acompanhar os avanços digitais. É imprescindível refletir sobre o desenvolvimento de sistemas de IA sob a ótica da Teoria Crítica dos Direitos Humanos, de modo a garantir que os novos mecanismos tecnológicos não excluam os mais vulneráveis do exercício da cidadania (Flores, 2009; Rodrigues; Santos, 2020; Leiner et al., 2025).

A implementação de um *design* acessível é, portanto, condição essencial para assegurar a universalidade e a não discriminação no uso da IA no Judiciário. Tal *design* deve ser guiado por princípios de inclusão, usabilidade e adaptabilidade, incorporando interfaces intuitivas, linguagem clara, suporte multicanal e recursos de acessibilidade digital voltados a pessoas com deficiência, idosos, populações periféricas e grupos em situação de vulnerabilidade social. Sistemas de IA concebidos com essas premissas ampliam o alcance do acesso à justiça e fortalecem a legitimidade democrática do Poder Judiciário.

Além disso, a acessibilidade deve abranger também o corpo funcional do Poder Judiciário. Os sistemas de IA desenvolvidos para auxiliar magistrados, servidores e equipes técnicas no desempenho de suas atividades rotineiras precisam ser concebidos de modo a assegurar clareza, simplicidade de uso e compreensão das operações realizadas pelo sistema. A adoção de interfaces amigáveis e processos de interação intuitivos é fundamental para que a tecnologia se torne uma ferramenta de apoio e não de alienação do trabalho judicial.

A acessibilidade funcional, nesse sentido, está diretamente relacionada à capacitação dos profissionais da justiça para o uso apropriado e crítico das ferramentas de IA. Magistrados e servidores devem receber formação continuada que lhes permita compreender os fundamentos, limitações e riscos dos sistemas empregados, garantindo autonomia e segurança

no uso das tecnologias. Sem capacitação adequada, há o risco de que a IA seja utilizada de forma mecânica e acrítica, o que enfraquece a função decisória humana e compromete a legitimidade das decisões automatizadas. A formação técnica e ética dos usuários é, portanto, parte constitutiva do próprio *design* acessível, que deve ser pensado como um ecossistema humano-tecnológico, e não apenas como um artefato digital.

A efetiva concretização do acesso a uma ordem jurídica justa depende, assim, da criação de instituições e mecanismos eficazes, corresponsáveis e inclusivos. Como já advertia Bobbio (1992, p. 25), a questão central contemporânea não é mais definir quais são os direitos humanos, mas “como garanti-los de forma segura e impedir que, apesar das solenes declarações, eles sejam continuamente violados”. Aplicado ao contexto tecnológico, isso implica a regulamentação de políticas públicas que assegurem capacitação digital, infraestrutura tecnológica acessível e medidas de inclusão social, sob pena de gerar um apartheid tecnológico que selecione quem tem e quem não tem acesso à justiça digital (Rodrigues; Santos, 2020).

Por fim, cabe salientar que o *design* acessível de sistemas de IA não se restringe à interface técnica, mas reflete um compromisso ético com a igualdade substantiva e a não discriminação. Ao promover a acessibilidade como valor central do desenvolvimento tecnológico e funcional, o Poder Judiciário brasileiro poderá assegurar que as inovações digitais não aprofundem as assimetrias sociais nem profissionais, mas atuem como instrumentos de democratização, capacitação e efetivação dos direitos humanos – tanto para a sociedade quanto para aqueles que nela exercem a função jurisdicional (Leiner et al., 2025).

A adoção de um *design* acessível também desempenha um papel relevante na mitigação de vieses de gênero no Poder Judiciário. Quando os sistemas de IA são concebidos de forma inclusiva, considerando a diversidade de perfis, experiências e formas de interação de seus usuários, reduzem-se as assimetrias que tradicionalmente afastam mulheres e grupos sub-representados dos espaços de decisão tecnológica e institucional.

Interfaces intuitivas, linguagem clara e processos compreensíveis favorecem a apropriação equitativa das ferramentas digitais por magistradas, servidoras e demais profissionais da Justiça, ampliando sua participação efetiva no uso, supervisão e aprimoramento dos sistemas. Do mesmo modo, a acessibilidade dos dados e das funcionalidades dos sistemas permite maior transparência sobre seus critérios de decisão e facilita auditorias de equidade, possibilitando identificar e corrigir padrões algorítmicos que reproduzam estereótipos ou discriminações de gênero.

Assim, o *design* acessível não é apenas um requisito de inclusão digital, mas uma estratégia de justiça de gênero, que busca garantir que a transformação tecnológica do Judiciário ocorra sob parâmetros de igualdade, participação e não discriminação.

5.1.6 Acurácia e robusteza dos sistemas inteligentes

A acurácia e a robusteza são elementos essenciais para garantir a confiabilidade, a legitimidade e a segurança do uso da IA no âmbito do Poder Judiciário. Em termos técnicos, acurácia refere-se ao grau de correção das previsões ou classificações produzidas por um modelo de IA, isto é, à proximidade entre os resultados obtidos e os valores reais observados (Russell; Norvig, 2010). Representa a capacidade de um sistema de processar adequadamente os dados e emitir respostas coerentes com a realidade, sendo expressa por métricas quantitativas como *precision*³², *recall*³³, *F1-score*³⁴ e *AUC-ROC*³⁵. Já a robusteza diz respeito à capacidade do sistema de manter desempenho estável e confiável diante de ruídos, perturbações, erros de entrada, variações de contexto ou ataques adversariais, garantindo a integridade e previsibilidade dos resultados (Sammut; Webb, 2017).

No Poder Judiciário, a robusteza técnica deve ser entendida como sinônimo de segurança institucional e ética. Um sistema robusto é aquele que preserva o desempenho mesmo em situações imprevistas, evitando falhas que possam comprometer direitos fundamentais, a imparcialidade judicial e a legitimidade das decisões. De acordo com Amodei et al. (2016), a robustez está diretamente associada à confiabilidade e à segurança de sistemas de IA, devendo ser constantemente monitorada e ajustada ao longo de todo o ciclo de vida do modelo.

A Resolução CNJ n. 615/2025 (Conselho Nacional de Justiça, 2025c) reforça essa exigência ao determinar que todos os sistemas de IA em uso no Poder Judiciário sejam auditados e monitorados quanto à acurácia, robustez, confiabilidade e prevenção de vieses discriminatórios. O normativo estabelece que os dados de treinamento, validação e teste devem ser representativos, equilibrados e estatisticamente apropriados ao contexto de aplicação (art. 13, I), assegurando que os resultados reflitam a pluralidade de situações existentes e evitando

³² *Precision* (ou precisão) é uma métrica de exatidão nas respostas positivas do modelo. Ela indica a proporção de previsões positivas que realmente são corretas.

³³ *Recall* é uma métrica que mede a proporção de casos positivos reais que foram corretamente indicados pelo modelo.

³⁴ *F1-score* é a média harmônica entre precisão e recall. Ela equilibra as duas métricas.

³⁵ *AUC-ROC* (ou área sob a curva ROC) é uma métrica que mede a capacidade geral do modelo de distinguir entre classes positivas e negativas, independentemente do limiar de decisão adotado.

distorções decorrentes de sub ou super-representação de determinados grupos sociais. O texto também impõe o uso de ferramentas de registro automático (*logs*) que possibilitem a avaliação periódica da acurácia e robustez, bem como a detecção de resultados discriminatórios e a implementação de medidas corretivas (art. 13, V).

Ademais, o art. 26 da mesma Resolução dispõe que os dados utilizados no desenvolvimento das soluções de IA devem ser objeto de curadoria de qualidade, garantindo precisão, integridade e confiabilidade, o que constitui pré-requisito técnico para que se obtenham modelos efetivamente acurados. A Resolução ainda exige auditorias contínuas (arts. 12, 14 e 41) e avaliações de impacto algorítmico com indicadores claros de desempenho, a fim de assegurar que as soluções mantenham sua eficácia e correção ao longo do tempo. Em síntese, o CNJ transforma a acurácia e a robustez em parâmetros normativos obrigatórios, vinculando o desempenho técnico à observância de princípios constitucionais de eficiência, transparência, isonomia e não discriminação (Conselho Nacional de Justiça, 2025b).

A importância da adoção de métricas de acurácia no contexto judicial vai além do controle técnico: trata-se de um instrumento de garantia da justiça substantiva. Avaliar e monitorar sistematicamente a acurácia permite detectar vieses e falhas de generalização que poderiam afetar grupos vulneráveis ou produzir decisões desproporcionais.

A acurácia deve ser avaliada também de forma desagregada por grupos sensíveis (gênero, raça, idade, entre outros), de modo a evidenciar se há discrepâncias no desempenho do modelo (Mehrabi et al., 2021). Essa abordagem, integrada à curadoria e à validação contínua, permite prevenir discriminações indiretas e fortalecer a equidade de resultados.

Sistemas robustos e acurados também mitigam vieses ao impedir que pequenas distorções nos dados se amplifiquem em predições discriminatórias. A robustez algorítmica protege o sistema contra ruídos e padrões espúrios – frequentemente correlacionados a desigualdades históricas –, reforçando a imparcialidade e a previsibilidade decisória (Hooker, 2021). Modelos que são capazes de resistir a variações estatísticas e identificar inconsistências em bases desbalanceadas tendem a produzir decisões mais justas e menos sujeitas a discriminações de gênero, raça ou classe social.

Em destaque, observa-se que a acurácia e a robustez dos sistemas de IA não apenas asseguram o desempenho técnico, mas constituem mecanismos centrais para o enfrentamento dos vieses de gênero nas decisões judiciais automatizadas. Um sistema acurado, capaz de avaliar de forma equilibrada os dados e corrigir desvios estatísticos, permite identificar quando o gênero está influenciando de modo indevido um resultado, possibilitando a revisão e o

reequilíbrio das inferências. Já a robustez atua como salvaguarda ética, impedindo que pequenas distorções ou correlações espúrias reforcem estereótipos e desigualdades históricas. Assim, ao combinar precisão técnica e estabilidade operacional, sistemas robustos e acurados fortalecem a justiça algorítmica, promovendo decisões mais previsíveis, transparentes e equitativas.

A conexão entre robustez, acurácia e os princípios constitucionais da eficiência e da isonomia é direta. O art. 37, caput, da Constituição Federal impõe a eficiência como princípio da administração pública, exigindo que os sistemas adotados pelos órgãos judiciais operem com precisão, confiabilidade e transparência (Brasil, 1988). A Resolução CNJ n. 615/2025 consolida essa diretriz ao afirmar que os sistemas de IA devem garantir justiça, equidade, inclusão e não discriminação (art. 3º, I), eficiência e qualidade na entrega da prestação jurisdicional (art. 3º, IV) e supervisão humana efetiva para garantir a correção e a confiabilidade das decisões (art. 3º, VII) (Conselho Nacional de Justiça, 2025c).

Dessa forma, a robustez e a acurácia dos sistemas de IA no Poder Judiciário não são apenas requisitos técnicos, mas condições de legitimidade institucional e de justiça algorítmica. Elas asseguram que as decisões automatizadas sejam auditáveis, previsíveis e justas, reforçando a confiança pública nas tecnologias judiciais e promovendo uma justiça mais eficiente, imparcial e equitativa. Como afirma Amodei et al. (2016), a segurança e a previsibilidade da IA dependem tanto do desenho técnico dos algoritmos quanto da capacidade institucional de supervisionar e corrigir continuamente seu desempenho – princípio que a Resolução n. 615/2025 transforma em obrigação concreta de governança (Conselho Nacional de Justiça, 2025c).

5.1.7 Segurança cibernética e segurança da informação

A cibersegurança constitui um dos pilares fundamentais da governança ética e técnica da IA no Poder Judiciário. Trata-se de um conjunto estruturado de práticas, técnicas e políticas voltadas à proteção da integridade, confidencialidade e disponibilidade dos sistemas digitais, assegurando que dados, modelos e infraestruturas tecnológicas permaneçam protegidos contra acessos indevidos, manipulações e ataques maliciosos. Em sistemas de IA, a cibersegurança assume caráter crítico, uma vez que tais sistemas frequentemente processam informações sensíveis, tomam decisões automatizadas e produzem efeitos diretos sobre direitos fundamentais.

Ataques cibernéticos contra sistemas de IA podem gerar consequências severas, que vão desde vazamentos de dados pessoais e manipulação de resultados até discriminação indevida e decisões incorretas com impactos sociais e jurídicos significativos. Vulnerabilidades técnicas em modelos de aprendizado de máquina podem ser exploradas por meio de ataques adversariais, nos quais pequenas alterações nos dados de entrada produzem distorções sistemáticas nas saídas do modelo. Em contextos judiciais, tais interferências podem comprometer a imparcialidade decisória e a credibilidade institucional.

Nesse sentido, os sistemas de IA do Poder Judiciário devem observar requisitos mínimos de segurança cibernética, incluindo:

- a)** avaliação e monitoramento contínuo de vulnerabilidades, com auditorias técnicas periódicas;
- b)** uso de criptografia e anonimização de dados pessoais, conforme os princípios da LGPD (arts. 46 e 47), que determinam a adoção de medidas técnicas e administrativas aptas a proteger os dados contra acessos não autorizados, destruição acidental, perda, alteração ou qualquer forma de tratamento inadequado (Brasil, 2018a);
- c)** mecanismos de detecção e resposta rápida a incidentes, mediante sistemas de monitoramento inteligente e protocolos de contingência;
- d)** atualização regular dos softwares e componentes utilizados, garantindo proteção contra ameaças emergentes;
- e)** robustez contra ataques adversariais e envenenamento de dados, preservando a integridade dos modelos e dos conjuntos de treinamento;
- e**
- f)** auditabilidade das decisões algorítmicas, permitindo rastrear *logs* e detectar comportamentos anômalos que possam indicar comprometimentos de segurança.

No contexto do Judiciário brasileiro, a Resolução n. 332/2020 do CNJ (Conselho Nacional de Justiça, 2020b), ao estabelecer princípios e diretrizes éticas para o uso da IA no sistema judicial, já previa que os modelos deveriam atender a critérios de transparência, previsibilidade, auditabilidade e garantia de imparcialidade, bem como critérios de segurança (vide arts. 13 a 15 da referida Resolução). Ainda, o art. 8º, inciso III, do referido normativo, reforçava a necessidade de documentação dos riscos identificados e indicação dos instrumentos

de segurança da informação adotados, aproximando o conceito de segurança digital da noção de responsabilidade institucional.

Já a Resolução CNJ nº 615/2025 amplia e aprofunda a regulamentação anterior ao inserir a cibersegurança no centro da política de governança digital do Poder Judiciário, reconhecendo a segurança cibernética como fundamento essencial para o uso da IA (art. 2º, XI) e a segurança da informação (juntamente com a segurança jurídica) como princípio basilar da ética e da governança digital (art. 3º, III) (Conselho Nacional de Justiça, 2025c).

O art. 7º dispõe sobre a cautela necessária em relação ao segredo de justiça e aos dados pessoais, ao passo que o art. 10 veda o desenvolvimento e a utilização de IA que acarrete risco excessivo à segurança da informação, prevenindo vulnerabilidades estruturais e falhas críticas. Já o art. 12 impõe a criação de processos internos para a garantia da segurança dos sistemas inteligentes, enquanto o art. 14 determina a realização de avaliações periódicas de risco e impacto algorítmico, que devem incluir a verificação de incidentes de segurança e a conformidade com padrões de proteção de dados pessoais (Conselho Nacional de Justiça, 2025c).

A Resolução n. 615/2025 estabelece um marco normativo abrangente e rigoroso para a proteção de dados, a integridade das informações e a segurança cibernética das soluções de IA. O texto reconhece que a confiança institucional e a legitimidade do uso da IA no sistema de justiça dependem da adoção de mecanismos que assegurem a qualidade, rastreabilidade e proteção integral dos dados utilizados em todo o ciclo de vida dos sistemas, desde o desenvolvimento até a execução e o descarte (Conselho Nacional de Justiça, 2025c).

O art. 26 determina que os dados empregados no desenvolvimento de soluções de IA devem ser preferencialmente provenientes de fontes públicas ou governamentais, submetidos a curadoria de qualidade que garanta precisão, integridade, equilíbrio e confiabilidade. O uso de dados não governamentais é permitido apenas mediante verificação rigorosa de qualidade e segurança, sempre em conformidade com a LGPD (Brasil, 2018a). Reforçando o princípio da minimização de dados, o dispositivo estabelece que apenas as informações estritamente necessárias ao treinamento dos modelos podem ser coletadas, vedando o armazenamento de conjuntos de dados sem controle ou finalidade definida (Conselho Nacional de Justiça, 2025c).

O art. 27 complementa esse dispositivo ao exigir que os sistemas adotem mecanismos de controle de versões, rastreabilidade e integridade dos dados, de modo a impedir alterações indevidas antes de sua utilização. Também impõe a manutenção de cópias seguras e criptografadas dos *datasets* relevantes, com acesso restrito e planos de descarte regulados,

garantindo auditabilidade e conformidade com as normas de proteção de dados. Essa estrutura assegura que, em caso de incidentes, os registros possam ser revisados e auditados, evitando manipulações e perdas de informações (Conselho Nacional de Justiça, 2025c).

O art. 28 trata do armazenamento e execução das soluções de IA, que podem ocorrer em *datacenters* próprios, serviços de nuvem ou APIs. O texto impõe o uso de mecanismos de isolamento, segregação e criptografia, garantindo que os dados do Judiciário não sejam acessados ou manipulados por terceiros não autorizados. Exige-se, ainda, que provedores contratados estejam em conformidade com a legislação brasileira e possuam certificações internacionais de segurança da informação, além de criptografia robusta e políticas transparentes de retenção e descarte de dados (Conselho Nacional de Justiça, 2025c).

No mesmo sentido, o art. 29 reforça que todos os dados utilizados no desenvolvimento e na execução das soluções de IA devem ser protegidos contra destruição, modificação, extravio e acessos não autorizados, mediante criptografia, controle de acesso, auditorias regulares e monitoramento contínuo. O artigo exige ainda a adoção de ferramentas proativas de detecção e resposta a incidentes, garantindo uma resposta ágil e eficaz a eventuais violações (Conselho Nacional de Justiça, 2025c).

O art. 30 inova ao proibir o compartilhamento de dados custodiados pelo Judiciário em plataformas, aplicativos ou APIs, salvo quando esses dados forem anonimizados ou pseudoanonimizados na origem, em conformidade com a LGPD. Essa regra protege a identidade dos titulares e impede a reidentificação de indivíduos, além de impor mecanismos de auditoria e controle contínuo para garantir conformidade com as normas de proteção de dados durante o treinamento ou a readequação de modelos de IA (Conselho Nacional de Justiça, 2025c).

O art. 31, por sua vez, consolida as diretrizes sobre segurança da informação nos ambientes de armazenamento e execução das soluções de IA, exigindo controles de acesso rigorosos, criptografia em repouso e em trânsito, bem como políticas de governança e gestão de vulnerabilidades. O dispositivo destaca também a necessidade de educação continuada das equipes em segurança da informação e privacidade, a eliminação de dados pessoais não anonimizados após o treinamento dos modelos e a curadoria constante da qualidade e da segurança dos dados (Conselho Nacional de Justiça, 2025b).

Essas disposições alinham-se à LGPD, que no art. 6º, VII, consagra o princípio da segurança, e no art. 46 impõe que controladores e operadores adotem medidas técnicas e administrativas aptas a proteger dados pessoais contra qualquer forma de acesso indevido ou

ilícito. No âmbito judicial, isso implica uma obrigação ampliada, visto que o Poder Judiciário atua como guardião de informações de alta sensibilidade, como dados processuais sigilosos, informações biométricas e registros de decisões judiciais automatizadas.

A segurança cibernética, portanto, transcende o campo técnico e se projeta como um dever ético, jurídico e social. Ela é componente essencial da responsabilidade institucional e da confiança pública nas soluções tecnológicas do sistema de justiça. Além das medidas técnicas, é indispensável promover capacitação contínua de magistrados, servidores e operadores do sistema judicial, para que compreendam os riscos associados ao uso de IA e saibam adotar condutas seguras em seu manejo.

A ausência de práticas adequadas de cibersegurança não apenas fragiliza a infraestrutura tecnológica, mas amplifica riscos de discriminação e violações de direitos fundamentais, especialmente quando vulnerabilidades são exploradas para manipular resultados, corromper dados sensíveis ou introduzir distorções nos modelos de decisão. A segurança digital, portanto, é uma condição prévia da equidade algorítmica: apenas sistemas protegidos e confiáveis podem garantir decisões justas, imparciais e livres de interferências indevidas.

Sob a perspectiva de gênero, a cibersegurança também desempenha um papel estratégico na mitigação de vieses estruturais. A proteção dos dados e a integridade dos modelos de IA asseguram que informações sensíveis – como gênero, raça ou condição socioeconômica – não sejam indevidamente acessadas, manipuladas ou exploradas de forma discriminatória. A ausência de medidas robustas de segurança pode permitir que atores externos alterem parâmetros ou conjuntos de treinamento, inserindo distorções que reforcem estereótipos de gênero ou excluam mulheres de representações justas nos sistemas.

Sob esse fundamento, reforça-se que o Poder Judiciário deve comprovar de forma efetiva a adoção de padrões satisfatórios de cibersegurança, demonstrando o cumprimento de protocolos técnicos, certificações reconhecidas internacionalmente e práticas de auditoria contínua que assegurem a integridade, a confidencialidade e a disponibilidade das informações judiciais. Tal comprovação é indispensável não apenas para atender às exigências normativas, mas também para preservar a confiança pública e a legitimidade institucional no uso da IA.

Em última análise, a cibersegurança é mais do que um requisito técnico: é um compromisso ético e democrático com a proteção de direitos fundamentais e a transparência na aplicação da tecnologia. Garantir a cibersegurança é também garantir equidade e proteção contra discriminação algorítmica, pois impede que vulnerabilidades técnicas se convertam em

injustiças sociais. Ao blindar os sistemas contra interferências maliciosas, o Poder Judiciário preserva a imparcialidade de suas decisões automatizadas e reforça seu compromisso com a igualdade de gênero e os direitos humanos.

5.1.8 Compreensão dos limites da linguagem natural

O uso de técnicas de aprendizagem de máquina no campo do Direito exige reconhecer e compreender os limites da linguagem natural, isto é, a linguagem falada e escrita cotidianamente pelos seres humanos. É por meio dela que o ordenamento jurídico é construído, que as normas são interpretadas e que as decisões judiciais são formuladas. Como observa Boeing (2019, p. 33), “é através da linguagem natural que o ser humano se comunica e, portanto, é com ela que se concebe o fenômeno jurídico”. Assim, qualquer sistema de IA que opere sobre textos jurídicos – sejam leis, petições, sentenças ou acórdãos – precisa lidar com as ambiguidades, polissemias e sutilezas próprias da linguagem humana.

No contexto da IA aplicada à justiça, essa limitação tem consequências diretas sobre a precisão e a imparcialidade das decisões automatizadas. A linguagem natural, por sua natureza, é aberta, interpretativa e dependente de contexto. Palavras e expressões jurídicas podem ter significados distintos conforme o caso, o tempo, o tribunal ou a doutrina. Mesmo para os operadores do Direito (juristas, magistrados, promotores, advogados, defensores, assessores jurídicos etc.), é comum que um mesmo dispositivo legal produza interpretações diversas. Assim, tanto a atividade jurisdicional conduzida por humanos quanto aquela mediada por sistemas de IA está sujeita à imprecisão semântica e pragmática da linguagem, o que pode gerar resultados diferentes para situações semelhantes.

Os modelos de linguagem natural (Large Language Models – LLMs), amplamente utilizados na atualidade, buscam interpretar e reproduzir o discurso humano com base em grandes volumes de textos. Contudo, eles não compreendem o significado jurídico em sentido profundo; apenas identificam padrões estatísticos de coocorrência linguística. Isso significa que o raciocínio jurídico automatizado pode reproduzir vieses, distorções ou interpretações equivocadas, especialmente quando os dados de treinamento refletem desigualdades estruturais ou omitem grupos historicamente marginalizados. Nesse sentido, compreender os limites da linguagem natural é também reconhecer que as IAs não interpretam o Direito, mas apenas o representam de forma probabilística.

Evitar equívocos de linguagem e interpretação é, portanto, essencial para mitigar vieses algorítmicos no campo judicial. Um sistema de IA que não leva em conta a polissemia dos termos jurídicos pode, por exemplo, associar de forma incorreta conceitos relacionados ao gênero, à raça ou à classe social, reforçando estereótipos ou desigualdades já presentes nos dados textuais. Além disso, a ausência de revisão humana qualificada sobre as saídas desses modelos pode transformar erros semânticos em decisões injustas. Diante disso, torna-se imprescindível que magistrados, servidores e desenvolvedores compreendam as limitações linguísticas da IA, promovendo revisões interpretativas, ajustes contextuais e curadorias constantes dos dados e resultados.

A mitigação de vieses de gênero, nesse contexto, depende da alfabetização linguística e semântica dos sistemas inteligentes. LLMs devem ser treinados com bases de dados amplas, diversificadas e balanceadas, que implementem e efetivamente representem adequadamente a pluralidade social e de gênero no âmbito jurídico. Além disso, deve-se adotar metodologias de explicabilidade e auditoria textual que permitam identificar como e por que certos termos, expressões ou contextos estão sendo interpretados de maneira enviesada. A compreensão e o controle desses limites linguísticos favorecem a transparência, a correção semântica e a imparcialidade das decisões automatizadas, fortalecendo a confiança social na IA judicial.

Portanto, reconhecer os limites da linguagem natural não é apenas uma questão técnica, mas uma exigência epistemológica e ética. No âmbito do Poder Judiciário, isso significa adotar uma postura crítica diante da automatização da linguagem jurídica, garantindo que a tecnologia seja usada para ampliar o acesso à justiça e não para cristalizar desigualdades. Compreender o alcance e as limitações da linguagem natural é, assim, um passo necessário para a construção de uma IA justa, responsável e sensível às nuances do discurso jurídico e às diversidades humanas que ele representa.

5.1.9 Personalização e atualização constante dos sistemas inteligentes

A incorporação da IA ao campo jurídico impõe o desafio de conciliar a lógica matemática³⁶ dos algoritmos com o discernimento subjetivo e interpretativo dos juízes

³⁶ Além da lógica matemática, atualmente é preciso considerar outras bases fundamentais para a IA, tal como defendido nessa tese. É preciso levar em conta como fundamento básico da IA no Poder Judiciário:

1. Ética: é fundamental garantir que sistemas inteligentes operem dentro de padrões éticos claros, evitando discriminações, vieses e danos sociais.
2. Direito e regulação: a IA exige fundamentação jurídica sólida, com normas capazes de orientar e responsabilizar o uso desses sistemas, protegendo direitos fundamentais e prevenindo abusos.

humanos. Enquanto os magistrados decidem com base em princípios, valores e experiências, os sistemas de IA operam segundo padrões estatísticos e relações probabilísticas.

Como observa Fonseca (2022, p. 301-302), à medida que os algoritmos judiciais se tornam mais complexos, suas decisões também se tornam mais personalizadas, uma vez que passam a considerar um número maior de detalhes e variáveis contextuais. No entanto, essa personalização só será legítima se for acompanhada de atualização constante, supervisão humana qualificada e mecanismos de revisão e correção de erros, garantindo que os sistemas permaneçam fiéis à evolução do Direito e às transformações sociais.

A Resolução n. 615/2025 do CNJ (Conselho Nacional de Justiça, 2025b) reconhece essa necessidade ao estabelecer, em seu art. 5º, §2º e parágrafo único, que os sistemas de IA devem ser periodicamente revisados e aprimorados, assegurando que se mantenham em conformidade com as normas legais, com a ética pública e com os direitos fundamentais. Os arts. 16, X e 18, II, determinam indiretamente que os modelos de IA empregados no Poder Judiciário sejam continuamente avaliados quanto à precisão técnica e ao impacto ético e social de suas decisões.

A Resolução prevê o monitoramento contínuo e o ajuste dinâmico das soluções de IA, com base em relatórios de desempenho e resultados observados na prática judicial (Conselho Nacional de Justiça, 2025c). No entanto, o CNJ não especifica, de modo detalhado, como esses processos devem ocorrer na prática, deixando lacunas quanto aos critérios técnicos, à periodicidade e à metodologia de auditoria e readequação. Essa indefinição pode comprometer a efetividade do monitoramento, dificultando a detecção de erros sistemáticos ou de vieses reproduzidos nas decisões automatizadas.

A interpretação jurídica, por natureza, é dinâmica e histórica. O conteúdo normativo e jurisprudencial muda ao longo do tempo, refletindo transformações sociais, políticas e culturais. Nesse sentido, os sistemas de IA aplicados ao Judiciário não podem permanecer estáticos, sob pena de fossilizarem o Direito e perpetuarem entendimentos desatualizados.

3. Explicabilidade e Transparência: garantir que a operação e as decisões dos sistemas sejam compreensíveis e auditáveis, facilitando questionamentos e revisões por parte das pessoas afetadas.

4. Responsabilidade (Accountability): quem desenvolve e opera IA precisa assumir responsabilidade clara pelos impactos sociais, econômicos e éticos gerados pelas decisões automatizadas.

5. Cybersegurança: é crucial assegurar proteção contra ataques digitais, garantindo a segurança dos dados e a confiança nos sistemas inteligentes.

6. Equidade (Fairness): garantir que a IA trate diferentes indivíduos e grupos sociais de forma justa, prevenindo discriminações e exclusões.

7. Aspectos sociais e culturais: levar em consideração o contexto social, político e cultural no desenvolvimento e uso da IA, respeitando as particularidades locais.

Esses elementos são indispensáveis para que a IA seja desenvolvida e utilizada de maneira socialmente responsável e sustentável no contexto atual.

Embora a padronização algorítmica possa favorecer a segurança jurídica, ela também pode limitar o desenvolvimento da jurisprudência e restringir a criatividade judicial, transformando a IA em um instrumento de reprodução mecânica de decisões. Por isso, a personalização e a atualização constante dos sistemas são fundamentais para assegurar que as decisões automatizadas acompanhem a evolução do pensamento jurídico e os valores contemporâneos da sociedade.

As atualizações periódicas dos sistemas de IA também exercem papel importante na correção e mitigação de vieses. À medida que novos dados são incorporados, é possível reequilibrar conjuntos de treinamento, eliminar correlações discriminatórias e ajustar parâmetros de inferência que antes produziam resultados enviesados. Por exemplo, um sistema que reproduz padrões de desigualdade de gênero em decisões sobre guarda de filhos ou progressão de regime prisional pode ser reprogramado e recalibrado após auditorias, de modo a eliminar tais distorções. Além disso, as atualizações permitem incorporar avanços metodológicos, métricas de *fairness* e práticas de governança mais modernas, garantindo que a IA atue de forma mais justa, transparente e representativa. Dessa maneira, o ciclo de atualização não apenas mantém os sistemas tecnicamente eficazes, mas também corrige suas distorções éticas e sociais, reforçando a legitimidade das decisões automatizadas.

A personalização dos sistemas também é essencial para evitar a padronização excessiva das decisões judiciais e o consequente risco de julgamentos injustos. É necessário que os algoritmos sejam sensíveis às circunstâncias específicas de cada processo, considerando fatores regionais, sociais e humanos. Essa sensibilidade deve incluir a adaptação dos sistemas aos diferentes contextos socioculturais e jurídicos do Brasil, reconhecendo que as realidades enfrentadas pelos tribunais variam significativamente entre regiões. Uma IA utilizada no TJSP, por exemplo, deve operar com parâmetros distintos de uma IA empregada no TJAC, em razão das diferenças de volume processual, perfil das demandas e contextos sociais.

Por fim, a personalização e a atualização contínua das soluções de IA não são apenas requisitos técnicos, mas garantias éticas e democráticas. Elas asseguram que a tecnologia não substitua o raciocínio humano, mas o complemente, preservando o espaço da interpretação, da equidade e da humanidade nas decisões judiciais.

A IA no Poder Judiciário deve ser capaz de evoluir junto com o Direito e com a sociedade, incorporando novos valores, corrigindo injustiças e fortalecendo a confiança pública. A falta de atualização, ao contrário, pode cristalizar erros, automatizar discriminações e comprometer a legitimidade do sistema de justiça. Portanto, investir em atualização constante

e personalização inteligente é investir na justiça algorítmica enquanto prática viva, adaptativa e inclusiva.

5.1.10 Interpretabilidade e auditabilidade dos sistemas inteligentes e prestação de contas (accountability) quanto ao seu desenvolvimento e uso

A adoção de sistemas de IA no Poder Judiciário exige que interpretabilidade, auditabilidade e *accountability* sejam tratadas como garantias estruturantes e não como meros atributos desejáveis. A interpretabilidade significa tornar inteligível a lógica que conduz dos dados de entrada aos resultados produzidos; a auditabilidade assegura que processos, dados, versões de modelos e decisões sejam rastreáveis e verificáveis por instâncias internas e independentes; e a *accountability* (prestação de contas) impõe a obrigação de responder por atos e omissões em todo o ciclo de vida do sistema – do desenho e desenvolvimento à implementação e ao uso cotidiano. Sem essas três camadas, não se protegem direitos humanos e direitos fundamentais diretamente envolvidos no processo judicial, como o acesso à justiça, a motivação das decisões, a ampla defesa, o contraditório e o devido processo legal.

Essas exigências encontram respaldo direto na Constituição Federal (Brasil, 1988), que estabelece no art. 37 o dever de publicidade, moralidade e eficiência como princípios da administração pública, e no art. 93, IX, a obrigatoriedade de fundamentação das decisões judiciais. Da mesma forma, a LGPD (Brasil, 2018a), em seus arts. 6º, X e 46, impõem a adoção de medidas técnicas e administrativas que assegurem a transparência, a responsabilização e a prestação de contas (*accountability*) dos agentes de tratamento, princípios que devem ser integralmente observados no desenvolvimento e uso de sistemas de IA pelo Poder Judiciário.

Sistemas de IA devem ser auditáveis, garantindo responsabilidade pelo desenvolvimento, implementação e resultados. A auditabilidade possibilita examinar dados utilizados, técnicas/algoritmos, hiperparâmetros, versões de modelos e métricas de desempenho, permitindo avaliar a justeza (*fairness*) e o potencial discriminatório do sistema. Quando falamos em prestação de contas (*accountability*), referimo-nos à obrigação de responder por atos praticados e assumir consequências jurídicas e éticas decorrentes das ações ou omissões – inclusive quando automatizadas. Como observa Salomão (2022, p. 32), avaliar a *accountability* de um modelo é identificar se ele toma decisões responsáveis conforme parâmetros predefinidos e publicamente controláveis.

As estratégias para tornar os sistemas de IA interpretáveis envolvem múltiplas frentes complementares. Uma delas consiste em simplificar os modelos ou adotar métodos específicos de avaliação da capacidade discriminatória, de modo a identificar e corrigir vieses nos dados e nos resultados. Outras abordagens defendem a criação de mecanismos regulatórios e de transparência reforçada, especialmente diante de algoritmos opacos cujo funcionamento interno não é facilmente compreensível (Monteiro, 2021, p. 220).

Também se propõe a realização de auditorias técnicas independentes, conduzidas por especialistas externos, a fim de garantir imparcialidade na verificação da equidade e do desempenho do sistema. Além disso, diversos autores salientam a importância de promover a educação digital e ética de amplos segmentos da sociedade, capacitando cidadãos, operadores do direito e gestores públicos a reconhecer e avaliar o potencial discriminatório e os impactos sociais dos sistemas inteligentes (Monteiro, 2021, p. 220).

A responsabilidade, por sua vez, demanda mecanismos efetivos de prevenção e reparação de danos, o que pressupõe transparência: sistemas devem explicar o que fazem e justificar por que determinado resultado foi proposto. Em outras palavras, os sistemas devem ter a capacidade de explicar suas ações às pessoas e capaz de justificar por que uma decisão específica foi tomada.

No âmbito normativo, a Resolução n. 615/2025 do CNJ positivou esse tripé (Conselho Nacional de Justiça, 2025c). Logo no art. 1º, faz-se referência à auditoria e monitoramento das soluções de IA empregadas no Poder Judiciário. No art. 2º (fundamentos) e no art. 3º (princípios) estabelece a contestabilidade, a auditabilidade e a transparência como eixos da ética digital e da justiça algorítmica.

Quanto aos vieses, o art. 8º, §1º observa a necessidade de se evitar o surgimento de vieses discriminatórios, o que torna imperiosa a validação contínua das soluções de IA e a auditoria ou monitoramento de suas decisões ao longo de todo o ciclo de vida da aplicação (Conselho Nacional de Justiça, 2025c).

Os arts. 11 e 14 indicam a necessidade de autorias, monitoramentos e avaliações periódicas de risco e impacto algorítmico, com verificação de incidentes e conformidade com proteção de dados, enquanto os arts. 22 a 25 criam um sistema de registro e publicidade das soluções (cadastro na Sinapses, finalidades, grau de risco, responsáveis), e o art. 39 consagra a “total transparência na prestação de contas”. Em conjunto, tais dispositivos convertem a interpretabilidade em dever institucional, viabilizam a auditoria técnica e social e dão conteúdo operacional à *accountability* (Conselho Nacional de Justiça, 2025c).

Ao tratar sobre medidas de governança, o art. 12, VIII da Resolução (Conselho Nacional de Justiça, 2025c) garante acesso à Ordem dos Advogados do Brasil (OAB), à advocacia pública, ao Ministério Público e às Defensorias, conforme o caso, aos relatórios de auditoria e monitoramento e à parametrização ao longo do ciclo de vida da solução que envolver o uso de IA.

Além dos dispositivos já mencionados, os arts. 39 a 42 consolidam o dever de transparência e prestação de contas. O art. 39 determina que toda solução de IA deve assegurar “total transparência na prestação de contas”, que deve incluir os responsáveis, custos de desenvolvimento, parcerias públicas ou privadas, resultados pretendidos e alcançados, bem como a publicidade acessível sobre as técnicas empregadas, o desempenho do sistema e seus riscos de erro. Além disso, as informações devem ser divulgadas em formato claro e acessível, com atualizações regulares e possibilidade de interação pública, podendo ainda ser submetidas a auditoria externa determinada pelo tribunal ou pelo Comitê Nacional de IA do Judiciário (Conselho Nacional de Justiça, 2025c).

Já os arts. 40 a 42 disciplinam o monitoramento, a auditoria e a gestão de incidentes. O art. 40 prevê que o Comitê Nacional de IA fiscalizará o uso de sistemas em desacordo com a norma, podendo indicar auditorias sobre práticas inadequadas, uso indevido de dados ou falta de transparência. O art. 41 estabelece a criação de protocolos formais de auditoria e monitoramento, conduzidos por comissões técnicas qualificadas, responsáveis por verificar riscos, salvaguardas e conformidade das soluções de IA, com prazos para correção de falhas. Por fim, o art. 42 impõe o dever de comunicação obrigatória de incidentes ou eventos adversos relacionados ao uso da IA, em até 72 horas após sua identificação, assegurando resposta imediata e correção dos problemas. (Conselho Nacional de Justiça, 2025c).

Essas garantias são decisivas para assegurar a não discriminação algorítmica. Modelos opacos dificultam detectar tratamentos desiguais entre grupos (por exemplo, mulheres e homens), enquanto trilhas de auditoria (dados de treino validados, versões de modelos, justificativas de previsões, métricas de *fairness* e de erro segmentadas) permitem identificar e corrigir padrões de erro e viés indireto.

A interpretabilidade torna visíveis as variáveis e relações ponderadas pelo modelo; a auditabilidade permite reconstituir o processo decisório, comparar saídas entre grupos sensíveis e aferir paridade/igualdade de desempenho; e a *accountability* vincula pessoas e instituições às consequências das decisões automatizadas, incentivando correções de rota, reparação de danos e melhoria contínua. Sem interpretabilidade e auditoria, não há como demonstrar a justeza, a

proporcionalidade e a motivação das decisões assistidas por IA; sem *accountability*, não há como atribuir responsabilidade por omissões, negligências ou projetos tecnicamente deficientes.

A literatura converge nesse ponto: é “imperativo que sistemas inteligentes atendam a regularidade procedimental” (Monteiro, 2021, p. 103). Diante da crescente automação decisória, Citron e Pasquale (2014, p. 19) defendem cercar os algoritmos de *accountability*, enfrentando a opacidade algorítmica por meio de devido processo informacional e tecnológico – isto é, regras que submetam modelos preditivos à revisão quanto à justeza e à acurácia. A responsabilidade pode exigir rastreabilidade fina para distribuir deveres entre programadores, designers, integradores, fornecedores e magistrados (Caballero Trenado, 2019, p. 369).

Contudo, como lembram Cárdenas e Molano (2021, p. 16, tradução e adaptação nossas³⁷), em consonância com a Declaração de Montreal (Montréal Declaration for a Responsible Development of Artificial Intelligence, 2018), a responsabilidade última não deve ser transferida para a IA:

1. apenas humanos podem ser responsáveis por decisões derivadas de ferramentas de IA;
2. decisões que afetem vida, qualidade ou reputação devem ser tomadas por humanos, de forma livre e informada;
3. decisões sobre a vida não podem ser delegadas;
4. quem autoriza o uso da IA para cometer ilícitos age com negligência; e
5. quando a IA é confiável e usada conforme o previsto, não é razoável culpar desenvolvedores ou usuários se, ainda assim, ocorrer dano – reforçando a necessidade de governança, supervisão e trilhas de decisão em todo o ciclo.

No plano operativo, a prestação de contas exige procedimentos e evidências: políticas de governança publicadas, matrizes de responsabilidade, relatórios de impacto algorítmico, inventários de dados e suas fontes, métricas de desempenho e *fairness* (com recortes por grupos sensíveis), registros de treinamento e validação, *logs* de inferência, relatórios de incidentes, planos de mitigação e cronogramas de atualização.

³⁷ No original: “1. Solo los seres humanos pueden ser responsables de las decisiones que 1. provienen de las recomendaciones formuladas por herramientas de inteligencia artificial y de las acciones que de ellas se originen. 2. En la totalidad de las áreas donde se debe tomar una decisión que puede afectar la vida, calidad o reputación de una persona, la decisión final debe ser tomada por un ser humano, siendo esta una decisión libre e informada. 3. Las decisiones que atenten contra la vida deben ser tomadas siempre por seres humanos, sin que se pueda transferir la responsabilidad de tal decisión a la inteligencia artificial. 4. Las personas que autorizan a las herramientas de inteligencia artificial a cometer delitos actúan con negligencia al permitir que la IA los cometa, así, son los responsables de tal delito. 5. Cuando la inteligencia artificial es confiable y se ha utilizado según lo previsto, pero aun así genera daños, no es razonable culpar a las personas involucradas en su desarrollo o uso”.

Boas práticas incluem auditorias internas (contínuas) e independentes, com escopos técnicos e jurídico-éticos; comitês interdisciplinares; e mecanismos de contestação acessíveis aos jurisdicionados e defensores. Sistemas responsáveis e *accountables* exigem treinamento adequado dos operadores, definição prévia de responsabilidades, procedimentos para identificação e mitigação de falhas e vias formais de contestação do uso e dos resultados (Accessnow.Org, 2018, p. 33–34).

No caso do Poder Judiciário brasileiro, a auditabilidade deve ser implementada de forma aberta, sistemática e eficiente. Isso implica que os tribunais disponibilizem publicamente relatórios de auditoria algorítmica, descrevendo a metodologia empregada, os dados testados, os parâmetros avaliados, as métricas de desempenho e as medidas de mitigação adotadas. Além disso, deve haver integração com o princípio constitucional da publicidade e com o dever de transparência previsto na LGPD, assegurando que tais informações sejam acessíveis à sociedade civil, às partes processuais e aos órgãos de controle. A abertura desses relatórios amplia o controle social, reforça a legitimidade institucional e permite que a sociedade conheça as limitações, riscos e resultados dos sistemas inteligentes utilizados pelo Judiciário.

Por fim, é imperiosa a publicidade da prestação de contas. A Resolução n. 615/2025 (Conselho Nacional de Justiça, 2025c), ao articular transparência, explicabilidade, auditabilidade e *accountability*, converge com a exigência democrática de controle social: relatórios, catálogos públicos, linguagem acessível, indicadores e documentação verificável.

A publicidade efetiva das auditorias e relatórios técnicos é requisito indispensável para que a auditabilidade cumpra seu papel democrático e de proteção aos direitos fundamentais, transformando a IA em um instrumento legítimo de justiça e não em uma “caixa-preta” institucional. Essa publicidade – que não é meramente formal – viabiliza escrutínio externo, reduz assimetrias informacionais, dissuade práticas discriminatórias e fortalece a confiança da sociedade. Interpretáveis, auditáveis e responsáveis, os sistemas de IA no Judiciário deixam de ser “caixas-pretas” e passam a integrar, com legitimidade, a arquitetura do devido processo e da igualdade no Estado Democrático de Direito.

5.1.11 Quadro resumo das garantias técnicas para uma inteligência artificial ética e feminista no Poder Judiciário brasileiro

Quadro 6. Garantias técnicas para uma IA ética, confiável, segura e alinhada aos direitos humanos e à justiça social.

GARANTIAS TÉCNICAS PARA UMA IA ÉTICA NO JUDICIÁRIO	DESCRIÇÃO CONCEITUAL E TÉCNICA
<i>Fairness</i> (justiça ou equidade algorítmica)	Conjunto de métricas que assegura decisões não discriminatórias, garantindo tratamento equânime entre grupos sensíveis (gênero, raça, idade etc.). Engloba paridade demográfica, igualdade de oportunidades, probabilidades igualadas e paridade de previsão, dentre outras métricas. É subjetivo, político e dependente dos valores sociais, exigindo equipes diversas e análises profundas, pois métricas técnicas isoladas são insuficientes para eliminar preconceitos estruturais.
Transparência	Capacidade do público de compreender facilmente como os sistemas de IA funcionam e tomam decisões. A transparência evita “caixas-pretas”, aumenta a confiança pública e envolve relatórios claros, códigos abertos e auditorias independentes frequentes.
Explicabilidade	Significa que os sistemas de IA devem demonstrar claramente como tomam decisões, permitindo aos usuários contestá-las. Vai além da mera transparência, exigindo justificativas compreensíveis, úteis e relevantes. Está relacionada diretamente ao direito à informação e à prevenção de práticas discriminatórias ou abusivas. Integra o campo da <i>Explainable AI (XAI)</i> .
Diversidade e não discriminação	Assegura pluralidade nas equipes de desenvolvimento e representatividade nos dados de treinamento; combate homogeneidade técnica e cultural.
<i>Design</i> acessível	O <i>design</i> dos sistemas de IA deve garantir que todos possam utilizá-los, independentemente de idade, deficiência ou status social, assegurando acessibilidade universal.
Acurácia e robusteza	Acurácia: grau de correção das previsões. Robustez: estabilidade e segurança do sistema diante de ruídos e ataques. Sistemas de IA devem ser confiáveis e seguros, capazes de classificar informações corretamente e de fazer

	previsões ou decisões precisas com base nos dados e modelos utilizados.
Segurança cibernética e da informação	Conjunto de práticas para proteger dados e sistemas contra acessos indevidos, vazamentos e ataques adversariais. A IA deve conter requisitos mínimos para prevenir ataques cibernéticos, protegendo a segurança dos dados e a integridade dos sistemas contra invasões e manipulações externas.
Compreensão dos limites da linguagem natural	Reconhece que o uso da IA no campo do Direito enfrenta desafios pela imprecisão inerente à linguagem humana, o que pode levar a interpretações divergentes do fenômeno jurídico.
Personalização e atualização contínua	Revisão periódica dos modelos e adaptação a novos contextos legais, sociais e regionais; incorpora correções de viés e avanços metodológicos. É necessário atualizar constantemente os algoritmos, acompanhando mudanças jurídicas e sociais para evitar decisões rígidas e desumanizadas.
Interpretabilidade, auditabilidade e prestação de contas (<i>accountability</i>)	Interpretabilidade: tornar compreensível o raciocínio do sistema. Auditabilidade: permitir rastreamento e verificação independente. <i>Accountability</i> : assegurar responsabilidade e prestação de contas. Sistemas inteligentes devem permitir auditorias independentes, garantindo que sejam explicáveis, responsáveis e rastreáveis. Envolve treinamento adequado dos operadores, procedimentos de responsabilização claros, transparência decisória e mecanismos eficazes para contestar e corrigir decisões injustas ou incorretas.

Fonte: a autora.

As garantias técnicas e éticas formam um ecossistema interdependente: *fairness* mede a equidade; transparência e explicabilidade tornam as decisões visíveis e compreensíveis; diversidade e *design* acessível democratizam o desenvolvimento e o uso; acurácia e robustez asseguram precisão e confiabilidade; cibersegurança protege a integridade e a privacidade dos dados; atualização e interpretabilidade mantêm o sistema vivo, auditável e responsável.

Juntas, elas mitigam vieses de gênero em múltiplos níveis – desde a coleta e representação dos dados até a inferência, a supervisão humana e a prestação de contas –, construindo uma IA judicial justa, inclusiva e eticamente orientada.

5.2 Garantias regulatórias para uma inteligência artificial ética e feminista no Poder Judiciário brasileiro

As garantias regulatórias correspondem ao conjunto de instrumentos jurídicos, normativos e institucionais destinados a assegurar que o desenvolvimento, a implementação e o uso da IA no Poder Judiciário brasileiro ocorram em conformidade com os direitos humanos, os princípios constitucionais e as diretrizes de ética pública. Diferentemente das garantias técnicas, que se concentram na estrutura operacional e algorítmica dos sistemas, as garantias regulatórias estabelecem o arcabouço jurídico e procedimental que orienta sua legitimidade e controle democrático. Elas constituem, portanto, o eixo normativo da governança da IA judicial, definindo os parâmetros éticos, jurídicos e administrativos que vinculam o uso da tecnologia à proteção da dignidade humana, à igualdade de gênero e ao devido processo legal.

Essas garantias abrangem dimensões complementares: o reconhecimento dos direitos humanos como fundamento universal para a regulação tecnológica; a consolidação de uma ética da IA, pautada pela transparência, pela não discriminação e pela responsabilidade institucional; o fortalecimento da governança de dados, assegurando qualidade, segurança e uso responsável das informações; e a construção de um marco regulatório robusto, com aprimoramento da Resolução n. 615/2025 do CNJ (Conselho Nacional de Justiça, 2025c).

No âmbito do Poder Judiciário, as garantias regulatórias também se articulam com princípios estruturantes do ordenamento jurídico, como o devido processo legal, o juiz natural e a fundamentação das decisões judiciais, que funcionam como barreiras contra a delegação indevida da decisão judicial a sistemas automatizados e como salvaguardas da autonomia e da imparcialidade da magistratura. Do mesmo modo, instrumentos específicos, como o Protocolo para Julgamento com Perspectiva de Gênero, integram esse conjunto de garantias ao promover a incorporação de critérios de justiça de gênero nos processos decisórios e na interpretação jurídica.

Assim, as garantias regulatórias não se limitam ao cumprimento formal de normas, mas constituem um pilar ético e jurídico indispensável para a confiabilidade e a legitimidade da IA aplicada à justiça. Quando orientadas pela perspectiva feminista e pelos valores

constitucionais de igualdade e dignidade da pessoa humana, elas tornam-se instrumentos de prevenção e correção de vieses estruturais, assegurando que o avanço tecnológico não reproduza as assimetrias históricas do sistema de justiça, mas contribua para uma transformação institucional inclusiva, transparente e democrática.

5.2.1 O papel dos direitos humanos

A incorporação da IA ao campo jurídico inaugura novos desafios éticos e políticos, entre os quais se destaca a necessidade de assegurar que o desenvolvimento e o uso dessas tecnologias estejam integralmente alinhados aos direitos humanos. A IA, enquanto ferramenta de poder e decisão, tem potencial para aprofundar desigualdades e gerar novas formas de opressão social, afetando de maneira desproporcional os grupos mais vulneráveis e hipossuficientes.

Nesse sentido, os princípios de direitos humanos funcionam como o alicerce normativo das garantias éticas anteriormente discutidas – como a *fairness*, a transparência, a explicabilidade, a robustez e a *accountability* –, pois são eles que conferem densidade jurídica aos valores de equidade, dignidade e justiça social que orientam o desenvolvimento ético da IA.

O conceito de direitos humanos aborda disparidades de poder, servindo de instrumento para indivíduos e organizações que os representam contestarem as ações de atores mais poderosos, como Estados e grandes corporações. Os direitos humanos são inalienáveis e vinculativos, estando codificados em uma série de documentos internacionais, incluindo a Declaração Universal dos Direitos Humanos (Organização das Nações Unidas, 1948).

O respeito aos direitos humanos é uma obrigação de todos; isso inclui entes públicos e atores privados, tais como governos e empresas, como estabelecido no artigo 1º da Declaração Universal: “todos os seres humanos nascem livres e iguais em dignidade e direitos”. Governos têm a responsabilidade adicional de proteger e garantir os direitos humanos, conforme estipulado no artigo 3º, que declara que “todo indivíduo tem direito à vida, à liberdade e à segurança pessoal” (Organização das Nações Unidas, 1948).

Esses dispositivos impõem ao Estado o dever de assegurar que qualquer inovação tecnológica, especialmente aquelas aplicadas à administração da justiça, respeite a dignidade humana e não produza discriminação direta ou indireta. Assim, o desenvolvimento e a adoção de sistemas de IA no Poder Judiciário devem ser acompanhados de salvaguardas normativas e

operacionais que assegurem a igualdade de gênero, a imparcialidade e o controle humano sobre as decisões automatizadas.

Um amplo sistema de entidades regionais, internacionais e nacionais está em vigor para oferecer soluções e aplicar a legislação dos direitos humanos a contextos em evolução, inclusive diante dos avanços tecnológicos. A proteção dos direitos humanos, com base nos princípios fundamentais da igualdade, liberdade e dignidade, pode ajudar a mitigar os danos sociais causados pela IA e a prevenir futuras violações.

No contexto judicial, isso significa que os sistemas de IA devem ser avaliados sob a ótica dos direitos humanos antes, durante e após sua implementação, garantindo que não reproduzam padrões discriminatórios ou reforcem as desigualdades de gênero presentes nos próprios bancos de dados e decisões históricas da Justiça.

Vale dizer que direitos humanos, via de regra, são mais abrangentes e definidos do que princípios éticos, além de preverem questões relacionadas à responsabilização de agentes e à reparação para eventuais danos decorrentes da sua violação. Dessa forma, direitos humanos e princípios éticos podem (e devem) se complementar mutuamente quando se pretende o desenvolvimento e o uso de uma IA segura (Accessnow.Org, 2018, p. 17). Enquanto os princípios éticos orientam o comportamento moral e técnico dos sistemas de IA, os direitos humanos estabelecem obrigações jurídicas e mecanismos de reparação em caso de violação, transformando a ética em dever institucional e não apenas em escolha moral.

A integração entre direitos humanos e a ética da IA pode ser exemplificada da seguinte forma: uma organização pode adotar diretrizes éticas para o uso de IA que incluam a prevenção de preconceitos sociais e a garantia de que seus sistemas possam ser supervisionados e responsabilizados. Direitos humanos como o da privacidade e o da não discriminação podem fortalecer essas diretrizes, enquanto o regime internacional de proteção aos direitos humanos oferece mecanismos para reagir a eventuais violações. Ademais, quando o uso de uma IA se mostra contrário a princípios éticos, é provável que também configure uma violação de direitos humanos – situação em que os instrumentos e procedimentos do direito internacional podem ser acionados para conter e corrigir práticas indevidas. (Accessnow.Org, 2018, p. 17).

No caso brasileiro, essa integração é particularmente relevante diante da composição desigual do Poder Judiciário, marcada pela sub-representação feminina e por assimetrias institucionais que podem se refletir nas decisões e nas bases de dados judiciais. Incorporar os princípios de direitos humanos ao desenvolvimento e à governança da IA no Judiciário significa criar mecanismos concretos de prevenção de vieses de gênero – por exemplo, exigindo

diversidade nas equipes técnicas, curadoria ética dos dados e auditorias periódicas voltadas à detecção de padrões discriminatórios.

Os direitos humanos devem servir de parâmetro para a identificação de riscos no desenvolvimento e utilização de sistemas de IA. Por sua vez, uma IA confiável e ética deve respeitar a dignidade humana, a liberdade individual, a democracia, a justiça e o Estado de Direito. Deve atender aos princípios da equidade, da não discriminação, da solidariedade, bem como observar direitos das minorias (High-Level Expert Group on Artificial Intelligence, 2018, p. 7). No âmbito do Poder Judiciário, isso implica garantir que sistemas de IA aplicados a decisões judiciais ou administrativas sejam compatíveis com o princípio da isonomia (art. 5º, caput, da Constituição Federal), com a dignidade da pessoa humana (art. 1º, III) e com o direito à revisão de decisões automatizadas (art. 20 da LGPD) (Brasil, 1988; 2018).

No processo de avaliação dos impactos da IA nos direitos humanos, que deve se dar de maneira contínua durante o ciclo de vida da IA, deve-se considerar a realização de testes e auditorias por especialistas independentes a fim de avaliar a sua conformidade, o estabelecimento de medidas de prevenção e mitigação de riscos envolvendo a violação de direitos humanos, bem como a criação de medidas de segurança capazes de interromper o uso de sistemas que violem tais direitos (Accessnow.Org, 2018, p. 32-33). Essas auditorias, quando aplicadas ao contexto judicial, devem incluir indicadores de gênero, permitindo identificar discrepâncias entre decisões automatizadas aplicadas a mulheres e homens em situações semelhantes. Assim, a proteção dos direitos humanos se traduz em instrumento técnico de monitoramento da equidade algorítmica.

Vale mencionar que *experts* em IA que elaboraram o *Draft Ethics Guidelines for Trustworthy AI* da Comissão Europeia elencaram cinco princípios e valores que devem ser observados a fim de assegurar o desenvolvimento de sistemas inteligentes centrados no ser humano. O primeiro deles seria o princípio do benefício (fazer o bem), que consiste em desenvolver uma IA que aumente o bem-estar individual e coletivo, gerando prosperidade, criando valores e promovendo a sustentabilidade (High-Level Expert Group on Artificial Intelligence, 2018, p. 8-9, tradução nossa³⁸):

³⁸ No original: [...] beneficent AI systems can contribute to wellbeing by seeking achievement of a fair, inclusive and peaceful society, by helping to increase citizen's mental autonomy, with equal distribution of economic, social and political opportunity. AI systems can be a force for collective good when deployed towards objectives like: the protection of democratic process and rule of law; the provision of common goods and services at low cost and high quality; data literacy and representativeness; damage mitigation and trust optimization towards users; achievement of the UN Sustainable Development Goals or sustainability understood more broadly, according to the pillars of economic development, social equity, and environmental protection¹⁰. In other words, AI can be a tool to bring more good into the world and/or to help with the world's greatest challenges.

[...] sistemas de IA benéficos podem contribuir para o bem-estar, buscando alcançar uma sociedade justa, inclusiva e pacífica, ajudando a aumentar a autonomia mental dos cidadãos, com distribuição igualitária das oportunidades econômicas, sociais e políticas. Os sistemas de IA podem representar uma força a favor do bem coletivo quando utilizados em objetivos como: proteção do processo democrático e do Estado de Direito; provisão de bens e serviços comuns com baixo custo e alta qualidade; alfabetização e representatividade em relação aos dados; mitigação de danos e otimização da confiança perante os usuários; alcance dos Objetivos de Desenvolvimento Sustentável da ONU ou promoção da sustentabilidade compreendida de maneira ampla, segundo os pilares de desenvolvimento econômico, equidade social e proteção ambiental[...]. Em outras palavras, a IA pode ser uma ferramenta para promover mais benefícios à humanidade e/ou auxiliar no enfrentamento dos maiores desafios mundiais.

O segundo, princípio da não maleficência (não fazer o mal), assegura que sistemas inteligentes não façam mal a indivíduos e que respeitem a dignidade, integridade, liberdade, privacidade e segurança das pessoas. Deve-se adotar medidas que evitem a discriminação e que deem especial atenção às minorias, promovendo a inclusão social e o respeito à diversidade (High-Level Expert Group on Artificial Intelligence, 2018, p. 9).

Por sua vez, o princípio da autonomia (preservar a autonomia humana) dispõe que seres humanos não devem estar na condição de subordinados em relação à IA, nem sofrer coerções em função dela (High-Level Expert Group on Artificial Intelligence, 2018, p. 9-10).

O quarto, princípio da justiça (ser justa), determina que o desenvolvimento, o uso e a regulamentação da IA devem ser justos, afastando discriminações, estigmatizações e vieses. Ele implica que os sistemas de IA forneçam aos usuários mecanismos eficazes de reparação de danos e impõe aos responsáveis pelo desenvolvimento ou implementação da IA altos padrões de responsabilidade (High-Level Expert Group on Artificial Intelligence, 2018, p. 10).

Por fim, o princípio da explicabilidade (transparência) busca garantir que sistemas de IA sejam auditáveis e compreensíveis pelo ser humano (High-Level Expert Group on Artificial Intelligence, 2018, p. 10).

Outrossim, a proteção de dados e o direito à privacidade devem ser garantidos em todo o ciclo de desenvolvimento e aplicação de sistemas de IA, em consonância com a LGPD (Brasil, 2018a), que consagra, em seu art. 6º, o princípio da segurança, da prevenção e da não discriminação, além de prever, no art. 20, o direito à revisão de decisões automatizadas. Esses dispositivos constituem marcos legais indispensáveis à garantia de que os sistemas de IA respeitem os direitos fundamentais e não produzam discriminações indiretas.

A regulamentação do desenvolvimento e aplicação de técnicas de IA é de suma importância para a garantia da observância de técnicas que não afrontem direitos humanos. É através dela que se garante e se promove o desenvolvimento saudável e sustentável de sistemas

inteligentes (Cui, 2020, p. 188). Não obstante, as mudanças sociais que as tecnologias de IA trazem são profundas, diversas e não totalmente conhecidas. Deve-se atentar para que regulamentações não sejam rígidas demais e alimentem a oposição ao uso de mecanismos que tragam benefícios sociais, sufocando a inovação.

Merece destaque também a proposta de princípios éticos da Declaração de Montreal para o Desenvolvimento Responsável da IA (2018). Vide Figura 33 que esclarece, sucintamente, cada um dos princípios afirmados pela Declaração, os quais, de maneira traduzida, são os seguintes (Declaração de Montreal para o Desenvolvimento Responsável da IA (2018, tradução e adaptação nossa):

- a)** Princípio do Bem-Estar: o desenvolvimento e o uso de sistemas de inteligência artificial (IA) devem promover o crescimento do bem-estar de todos os seres sencientes.
- b)** Princípio do Respeito à Autonomia: os sistemas de IA devem ser desenvolvidos e utilizados com respeito à autonomia e à autodeterminação dos indivíduos, assegurando que as pessoas mantenham o controle sobre suas próprias vidas e decisões.
- c)** Princípio da Proteção da Privacidade e da Intimidade: a privacidade e a intimidade devem ser protegidas contra a intrusão e o uso abusivo de dados por sistemas de inteligência artificial.
- d)** Princípio da Solidariedade: o desenvolvimento de sistemas de IA deve ser compatível com o fortalecimento dos laços de solidariedade entre pessoas e gerações.
- e)** Princípio da Participação Democrática: os sistemas de IA devem ser projetados para garantir inteligibilidade, justificabilidade e acessibilidade, de modo a promover a participação democrática, o diálogo e o controle social.
- f)** Princípio da Equidade: o desenvolvimento e o uso da IA devem contribuir para a construção de uma sociedade justa e equitativa.
- g)** Princípio da Inclusão e Diversidade: o desenvolvimento e o uso da IA devem ser compatíveis com a diversidade, assegurando que todos os indivíduos – independentemente de suas origens ou experiências pessoais – possam se beneficiar igualmente dos avanços tecnológicos.

- h) Princípio da Prudência:** toda pessoa envolvida no desenvolvimento de sistemas de IA deve agir com cautela e responsabilidade moral, antecipando as consequências de suas ações e tomando medidas apropriadas para evitar danos.
- i) Princípio da Responsabilidade:** o desenvolvimento e o uso da IA não devem contribuir para a diminuição da responsabilidade humana diante de suas ações e decisões.
- j) Princípio do Desenvolvimento Sustentável:** o desenvolvimento e o uso da IA devem ser conduzidos de modo a sustentar um crescimento econômico, social e ambiental equilibrado e sustentável para as gerações presentes e futuras.

Figura 33. Princípios da Declaração de Montreal para o Desenvolvimento Responsável da Inteligência Artificial.



Fonte: Montréal Declaration Responsible AI, [2018?].

Tal como defende García (2019, p. 18-19), é necessário estabelecer regras e princípios que guiem o desenvolvimento e o uso da IA de maneira ética, minimizando os riscos inerentes à utilização de tais sistemas. O autor, com base em organismos internacionais, propõe como princípios os expostos no Quadro 7.

Quadro 7. Princípios-guia para o desenvolvimento e uso ético de IA, segundo García.

Princípio	Descrição e Aplicações
a) Respeito à autonomia humana	Os sistemas de IA devem sempre respeitar a autonomia e os direitos humanos, garantindo a preservação da vida e da dignidade sem qualquer forma de discriminação. Devem operar de modo a não interferir injustamente na capacidade das pessoas de tomar decisões próprias. O <i>design</i> e a programação devem proteger direitos fundamentais como privacidade, liberdade de expressão e segurança, assegurando a ausência de vieses que possam gerar tratamento desigual entre indivíduos ou grupos.
b) Transparência	É fundamental que os usuários compreendam como e por que um sistema de IA chega a determinadas decisões. A transparência envolve a explicabilidade e a rastreabilidade do processo algorítmico, permitindo acompanhar os dados utilizados, o funcionamento do algoritmo e as etapas que conduzem aos resultados. Todo o processo deve ser compreensível tanto sob o ponto de vista técnico (programação) quanto humano (<i>design</i>). Tecnologias imprevisíveis ou ininteligíveis violam o princípio da autonomia humana, razão pela qual sua operação deve ser clara e acessível.
c) Responsabilidade e prestação de contas (<i>accountability</i>)	Considerando que sistemas de IA podem causar impactos significativos sobre pessoas e grupos, é indispensável que haja atribuição clara de responsabilidades e obrigações. Devem existir estruturas de auditoria e mecanismos de controle que garantam a possibilidade de revisão e correção de decisões automatizadas. A suposta autonomia da IA não pode servir de pretexto para a irresponsabilização de seus criadores, operadores ou usuários institucionais.
d) Robustez e segurança	Para que sejam confiáveis, os sistemas de IA devem ser robustos e seguros, capazes de resistir a falhas e ataques. É essencial que permitam a correção de erros e inconsistências ao longo de todo o ciclo de vida do sistema. Isso inclui medidas de cibersegurança, atualização contínua e monitoramento constante, assegurando a integridade dos dados e a confiabilidade dos resultados.

e) Justiça e não discriminação	Os sistemas de IA devem ser inclusivos e equitativos, prevenindo práticas discriminatórias. Os vieses podem surgir tanto nos dados históricos de treinamento quanto nas estruturas algorítmicas. No primeiro caso, refletem desigualdades pré-existentes na sociedade; no segundo, decorrem de concepções enviesadas de programadores ou falhas de código. Por isso, devem ser implementadas práticas de auditoria de dados, curadoria ética e diversidade nas equipes de desenvolvimento para garantir decisões justas e imparciais.
---------------------------------------	---

Fonte: a autora (adaptado de García, 2019, p. 18-19, tradução e adaptação nossas³⁹)

Pode-se dizer que a Constituição Federal (Brasil, 1988) reafirma esses princípios no plano interno ao instituir a dignidade da pessoa humana (art. 1º, III) e a igualdade de todos perante a lei (art. 5º, caput) como fundamentos da ordem jurídica. Esses preceitos vinculam o uso de tecnologias públicas à promoção da igualdade substantiva e da não discriminação, especialmente de gênero. Assim, toda política de IA no Judiciário deve estar subordinada à concretização desses valores constitucionais.

No âmbito do Poder Judiciário, a Resolução n. 615/2025 do CNJ (Conselho Nacional de Justiça, 2025c) representa um marco regulatório relevante, mas ainda insuficiente para assegurar a efetiva observância dos direitos humanos sob a perspectiva de gênero. Embora

³⁹ No original: “1. Respeto de la autonomía humana: los sistemas inteligentes deben respetar en todo momento la autonomía y los derechos fundamentales de las personas. Su diseño y programación debe respetar, por tanto, la vida y los derechos humanos sin ningún tipo de discriminación.

2. Transparencia: en el caso de los sistemas provistos de IA, la transparencia atañe principalmente a la explicabilidad y la trazabilidad de dichos sistemas. Dado que el diseño de estos dispositivos contempla que tomen decisiones automáticamente con base en distintos cálculos y proyecciones, debe ser posible en todo momento trazar el razonamiento seguido por el sistema y explicar las consecuencias alcanzadas. En concreto, ha de ser posible trazar el conjunto de datos empleados en el razonamiento, el funcionamiento del algoritmo y los pasos seguidos para alcanzar los resultados. Todo este proceso debe ser, además, explicable desde los puntos de vista técnico de la programación y humano del diseño. El diseño y el empleo de una tecnología impredecible son incompatibles con la defensa de la autonomía humana. Es absolutamente necesario que la actividad de todos estos dispositivos sea de fácil comprensión y acceso.

3. Responsabilidad y rendición de cuentas: estrechamente relacionado con el principio anterior, el diseño y el empleo de sistemas inteligentes deben estar precedidos por una clara asignación de responsabilidades ante los posibles daños y perjuicios que estos puedan ocasionar. La presunta autonomía de estos sistemas no puede servir de pretexto para la dilución de responsabilidades. Al contrario, será preciso incluir los mecanismos adecuados (auditoría, informe de errores, penalizaciones, etc.) para asegurar que las responsabilidades y obligaciones en relación con el funcionamiento de estos sistemas queden bien definidas.

4. Robustez y seguridad: la fiabilidad de la IA exige que los algoritmos sean suficientemente seguros, fiables y sólidos para operar de manera precisa y segura, y para resolver errores o incoherencias durante todas las fases del ciclo de vida útil de los dispositivos. Este principio exige, además, que los sistemas se diseñen y desarrollen contemplando la posibilidad de ciberataques y fallos técnicos.

5. Justicia y no discriminación: el diseño de estos sistemas debe contar con la participación de todos los grupos de interés con los que cada aplicación provista de IA se relacione. Además, estos dispositivos deben garantizar un empleo justo de los datos disponibles para evitar posibles discriminaciones hacia determinados grupos o distorsiones en los precios y en el equilibrio de mercado”.

reconheça a diversidade e a não discriminação (arts. 3º, I e 35), a norma não prevê mecanismos específicos de proteção às mulheres e minorias nos processos de concepção, treinamento e auditoria dos sistemas de IA, nem institui avaliações obrigatórias de impacto em direitos humanos. Diante da baixa representatividade feminina na magistratura e nas equipes técnicas, essa omissão pode perpetuar desigualdades estruturais e limitar a eficácia da equidade algorítmica.

Portanto, é imprescindível aprimorar a referida Resolução, incorporando salvaguardas vinculantes de gênero, mecanismos de revisão de vieses e medidas de transparência pública, de modo que a regulação da IA no Judiciário se converta em instrumento de governança algorítmica inclusiva e democrática.

Somente por meio da integração entre ética, técnica e direito – e de uma regulação viva, participativa e sensível às desigualdades – será possível garantir que a IA funcione como ferramenta de emancipação social e de concretização dos direitos humanos, e não como veículo de reprodução das assimetrias históricas do sistema de justiça.

5.2.2 O papel da regulamentação geral sobre ética e inteligência artificial

Normas gerais e documentos internacionais que discorram sobre ética algorítmica não devem ser ignorados. Além daqueles já trazidos nesta tese, outro exemplo de trabalho internacional de relevância sobre o tema é o desenvolvido pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE). A OCDE (OECD.AI Policy Observatory; GPAI Global Partnership, [20–]) assenta princípios – crescimento inclusivo, direitos humanos/valores democráticos (equidade e privacidade), transparência/explicabilidade, robustez/segurança e responsabilização – e recomenda políticas de P&D, ecossistemas inclusivos, governança interoperável, capacitação e cooperação internacional. A Figura 34 ilustra os princípios da OCDE que servem como orientação aos atores envolvidos no ecossistema de IA na promoção do desenvolvimento de sistemas confiáveis e, ao mesmo tempo, oferecem subsídios normativos aos formuladores de políticas públicas para a construção de diretrizes eficazes em matéria de regulação da IA.

Figura 34. Panorama dos princípios sobre IA da OCDE.



Fonte: OECD.AI Policy Observatory e GPAI, [20–].

Segundo os princípios recomendados pela OCDE, a IA deve promover o crescimento inclusivo, o desenvolvimento sustentável e o bem-estar das pessoas e do ambiente. Assim, a IA necessariamente tem de colocar os interesses humanos e ambientais em primeiro lugar, aumentando as capacidades humanas, a promoção do crescimento inclusivo e o desenvolvimento sustentável (Salomão, 2022, p. 49).

Além disso, os sistemas de IA devem ser desenvolvidos com base no respeito ao Estado de Direito, aos direitos humanos, aos valores democráticos e à diversidade. Devem incluir salvaguardas adequadas, como a possibilidade de intervenção humana quando necessário, para assegurar uma sociedade justa (OECD.AI Policy Observatory; GPAI Global Partnership, [20–]). A IA deve respeitar o direito à liberdade, dignidade, autonomia, privacidade e proteção de dados, igualdade, não discriminação e diversidade etc. (Salomão, 2022, p. 49).

Ainda, de acordo com os princípios elencados pelo Conselho, os sistemas de IA devem operar de maneira robusta, segura e protegida ao longo de seu ciclo de vida. Os riscos potenciais devem ser avaliados e gerenciados continuamente (OECD.AI Policy Observatory; GPAI Global Partnership, [20–]). Assim, deve-se assegurar a segurança do processo e das decisões tomadas pela IA de maneira contínua (Salomão, 2022, p. 49).

A respeito da responsabilização, adverte-se que as organizações e os indivíduos envolvidos no desenvolvimento, implementação ou operação de sistemas de IA devem ser responsabilizados pelo seu funcionamento adequado, em conformidade com os princípios acima mencionados (OECD.AI Policy Observatory; GPAI Global Partnership, [20–]). Tais atores devem ser responsáveis pelo funcionamento adequado do sistema inteligente, promovendo a sua harmonia com os princípios elencados, conforme seus papéis e contextos (Salomão, 2022, p. 49).

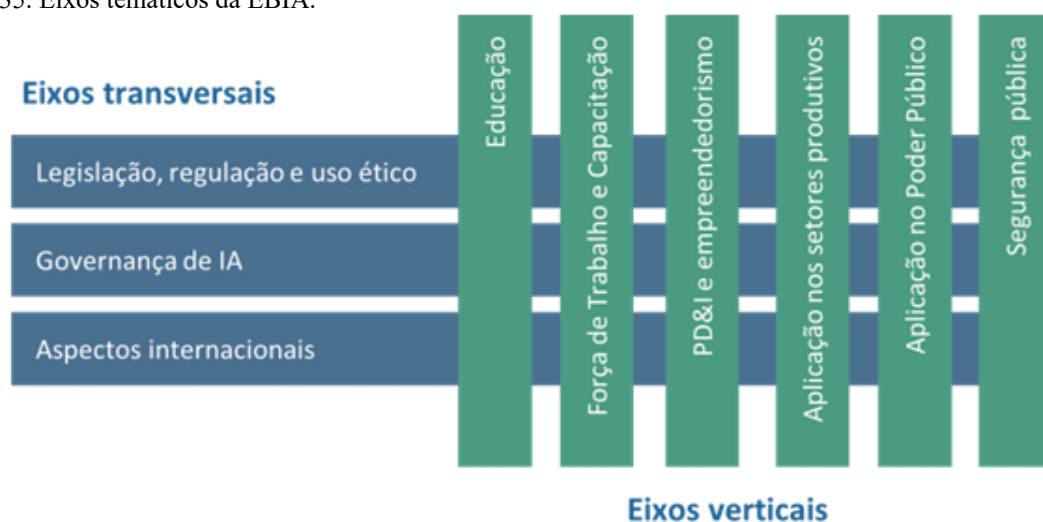
No Brasil, levando em consideração os princípios traçados pela OCDE, foi criada a Estratégia Brasileira de Inteligência Artificial (EBIA) em 2021 (Ministério da Ciência, Tecnologia e Inovações, 2021), que tem por objetivo potencializar o desenvolvimento e a utilização da IA para a promoção do avanço científico e a solução de problemas concretos (SALOMÃO, 2022, p. 50). Foram estabelecidos nove pilares temáticos que buscam contribuir para a elaboração de princípios éticos que permeiem o desenvolvimento e o uso de IA de maneira responsável, promovendo-se, concomitantemente, a inovação em IA.

Salomão (2022, p. 50), de maneira bastante didática, traz a Figura 35, que expõe três eixos transversais (legislação, regulação e uso ético; governança de IA; e aspectos internacionais) que perpassam os demais (educação, força de trabalho e capacitação, PD&I e empreendedorismo; aplicação nos setores produtivos; aplicação no Poder Público; e segurança pública):

O eixo transversal “legislação, regulação e uso ético” destaca a importância de estabelecer parâmetros legais, regulatórios e éticos para guiar o avanço e utilização da IA. Como explica Salomão (2022, p. 50), alguns assuntos a serem abordados nesse eixo incluem a salvaguarda da privacidade e proteção de dados pessoais, a mitigação de discriminação e preconceitos algorítmicos e da criação de diretrizes legais que assegurem a estabilidade jurídica para os diversos envolvidos no ciclo de vida dos sistemas.

O eixo intitulado “governança de IA” visa assegurar que princípios éticos permeiem o desenvolvimento e o uso de IA, seja no setor público, seja no privado (Salomão, 2022, p. 50). O terceiro eixo transversal, por sua vez, busca integrar o Brasil em organismos e fóruns internacionais que incentivem o uso ético da IA, intensificando o compartilhamento de conhecimento e boas práticas (Salomão, 2022, p. 50).

Figura 35. Eixos temáticos da EBIA.



Fonte: Ministério da Ciência, Tecnologia e Inovações, 2021.

Não obstante, para além da EBIA, não há no Brasil regulamentação específica sobre IA no Brasil. Ausência de uma lei específica impõe ao Poder Judiciário e às demais instituições públicas o desafio de desenvolver parâmetros próprios de governança, alinhados às boas práticas internacionais e aos princípios constitucionais de dignidade, igualdade e justiça. Embora existam normas dispersas que tangenciam o tema – como a LGPD (Brasil, 2018a) e a Resolução CNJ n. 615/2025 (Conselho Nacional de Justiça, 2025c) –, o país ainda carece de um marco legal abrangente que discipline o ciclo de vida da IA, desde o desenvolvimento até a aplicação, incluindo aspectos como ética, segurança, transparência, responsabilidade e prevenção de vieses.

Atualmente, tramita no Congresso Nacional o Projeto de Lei n. 2338, de 2023, de autoria do Senador Rodrigo Pacheco (PSD/MG), que propõe instituir o Marco Legal da Inteligência Artificial no Brasil (Brasil, 2023a). O texto busca definir princípios, direitos e deveres para o desenvolvimento e o uso responsável da IA, com base em diretrizes como a centralidade do ser humano, a não discriminação, a transparência, a governança e a prestação de contas. Apesar de representar um avanço, o projeto ainda demanda aprimoramentos significativos, sobretudo na incorporação explícita de salvaguardas relacionadas à equidade de gênero e à diversidade, elementos essenciais para prevenir a reprodução de vieses estruturais em sistemas automatizados.

Sem dúvidas, legisladores e formuladores de políticas públicas desempenham um papel crucial na regulamentação da IA, pois possuem a capacidade de criar estruturas legais e diretrizes que assegurem o uso ético dessas tecnologias, observando princípios e regras gerais

voltados à proteção dos direitos humanos e incorporando valores essenciais como inclusão social, responsabilidade, *accountability* e justiça (UNESCO; IRCAI, 2024, p. 4).

A criação de regulamentações que exijam transparência nos algoritmos e nos conjuntos de dados utilizados para o treinamento é fundamental para identificar e corrigir possíveis vieses. Isso implica estabelecer padrões claros de coleta, curadoria e uso de dados, bem como normas para o desenvolvimento de algoritmos que garantam equidade, auditabilidade e explicabilidade, evitando que novas discriminações sejam introduzidas ou perpetuadas.

É imprescindível, ainda, que haja supervisão regulatória rigorosa e independente, assegurando a conformidade das soluções de IA com princípios éticos e jurídicos. Essa supervisão deve envolver a realização de auditorias periódicas, a análise de impacto algorítmico e o monitoramento contínuo de práticas discriminatórias, de modo a garantir que a justiça e a igualdade sejam mantidas ao longo do tempo.

Além disso, os governos devem exigir que as empresas e órgãos públicos de tecnologia invistam em pesquisas que analisem os impactos da IA sobre diferentes grupos sociais, especialmente mulheres, pessoas negras e populações em situação de vulnerabilidade, para que o desenvolvimento tecnológico seja orientado por princípios éticos e pelo bem-estar coletivo.

A criação de parcerias multissetoriais, envolvendo especialistas em tecnologia, organizações da sociedade civil, universidades, operadores do direito e comunidades afetadas, é igualmente essencial para a construção de um ecossistema regulatório inclusivo e participativo. Essa abordagem colaborativa permite que múltiplas perspectivas sejam consideradas, tornando os sistemas de IA mais justos, auditáveis e menos suscetíveis a causar danos (UNESCO; IRCAI, 2024, p. 4).

No contexto do Poder Judiciário brasileiro, a regulamentação geral da IA deve ser vista como parâmetro orientador, a partir do qual se possam adaptar diretrizes específicas voltadas à justiça algorítmica e à mitigação de vieses de gênero. Isso inclui a incorporação, no âmbito judicial, de exigências de avaliação de impacto em direitos humanos, critérios de diversidade nas equipes técnicas e procedimentos de auditoria pública dos sistemas utilizados, garantindo que a automação não substitua o discernimento humano, mas o complemente sob a ótica da equidade e da responsabilidade institucional.

Dessa forma, a regulamentação geral sobre ética e IA não deve ser entendida como mero instrumento de controle tecnológico, mas como um pilar democrático que assegura que o avanço da IA ocorra em conformidade com os valores constitucionais e os compromissos internacionais de direitos humanos.

5.2.3 Governança dos sistemas inteligentes

A qualidade dos dados utilizados na alimentação de sistemas de IA é crucial para o desempenho e para a acurácia das soluções de aprendizado de máquina. Mesmo quando os dados são tratados de modo a preservar a privacidade, é indispensável cumprir requisitos rigorosos para assegurar uma IA de alta qualidade. Como visto, bases de dados inevitavelmente podem carregar vieses; por isso, é preciso identificá-los e mitigá-los antes do treinamento dos modelos e, quando necessário, ao longo do próprio processo de treinamento.

Embora a tomada de decisão mediada por IA possa ser altamente vantajosa, seus efeitos colaterais de longo prazo demandam esforço conjunto entre setor público e privado para que a sociedade – no presente e no futuro – não se depare com um cenário oposto ao desenvolvimento social desejado, em que desigualdades sejam exacerbadas (Pinto, 2020, p. 58). Qualquer que seja o grau de interação entre humanos e sistemas inteligentes, impõe-se o desenvolvimento de critérios éticos e de medidas de transparência que preservem e promovam direitos humanos.

Nesse contexto, governança de dados deve ser entendida como um arranjo regulatório contínuo aplicado ao ciclo de vida dos dados – da coleta ao descarte – que define papéis, responsabilidades, processos e controles para assegurar qualidade, segurança, legitimidade e finalidade do tratamento. Em termos práticos, a governança opera como uma “regulação sobre os dados”: estabelece políticas de minimização, padroniza metadados, cria trilhas de auditoria, define curvas de retenção/eliminação e institui mecanismos de validação, curadoria e acesso. Ao fazê-lo, materializa princípios e garantias éticas associados à IA (não discriminação, transparência, explicabilidade, prestação de contas, robustez e segurança), pois sem dados governados não há como demonstrar *fairness*, nem como auditar ou explicar decisões automatizadas.

A IA e outras tecnologias emergentes representam um marco transformador para o setor público. Ao mesmo tempo em que abrem oportunidades para otimizar serviços, criar soluções inovadoras e impulsionar o progresso social, trazem desafios que exigem profunda reavaliação de estruturas e processos institucionais. Para aproveitar os benefícios da IA de forma responsável e sustentável, as administrações públicas precisam adotar estratégia que integre novas tecnologias com governança sólida, garantindo desenvolvimento inclusivo e equitativo, combatendo desigualdades e promovendo acesso universal aos benefícios da automação.

No Judiciário, a IA pode ser utilizada, por exemplo, para automatizar tarefas repetitivas, liberando tempo de magistrados e servidores para atividades que exigem discernimento. Essa otimização tende a tornar o processo mais célere, eficiente e coerente. Mas tais ganhos só são legítimos se sustentados por governança de dados: catálogos e dicionários de dados, políticas de qualidade (*completeness, consistency, accuracy*), controle de versões de *datasets*, registros de linhagem (*data lineage*) e validações contínuas que evitem contaminações, *drifts* e usos indevidos.

É fundamental que o desenvolvimento e a implementação da IA no Poder Judiciário permaneçam consonantes com princípios basilares do Direito – igualdade, não discriminação, devido processo e presunção de inocência – e com o arcabouço de proteção de dados. A governança, ao padronizar fluxos, direitos de acesso e salvaguardas, operacionaliza tais princípios: impõe minimização e finalidade legítima; requer bases legais e registros de tratamento; prevê anonimização/pseudoanonimização quando couber; e vincula cada conjunto de dados a controles de qualidade, segurança e *accountability*.

Mais do que isso, impõe-se a construção de uma IA confiável, que atue dentro de princípios éticos previamente estabelecidos para proteger direitos e liberdades e maximizar o bem-estar comum (High-Level Expert Group on Artificial Intelligence, 2018, p. 5). É imprescindível que todos os participantes do ecossistema jurídico não abdicuem de um substrato ético mínimo no processo decisório: sem governança, o valor da Justiça pode ser reduzido a estatísticas incapazes de refletir a realidade que o Direito pretende regular.

As reflexões éticas ajudam a identificar como tecnologias impactam direitos fundamentais e orientam o que fazer com a tecnologia em prol do bem comum (High-Level Expert Group on Artificial Intelligence, 2018, p. 5). O guia ético da Comissão Europeia é claro: respeito a direitos fundamentais, princípios e valores – e a comprovação de conformidade – é elemento-chave de uma IA confiável (High-Level Expert Group on Artificial Intelligence, 2018, p. 6).

Decisões judiciais apoiadas por IA devem ser transparentes, previsíveis e auditáveis. A governança de dados fornece as condições de possibilidade para isso: sem catálogos, linhagem e *logs*, não há como provar imparcialidade nem evitar a reprodução de vieses sociais. Logo, defende-se um debate público e institucional sobre as questões éticas envolvidas tanto na criação quanto na aplicação de IA no Direito. A qualidade dos dados de treinamento é determinante para a confiabilidade das decisões; por isso, devem advir de fontes seguras, rastreáveis e auditáveis, sob protocolos rigorosos de segurança, de modo a evitar acessos não

autorizados e preservar a privacidade. Os indivíduos devem ter conhecimento e controle sobre seus dados, com transparência sobre o que é coletado, possibilidade de contestação e correção: a privacidade é um valor fundamental na era da IA.

A ausência de lei específica no Brasil exclusivamente dedicada à IA reforça o papel da governança de dados como primeira linha de regulação prática dentro das organizações públicas. Ela complementa normas existentes (por exemplo, proteção de dados e segurança da informação), oferecendo previsibilidade e responsabilidade no uso da IA e protegendo direitos dos cidadãos. Nesse sentido, uma abordagem de “tecnologia social” – participativa e inclusiva – deve embasar políticas de dados: definição colaborativa de padrões, métricas de equidade desagregadas por gênero/raça, canais de contestação e participação de sociedade civil e academia.

Navegar esse cenário requer que administrações públicas se mantenham atualizadas e proativas. A governança de dados – comitês, papéis (*data owners, stewards*), indicadores de qualidade, testes de viés, auditorias independentes e ciclos de melhoria – é o que transforma normas principiológicas em rotina institucional.

Nessa linha, Corrêa (2021, p. 199) distingue “regulação do algoritmo” (normas/diretrizes do Estado e de instituições para criar e usar algoritmos) e “regulação pelo algoritmo” (o algoritmo como produtor de normas). A governança de dados articula ambas: define o que é permitido (políticas) e como o sistema deve se comportar (regras e salvaguardas implementadas nos próprios pipelines).

Bostrom e Yudkowsky (2011, p. 201-202) ressaltam a necessidade de algoritmos “transparentes para inspeção” e previsíveis, especialmente em funções sociais; evocam, inclusive, o paralelo com o *stare decisis* para sublinhar que previsibilidade jurídica é valor institucional. Além da transparência, listam critérios desejáveis: responsabilidade, auditabilidade, incorruptibilidade, previsibilidade e propensão a não vitimar inocentes – exatamente o tipo de requisitos que a governança de dados busca concretizar por meio de controles técnicos e administrativos (Bostrom; Yudkowsky, 2011, p. 202-203).

Ser proativo é crucial. Como alerta Jacobstein (2014), é melhor antecipar riscos antes que tecnologias estejam plenamente difundidas, criando controles redundantes e em camadas. No campo judicial, a Carta Europeia de Ética sobre o Uso de IA em Sistemas Judiciais (CEPEJ; Council of Europe, 2018) oferece princípios valiosos – respeito a direitos fundamentais; não discriminação; qualidade e segurança; transparência, imparcialidade e justiça; e “sob controle do usuário” – todos dependentes de governança de dados robusta, já que é ela que garante fontes

seguras, qualidade verificável, explicabilidade e possibilidade de revisão humana (Alves, 2020, p. 35).

Convergente com isso, as “Diretrizes de IA Confiável” da Comissão Europeia (2018) afirmam a necessidade de levar em consideração dois componentes: (a) respeito a direitos humanos, normas aplicáveis, princípios e valores; e (b) solidez técnica. Entre os requisitos elencados estão responsividade, não discriminação, governança de dados, *design* inclusivo, respeito à autonomia e à privacidade, além de robustez, segurança e transparência (High-Level Expert Group on Artificial Intelligence, 2018, p. 14).

Ainda, a IEEE, em *Ethically Aligned Design* (Institute of Electrical and Electronic Engineers, 2017), reforça princípios como benefício humano, responsabilidade, transparência e educação, todos dependentes de práticas de dados verificáveis.

Especificamente no âmbito do Judiciário, como já se viu nesse trabalho, a Resolução CNJ n. 332/2020 inaugurou diretrizes de ética, transparência e governança (Conselho Nacional de Justiça, 2020b) e a Resolução n. 615/2025 consolidou um marco de segurança, transparência, eficiência e respeito a direitos fundamentais, incluindo padrões de qualidade, rastreabilidade, segurança e isolamento de dados, anonimização/pseudoanonimização, controle de versões e auditorias periódicas – todos instrumentos nucleares de governança de dados (Conselho Nacional de Justiça, 2025c). Ainda, a Portaria n. 271/2020 organizou projetos e definiu a plataforma Sinapses (Conselho Nacional de Justiça, 2020a).

O texto da Resolução n. 615/2025 do CNJ reforça que a governança das soluções de IA deve respeitar a autonomia dos tribunais, permitindo o desenvolvimento e a implementação de soluções inovadoras locais, ajustadas às realidades institucionais e territoriais de cada tribunal, desde que observados os padrões de auditoria, monitoramento e transparência definidos nacionalmente, sem prejuízo da atuação coordenadora do CNJ (Conselho Nacional de Justiça, 2025c).

Além de resguardar a autonomia institucional, a Resolução consolida um sistema de governança algorítmica voltado à integridade das decisões judiciais, à responsabilização pelo uso de tecnologias e à prevenção de danos decorrentes de vieses discriminatórios (arts. 12 e 13). Determina que cada tribunal deve estruturar processos internos para garantir a segurança e a transparência dos sistemas de IA, assegurando a publicação de relatórios que descrevam seu funcionamento, finalidades, dados utilizados e mecanismos de supervisão (art. 12, I). Também impõe a obrigação de prevenção e mitigação de vieses ilegais ou abusivos, por meio de

monitoramento contínuo, revisões periódicas e correção de eventuais desvios (art. 12, II) (Conselho Nacional de Justiça, 2025c).

De forma complementar, o normativo orienta que as soluções adotadas sejam interoperáveis e passíveis de adaptação a diferentes contextos, priorizando *softwares* de código aberto ou comerciais que garantam transparência e proteção de dados pessoais (art. 12, V). Determina ainda que as soluções de IA sigam boas práticas de gestão de produto, com fases claramente definidas – requisitos, desenvolvimento, testes, implementação, suporte e melhoria contínua –, incluindo comitês internos de fiscalização responsáveis por assegurar a observância das diretrizes éticas, de segurança e de governança (art. 12, VI e VIII) (Conselho Nacional de Justiça, 2025c).

Nos casos de modelos de IA classificados como de alto risco, a Resolução (Conselho Nacional de Justiça, 2025c) exige medidas reforçadas de governança (arts. 13 e 14), como: o uso de dados de treinamento representativos e equilibrados, que reflitam adequadamente os grupos afetados; o registro das fontes automatizadas e do grau de supervisão humana nos resultados apresentados; a indicação clara, em linguagem simples, dos objetivos e resultados pretendidos; a documentação completa e compreensível de todo o ciclo de vida do sistema; a manutenção de logs automáticos de operação, que permitam auditoria, rastreabilidade e detecção de resultados discriminatórios; mecanismos de explicabilidade que assegurem a compreensão dos resultados por usuários e magistrados; e a realização de avaliação de impacto algorítmico contínua, acompanhada de auditorias regulares, revisões periódicas e participação pública simplificada, garantindo o acesso de entidades como OAB, Ministério Público e Defensoria Pública aos relatórios e às medidas corretivas adotadas.

Não obstante, é preciso observar que, na prática, os desafios são expressivos. É indispensável transparência, explicabilidade, auditabilidade e possibilidade real de responsabilização por erros ou discriminações. Avaliar impacto ético é difícil – pela subjetividade embutida no *design*, pela opacidade de algoritmos de aprendizado e pela incerteza sobre critérios que evoluem em produção (Mittelstadt et al., 2016). Determinar se um erro é isolado ou sistêmico exige governança: *logs*, versões, *experiment tracking*, avaliação segmentada por grupos sensíveis e processos claros de correção. Ademais, decisões judiciais têm efeitos sociais além das partes (Fonseca, 2022, p. 299), o que eleva o padrão de cuidado.

Como ressalta Machado Segundo (2020), transparência algorítmica é requisito de legalidade e de controle democrático: governos devem agir com abertura e clareza; manter em segredo critérios de decisão automatizada é incompatível com o Estado de Direito. Assim, a

governança de dados – com políticas claras, documentação verificável, controles técnicos e participação social – é o meio institucional pelo qual o Judiciário transforma princípios éticos e direitos fundamentais em prática cotidiana de dados e, por consequência, em IA confiável, inclusiva e não discriminatória.

A efetiva implementação das medidas de governança previstas na Resolução n. 615/2025 é essencial para garantir que os sistemas de IA empregados no Poder Judiciário operem de maneira confiável, justa e responsável (Conselho Nacional de Justiça, 2025c). A simples previsão normativa de mecanismos de auditoria, transparência e avaliação de impacto algorítmico não é suficiente se tais dispositivos não forem postos em prática de forma sistemática, documentada e verificável.

A aplicação concreta dessas diretrizes – sobretudo a manutenção de registros automáticos (*logs*), as auditorias regulares, o monitoramento contínuo de vieses discriminatórios e a avaliação pública de impacto – permite identificar e corrigir desvios que poderiam perpetuar desigualdades estruturais, como os vieses de gênero e outros tipos de discriminação indireta. Esses instrumentos viabilizam o controle efetivo do ciclo de vida dos sistemas de IA, possibilitando que eventuais distorções nos dados de treinamento, nas variáveis ponderadas ou nos padrões de decisão sejam detectadas em tempo hábil.

Além disso, a operacionalização das medidas de governança fortalece o controle institucional e social sobre o uso de tecnologias automatizadas no sistema de justiça. Ao assegurar a participação de órgãos externos, como o Ministério Público, a Defensoria Pública e a OAB, o processo torna-se mais transparente e plural, permitindo o escrutínio público e o diálogo entre especialistas técnicos e jurídicos. Essa integração é fundamental para evitar que decisões judiciais baseadas em algoritmos se transformem em “caixas-pretas” imunes à contestação e, ao mesmo tempo, para garantir que o desenvolvimento tecnológico seja compatível com os princípios constitucionais da igualdade, da dignidade da pessoa humana e do devido processo legal.

Assim, colocar em prática os mecanismos de governança delineados pela Resolução não é apenas uma exigência técnica, mas um imperativo ético e democrático. Trata-se de assegurar que o Poder Judiciário disponha de ferramentas de IA que reforcem – e não comprometam – a imparcialidade, a equidade e a legitimidade de suas decisões, consolidando uma justiça algorítmica responsável e inclusiva.

5.2.4 A Resolução CNJ n. 615/2025

Dos pontos trabalhados neste capítulo, vê-se que a Resolução n. 615/2025 do CNJ representa um dos marcos mais relevantes na consolidação de um regime normativo voltado ao desenvolvimento, à governança e ao uso ético e responsável de soluções de IA no âmbito do Poder Judiciário brasileiro. Seu escopo abrange desde fundamentos éticos e direitos fundamentais até diretrizes técnicas e operacionais aplicáveis ao desenvolvimento, à auditoria, ao monitoramento e à utilização de sistemas automatizados, com atenção especial à segurança da informação, à transparência institucional, à autonomia da magistratura e à proteção dos jurisdicionados (Conselho Nacional de Justiça, 2025c).

O art. 1º da Resolução explicita que o seu objetivo é estabelecer normas para o desenvolvimento, auditoria, monitoramento e uso responsável de soluções de IA, assegurando inovação tecnológica com segurança, transparência, isonomia e respeito aos direitos fundamentais. O §1º do mesmo artigo reforça a autonomia dos tribunais na criação de soluções locais, desde que observem os padrões nacionais de governança, auditoria e transparência, o que revela a busca por um equilíbrio entre federalismo cooperativo e unidade institucional, sem prejuízo da coordenação pelo CNJ (Conselho Nacional de Justiça, 2025c).

O art. 2º elenca os fundamentos normativos do uso de IA no Judiciário, entre os quais se destacam o respeito aos direitos fundamentais e aos valores democráticos, a centralidade da pessoa humana, a promoção da igualdade e da justiça decisória, o desenvolvimento tecnológico colaborativo, a proteção de dados pessoais e a conscientização e capacitação contínua de magistrados e servidores (Conselho Nacional de Justiça, 2025c). Esses princípios dialogam com diretrizes internacionais – como a Recomendação da UNESCO sobre Ética da Inteligência Artificial (2022) e o High-Level Expert Group on AI (Comissão Europeia, 2018) – e também com a EBIA (2021).

O art. 3º reafirma os princípios estruturantes da governança algorítmica, entre os quais figuram a justiça, a equidade, a inclusão e a não discriminação, acompanhadas da transparência, explicabilidade, contestabilidade e auditabilidade. A Resolução ainda introduz a supervisão humana efetiva, a prevenção de riscos algorítmicos e a responsabilidade institucional no desenvolvimento e uso da IA no Judiciário (Conselho Nacional de Justiça, 2025c).

O Capítulo II, dedicado à compatibilidade com direitos fundamentais, reforça que toda solução de IA deve observar a Constituição Federal e os tratados internacionais de direitos humanos (art. 5º), exigindo validação contínua, revisão e auditoria periódica de seus modelos.

O art. 7º, em especial, impõe que modelos discriminatórios sejam identificados e corrigidos e, caso o viés não possa ser eliminado, o sistema deve ser descontinuado, medida essencial à prevenção de danos estruturais à igualdade material (Conselho Nacional de Justiça, 2025c).

O Capítulo III introduz uma classificação de risco para sistemas de IA, com base em critérios de impacto sobre direitos fundamentais, complexidade e sensibilidade dos dados. As soluções classificadas como de alto risco devem obrigatoriamente passar por avaliação de impacto algorítmico, conforme o art. 14, incluindo a participação de instituições externas como o Ministério Público, a Defensoria Pública e a OAB, que podem acompanhar auditorias, propor medidas corretivas e ter acesso aos relatórios. No entanto, essa previsão, embora inovadora, ainda carece de parâmetros interseccionais explícitos – por exemplo, indicadores específicos de gênero e raça nos relatórios de impacto (Conselho Nacional de Justiça, 2025c).

O Capítulo IV regula detalhadamente as medidas de governança exigidas dos tribunais desenvolvedores ou contratantes (arts. 12 e 13), impondo a criação de processos internos de supervisão e transparência, com a publicação de relatórios acessíveis ao público sobre os sistemas de IA utilizados. Esses dispositivos determinam a adoção de mecanismos de prevenção e mitigação de vieses, *logs* automáticos para rastreabilidade, comitês internos de fiscalização e documentação técnica em linguagem simples. Incentiva-se o uso de soluções interoperáveis e de código aberto, visando maior adaptabilidade e segurança e se estabelece requisitos adicionais para modelos de alto risco, incluindo dados de treinamento representativos, registro das fontes e do grau de supervisão humana e mecanismos de explicabilidade voltados à compreensão por magistrados e jurisdicionados (Conselho Nacional de Justiça, 2025c).

O Capítulo V institui o Comitê Nacional de Inteligência Artificial do Judiciário, composto por representantes da magistratura, do Ministério Público, da advocacia (OAB), da Defensoria Pública e da sociedade civil (arts. 15 e 16). O órgão é responsável por formular padrões técnicos, supervisionar auditorias, revisar critérios de risco e garantir a publicidade dos relatórios de avaliação (Conselho Nacional de Justiça, 2025c).

Embora o texto normativo da Resolução (Conselho Nacional de Justiça, 2025c) seja robusto, sob uma perspectiva feminista e interseccional, ele ainda apresenta lacunas significativas. Apesar de reafirmar a igualdade e a não discriminação (por exemplo, nos arts. 3º e 35), o CNJ não instituiu mecanismos operacionais específicos para o enfrentamento de vieses de gênero. O conceito de “viés discriminatório” permanece genérico, sem contemplar as múltiplas formas de desigualdade reproduzidas pela linguagem jurídica, pelos dados históricos e pela estrutura institucional do Judiciário.

O art. 35 (Conselho Nacional de Justiça, 2025c), que trata da diversidade nas equipes de desenvolvimento, menciona a busca por representatividade, mas sua aplicação é facultativa, podendo ser dispensada sob alegações de ausência de profissionais ou razões de eficiência. Essa flexibilização enfraquece o potencial transformador da norma, já que a composição homogênea das equipes de tecnologia tende a reproduzir os mesmos vieses de gênero e classe existentes no sistema de justiça.

Além disso, a Resolução (Conselho Nacional de Justiça, 2025c) não prevê auditorias sensíveis a gênero, protocolos éticos específicos nem a inclusão obrigatória de indicadores de gênero nos relatórios de impacto algorítmico, o que limita sua efetividade. Tal omissão perpetua o risco de que sistemas automatizados no Judiciário repitam padrões históricos de desigualdade, como a subvalorização da palavra feminina em casos de violência doméstica ou estereótipos de gênero em decisões sobre custódia e família.

É necessário, portanto, que o CNJ avance na implementação prática dos dispositivos da Resolução (Conselho Nacional de Justiça, 2025c). A execução efetiva dos mecanismos de auditoria, monitoramento contínuo, avaliação de impacto e publicidade dos relatórios é o caminho para garantir uma governança real e verificável. Essa operacionalização deve incorporar metodologias interseccionais, prevendo auditorias independentes, indicadores específicos de gênero e transparência pública sobre os resultados obtidos.

No Quadro 8, elencam-se os pontos fortes e os limites da Resolução aqui esmiuçada.

Quadro 8. Pontos fortes e limites da Resolução CNJ n. 615/2025.

Pontos fortes	Limites
Diretrizes para segurança, transparência e eficiência no uso da IA.	Ausência de abordagem interseccional.
Princípios de justiça algorítmica e de não discriminação.	Compromissos genéricos com a igualdade de gênero.
Rastreabilidade e explicabilidade de decisões automatizadas.	Falta de metodologias para avaliação de viés algorítmico.
Compatibilidade com normas internacionais.	Auditorias não obrigatórias sensíveis a gênero.
Instituição do Comitê Nacional de IA e previsão de avaliações de impacto.	Falta de protocolos éticos feministas e de inclusão obrigatória de indicadores de gênero.

Fonte: a autora.

Em suma, a Resolução n. 615/2025 do CNJ (Conselho Nacional de Justiça, 2025c) representa um avanço normativo incontestável, mas sua eficácia social e democrática dependerá de aperfeiçoamentos institucionais e da adoção de políticas de justiça algorítmica sensível a gênero. É preciso transformar a norma em prática concreta, tornando o Poder Judiciário brasileiro não apenas tecnologicamente eficiente, mas também eticamente comprometido com a igualdade substantiva, a diversidade e os direitos humanos.

5.2.5 Regras próprias envolvendo o Poder Judiciário: princípios do devido processo legal, do juiz natural e da fundamentação das decisões judiciais

O uso de sistemas de IA no Poder Judiciário deve se submeter, em todos os momentos, às garantias do devido processo legal, do juiz natural e da fundamentação das decisões judiciais, previstas na Constituição Federal (Brasil, 1988; Siqueira et al., 2022), as quais compõem o núcleo essencial do Estado Democrático de Direito e asseguram a legitimidade das decisões jurisdicionais.

O princípio do devido processo legal, consagrado no art. 5º, inciso LIV, da Constituição Federal (Brasil, 1988), dispõe que “ninguém será privado da liberdade ou de seus bens sem o devido processo legal”. Tal dispositivo garante que toda decisão judicial resulte de um processo regular, conduzido por autoridade competente, respeitando-se as normas procedimentais, o contraditório e a ampla defesa (art. 5º, LV, CF/88). No contexto da IA, esse princípio funciona como limite material e procedimental à adoção de tecnologias automatizadas na atividade jurisdicional. Qualquer uso de IA no processo judicial deve, portanto, preservar as condições mínimas de previsibilidade, transparência e revisibilidade, de modo a evitar decisões opacas, arbitrárias ou inacessíveis à compreensão das partes.

Dessa forma, o devido processo legal impõe que as decisões judiciais permaneçam sempre sob controle humano, sendo a IA apenas um instrumento auxiliar, jamais um substituto da função jurisdicional. Como afirmam Ribeiro e Cassol (2021), “o princípio do devido processo legal serve de amparo à limitação do uso da IA na tomada de decisões judiciais”. A tecnologia deve ser empregada apenas para apoiar o juiz – por exemplo, em tarefas de pesquisa jurisprudencial, análise de padrões decisórios, verificação de precedentes, organização de dados ou elaboração de minutas – sem, contudo, interferir na deliberação e no convencimento judicial, que são atos indelegáveis e personalíssimos.

Desse princípio deriva o princípio do juiz natural, previsto no art. 5º, inciso LIII, da CF/88 (Brasil, 1988), segundo o qual “ninguém será processado nem sentenciado senão pela

autoridade competente”. Tal preceito assegura que a jurisdição seja exercida por juízes constitucionalmente investidos, impedindo a designação arbitrária de julgadores e garantindo imparcialidade e independência.

A aplicação da IA no Judiciário, portanto, não pode violar o princípio do juiz natural, devendo manter-se como instrumento de apoio técnico sob supervisão do magistrado. Em outras palavras, a decisão judicial é ato humano e soberano, e qualquer algoritmo utilizado deve operar sob controle humano efetivo e rastreabilidade de suas recomendações (Resolução CNJ n. 615/2025, art. 3º, inciso IV (Conselho Nacional de Justiça, 2025c)).

A fundamentação das decisões judiciais, por sua vez, constitui um dever constitucional e democrático do julgador. O art. 93, inciso IX, da CF/88 (Brasil, 1988) determina que “todos os julgamentos dos órgãos do Poder Judiciário serão públicos, e fundamentadas todas as decisões”, garantindo transparência, controle social e possibilidade de impugnação. No contexto da IA, esse dever implica que o juiz, ao utilizar uma ferramenta automatizada, deve ser capaz de compreender, justificar e explicar o conteúdo das recomendações algorítmicas que influenciaram a decisão. É indispensável que a lógica empregada pela IA – isto é, os dados, os parâmetros e os critérios utilizados – seja explicável, auditável e passível de revisão humana, sob pena de violação do dever de motivação.

O contraditório e a ampla defesa, igualmente previstos no art. 5º, inciso LV, da CF/88 (Brasil, 1988), também se estendem ao uso de sistemas inteligentes. As partes devem ser informadas sobre a utilização de ferramentas de IA nos processos em que figuram e ter acesso às explicações sobre a forma como tais tecnologias atuam. Conforme observam Ribeiro e Cassol (2021), “ambas as partes deverão estar cientes das técnicas de IA disponíveis e aplicadas pelo Judiciário”, o que inclui o direito de questionar erros, vieses ou inconsistências do sistema. Assim, a transparência procedimental torna-se condição essencial para assegurar o equilíbrio processual e a confiança das partes na imparcialidade da decisão.

Nesse contexto, a aplicação dos princípios do devido processo legal, do juiz natural e da fundamentação das decisões se articula diretamente com as garantias técnicas e regulatórias analisadas nos itens anteriores desta tese – como a explicabilidade, a auditabilidade e a *accountability* –, consolidando o entendimento de que a tecnologia deve estar a serviço da justiça, e não o contrário. O magistrado, ao utilizar sistemas de apoio baseados em IA, deve agir como responsável final pela decisão, garantindo que o processo de julgamento continue regido pela racionalidade pública, pela argumentação jurídica e pela observância dos direitos fundamentais.

A Resolução n. 615/2025 do CNJ reforça essa exigência ao dispor que todas as soluções de IA devem preservar a supervisão humana efetiva e a possibilidade de contestação e revisão das decisões automatizadas, ficando o juiz como responsável final pela decisão judicial, ainda que apoiada por sistema inteligente (Conselho Nacional de Justiça, 2025c). Contudo, é necessário que o CNJ e os tribunais implementem mecanismos práticos para dar efetividade a essas previsões, garantindo que os magistrados recebam capacitação técnica adequada para compreender as limitações e os potenciais vieses dos algoritmos utilizados, além de assegurar que todas as ferramentas passem por auditorias regulares e independentes.

Em síntese, a incorporação da IA ao sistema de justiça não exime o Poder Judiciário de observar suas próprias garantias constitucionais. Ao contrário, exige sua reafirmação em novos termos. O devido processo legal, o juiz natural e a fundamentação das decisões continuam a ser os pilares da legitimidade jurisdicional, e sua aplicação à IA é o que diferencia uma justiça automatizada de uma justiça humana, democrática e constitucionalmente comprometida com a dignidade e a igualdade.

5.2.6 O Protocolo para Julgamento com Perspectiva de Gênero

Tal como já exposto no item 3.3 desta tese, o “Protocolo para Julgamento com Perspectiva de Gênero” (Conselho Nacional de Justiça, 2021a) tem como objetivo central é orientar magistradas e magistrados a incorporar a perspectiva de gênero em todas as etapas do processo judicial, de modo a garantir a igualdade substantiva e evitar que decisões judiciais reproduzam estereótipos, preconceitos ou estruturas discriminatórias. O documento propõe, assim, uma mudança paradigmática na cultura jurídica, convertendo a noção tradicional de imparcialidade – baseada na suposta neutralidade do julgador – em uma imparcialidade consciente, que reconhece a existência de vieses e atua de forma ativa para mitigá-los.

O Protocolo está organizado em três partes interconectadas, que articulam bases conceituais, metodologia e aplicação prática, constituindo um instrumento que alia rigor normativo e aplicabilidade cotidiana.

Num primeiro momento, o documento apresenta os fundamentos teóricos necessários à compreensão das desigualdades de gênero como construções históricas e culturais reproduzidas, inclusive, pelo próprio sistema jurídico. Essa base conceitual fornece ao julgador os parâmetros para identificar e questionar as formas de discriminação institucionalizadas (Conselho Nacional de Justiça, 2021a, p. 16-43). Em um segundo momento, ele oferece um

roteiro metodológico prático, orientando o(a) julgador(a) a aplicar a lente de gênero desde o início do processo até a elaboração da sentença. O guia propõe a reflexão sobre como as desigualdades estruturais podem ter condicionado o conflito e como devem ser adotadas medidas corretivas, como, por exemplo, a revisão ativa de estereótipos durante a valoração da prova e a fundamentação (Conselho Nacional de Justiça, 2021a, p. 44-64).

Na sequência, a terceira parte exemplifica como a perspectiva de gênero se aplica de modo distinto nos diversos ramos do Judiciário, reconhecendo que as desigualdades assumem formas e intensidades variadas em cada contexto institucional. Na Justiça Estadual, por exemplo, foca-se em casos de violência doméstica, feminicídio e crimes sexuais, ressaltando a importância da credibilidade da palavra da vítima, da análise das dimensões psicológicas e institucionais da violência e do combate a estereótipos que culpabilizam mulheres (Conselho Nacional de Justiça, 2021a, p. 83-101) e, na Justiça do Trabalho, o documento enfatiza as questões de assédio moral e sexual, desigualdade salarial, divisão sexual do trabalho e invisibilidade do trabalho reprodutivo, propondo que julgadores considerem os efeitos dessas desigualdades nas relações laborais e na reparação de danos (Conselho Nacional de Justiça, 2021a, p. 102-117). Essa diferenciação demonstra que o julgamento com perspectiva de gênero não se resume a um modelo único, mas constitui um conjunto de práticas contextualizadas, ajustadas às realidades e desafios específicos de cada ramo da Justiça.

Diante disso, nesta etapa da tese, reafirma-se a defesa de que os conceitos e as discussões trazidos pelo Protocolo também devem servir de orientação aos desenvolvedores de soluções de IA no âmbito do Judiciário. O roteiro de combate à desigualdade de gênero, concebido originalmente para orientar a conduta judicial humana, pode e deve ser adaptado às etapas de desenvolvimento, treinamento e uso da IA no sistema de justiça. Assim como o(a) juiz(a) é chamado(a) a reconhecer e corrigir vieses cognitivos e culturais, os sistemas algorítmicos – que reproduzem padrões históricos – precisam ser programados e supervisionados com base na mesma metodologia reflexiva. Isso implica:

- a)** incorporar variáveis e indicadores desagregados por sexo, raça e classe nos bancos de dados;
- b)** adotar métricas de explicabilidade, contestabilidade e supervisão humana contínua;
- c)** e realizar avaliações de impacto algorítmico que contemplem desigualdades interseccionais.

O documento deve ser convertido em referência metodológica para a governança de IA judicial, definindo o mesmo padrão ético e processual que deve orientar tanto a atividade jurisdicional quanto a tecnologia que a apoia.

Diante da possibilidade real de que a IA perpetue desigualdades existentes, é essencial que o Protocolo seja compreendido como parâmetro normativo e ético para o desenvolvimento e a utilização de tecnologias no sistema de justiça. Integrar o documento às soluções algorítmicas é uma medida indispensável para que o avanço tecnológico não retroceda na garantia dos direitos das mulheres e na luta contra o viés de gênero.

Nesse contexto, o Protocolo deve ser entendido como uma referência ética para a governança algorítmica. Se a ausência de perspectiva de gênero nas decisões humanas gera discriminações institucionais, a ausência dessa lente crítica na programação, nos dados e nos critérios de decisão das máquinas pode ampliar e automatizar os mesmos vieses. O documento, ao rejeitar a falsa neutralidade, oferece uma lente crítica que deve orientar também o *design*, a curadoria e a auditoria da IA judicial, vinculando o uso de sistemas inteligentes aos valores constitucionais de igualdade, dignidade da pessoa humana e não discriminação.

Com base nessa perspectiva, apresenta-se o Quadro 9, que resume as razões pelas quais o documento deve ser incorporado como diretriz ética e metodológica para o desenvolvimento e uso da IA no Poder Judiciário brasileiro:

Quadro 9. Razões para a incorporação do Protocolo para Julgamento com Perspectiva de Gênero na IA aplicada dentro do Poder Judiciário.

Razões	Resumo argumentativo	Argumento central
A IA como reprodutora (e amplificadora) de vieses estruturais	A IA é treinada com dados históricos que refletem desigualdades estruturais de gênero no Judiciário brasileiro. A composição hegemônica da magistratura influencia o padrão decisório, que é replicado e ampliado por sistemas algorítmicos.	Se as decisões judiciais estão marcadas por vieses de gênero, a IA pode não apenas perpetuá-los, mas ampliá-los em razão de sua escala e aparência de imparcialidade.
A ilusão da neutralidade algorítmica e o risco da falsa objetividade	O uso da IA tende a mascarar os vieses sob uma aparência tecnocientífica de objetividade. O protocolo reconhece que o Direito pode reproduzir desigualdades quando aplicado	O viés androcêntrico pode ser codificado sob a pretensa neutralidade da IA. O protocolo deve ser incorporado como diretriz ética e metodológica desde o

	sem criticidade, o que se agrava com a automação.	início dos sistemas algorítmicos.
A necessidade de uma lente interseccional e sensível às vulnerabilidades	O protocolo oferece uma abordagem interseccional que considera múltiplas formas de opressão. A IA, se desenvolvida sem essa lente, pode resultar em decisões homogeneizantes e insensíveis às desigualdades materiais dos sujeitos.	Sem uma perspectiva interseccional, a IA tende a ignorar as vulnerabilidades sociais. O protocolo é essencial para orientar a IA com base na justiça substantiva.
A responsabilidade institucional do Judiciário na promoção da justiça de gênero	O CNJ reconhece o dever da magistratura em promover a equidade de gênero. Esse dever deve se estender aos sistemas tecnológicos utilizados, assegurando que também eles estejam alinhados aos valores constitucionais de igualdade.	Sistemas de IA que auxiliam a jurisdição devem estar vinculados às mesmas diretrizes normativas e éticas que orientam a atuação dos(as) juízes(as), incluindo o Protocolo para Julgamento com Perspectiva de Gênero (Conselho Nacional de Justiça, 2021a).
O protocolo como marco ético-jurídico para a IA judicial	O protocolo não é apenas interpretativo, mas uma mudança paradigmática na forma de julgar. Incorporá-lo à IA significa conectar a inovação tecnológica ao compromisso com os direitos humanos e com a justiça social.	A adoção do protocolo na IA judicial permite alinhar tecnologia, responsabilidade democrática e justiça de gênero.
A perspectiva feminista crítica como chave para o controle democrático da IA	A abordagem feminista crítica reconhece que o Direito é estruturado por assimetrias de gênero. A IA pode reforçar esses filtros androcêntricos sob um verniz técnico, legitimando discriminações.	A incorporação do protocolo é essencial para evitar que a IA se torne um novo instrumento de exclusão e para que atue como ferramenta de transformação institucional comprometida com a justiça de gênero.

Fonte: a autora.

Por meio dessa integração, o Protocolo deixa de ser apenas uma ferramenta interpretativa e assume a função de marco de governança algorítmica feminista, promovendo um modelo de inovação tecnológica que se alinha à ética, aos direitos humanos e à responsabilidade democrática do Judiciário.

5.2.7 Quadro resumo das garantias regulatórias para uma inteligência artificial ética e feminista no Poder Judiciário

Quadro 10. Garantias regulatórias para uma IA ética, responsável e alinhada à proteção dos direitos humanos e princípios jurídicos fundamentais.

GARANTIAS REGULATÓRIAS PARA UMA IA ÉTICA NO JUDICIÁRIO	DESCRIÇÃO CONCEITUAL E TÉCNICA
O papel dos direitos humanos	Os direitos humanos constituem o alicerce jurídico e ético para o desenvolvimento e a aplicação de tecnologias de IA. São universais, inalienáveis e vinculativos, funcionando como mecanismos de contenção do poder tecnológico e de reparação em caso de violações. Sua incorporação assegura que a IA no Judiciário respeite a dignidade humana, a igualdade substantiva e a não discriminação. Devem orientar políticas de curadoria ética de dados, diversidade nas equipes técnicas e auditorias contínuas, permitindo identificar e mitigar vieses de gênero e outras formas de exclusão estrutural.
O papel da regulamentação geral sobre ética e inteligência artificial	A regulamentação geral é essencial para estabelecer princípios universais de governança da IA: centralidade do ser humano, inclusão social, responsabilidade, transparência, governança e <i>accountability</i>. No contexto brasileiro, a ausência de lei específica demanda que o Judiciário adote parâmetros próprios, adaptando as boas práticas internacionais às realidades institucionais e regionais, especialmente com foco em equidade de gênero e diversidade.
Governança dos sistemas inteligentes	A governança de dados constitui o núcleo operativo da regulação prática da IA, garantindo qualidade, integridade, rastreabilidade e segurança das informações. Define políticas, papéis e controles sobre todo o ciclo de vida dos dados – da coleta ao descarte – e viabiliza princípios de transparência, explicabilidade e justiça algorítmica. No Judiciário, é o mecanismo que assegura que decisões assistidas por IA sejam auditáveis, previsíveis e isentas de vieses discriminatórios. Inclui curadoria ética, <i>logs</i> automáticos, auditorias periódicas e participação pública. A ausência de governança implica risco de opacidade e arbitrariedade, contrariando o devido processo legal e o dever de fundamentação das decisões.

Resolução CNJ n. 615/2025	Estabelece normas sobre desenvolvimento, auditoria, monitoramento e uso responsável da IA, com base em princípios de segurança, transparência, justiça e supervisão humana. Determina a adoção de processos internos de fiscalização, auditorias regulares, avaliações de impacto e mecanismos de explicabilidade e institui o Comitê Nacional de IA do Judiciário. No entanto, carece de abordagens interseccionais de gênero, protocolos éticos feministas e obrigatoriedade de indicadores sensíveis ao gênero. Seu aprimoramento é indispensável para a efetivação de uma justiça algorítmica inclusiva, auditável e igualitária.
Regras próprias envolvendo o Poder Judiciário: devido processo legal, juiz natural e fundamentação das decisões judiciais	A aplicação da IA na jurisdição deve observar integralmente os princípios constitucionais que estruturam o Estado Democrático de Direito. O devido processo legal assegura que toda decisão siga rito regular e seja compreensível; o princípio do juiz natural preserva que apenas magistrados constitucionalmente investidos exerçam jurisdição; e o dever de fundamentação das decisões impõe transparência e controle social. A IA, portanto, deve atuar exclusivamente como ferramenta auxiliar, sob supervisão humana e rastreabilidade, sem substituir o discernimento judicial. O contraditório e a ampla defesa exigem que as partes sejam informadas sobre o uso de IA e possam contestar seus resultados, garantindo legitimidade e confiança pública.
O Protocolo para Julgamento com Perspectiva de Gênero	O Protocolo é instrumento normativo e pedagógico que orienta a magistratura a julgar com consciência das desigualdades estruturais de gênero, promovendo uma justiça emancipatória e interseccional. Estruturado em três partes – conceitos, guia metodológico e aplicações por ramos da justiça –, fornece parâmetros para identificar e corrigir vieses nos processos decisórios. No contexto da IA judicial, o Protocolo deve ser incorporado como referência ética e metodológica para o desenvolvimento, treinamento e auditoria de sistemas, prevenindo a reprodução automatizada de estereótipos e desigualdades históricas. Sua integração à governança algorítmica fortalece o compromisso do Judiciário com a igualdade de gênero, a diversidade e os direitos humanos, transformando a tecnologia em instrumento de inclusão e não de exclusão.

As garantias regulatórias constituem o eixo normativo da governança ética da IA no Poder Judiciário, pois traduzem os valores jurídicos fundamentais – como dignidade, igualdade e devido processo legal – em parâmetros concretos de desenvolvimento e uso da tecnologia. Essas garantias operam como um sistema de freios e contrapesos aplicado à IA, assegurando que sua incorporação às práticas judiciais ocorra sob supervisão humana, transparência e responsabilidade institucional. A centralidade dos direitos humanos estabelece a base dessa estrutura: a IA deve ser projetada e auditada para respeitar a dignidade e a diversidade humanas, prevenindo a reprodução de discriminações e assegurando a reparação em caso de violações. Esse compromisso ético se materializa na governança de dados, que define padrões de qualidade, integridade e rastreabilidade das informações, bem como na regulamentação geral da ética em IA, que consagra princípios universais como inclusão, explicabilidade e *accountability*. Assim, o arcabouço regulatório permite alinhar a inovação tecnológica ao Estado Democrático de Direito, garantindo que a automação judicial não comprometa direitos fundamentais, mas os fortaleça.

Em destaque, no contexto da presente tese, as garantias regulatórias também são interpretadas como mecanismos de enfrentamento dos vieses de gênero e de promoção da justiça substantiva. A Resolução CNJ n. 615/2025 representa um marco nesse processo ao determinar a obrigatoriedade de auditorias, avaliações de impacto algorítmico e supervisão humana, ainda que careça de uma abordagem interseccional sensível às desigualdades de gênero. Sua complementação pelo “Protocolo para Julgamento com Perspectiva de Gênero” amplia a dimensão ética da regulação, inserindo no debate tecnológico o compromisso com a equidade e a diversidade.

Incorporar o Protocolo à IA judicial significa fazer dos princípios constitucionais – como o devido processo legal, o juiz natural e a fundamentação das decisões – não apenas garantias formais, mas condições materiais de uma governança algorítmica feminista, que reconhece e corrige as assimetrias históricas reproduzidas pelos sistemas automatizados. Desse modo, as garantias regulatórias se consolidam como o elo entre técnica e ética, unindo a legitimidade jurídica da decisão judicial ao imperativo democrático de uma justiça digital inclusiva, auditável e comprometida com os direitos humanos.

5.3 Garantias políticas para uma inteligência artificial ética e feminista no Poder Judiciário brasileiro

A efetividade das garantias técnicas e regulatórias depende de um ambiente político-institucional capaz de sustentar, difundir e fiscalizar padrões éticos no ciclo de vida da IA. Por isso, este subcapítulo organiza um conjunto de garantias políticas que dão lastro à governança algorítmica e a conectam com valores democráticos, direitos humanos e justiça de gênero. Em síntese, trata-se de alinhar a transformação digital do Judiciário a práticas contínuas de formação crítica, transparência pública e responsabilidade institucional.

5.3.1 Conscientização e letramento tecnológico

A inclusão digital é parte indissociável do desenvolvimento humano em uma sociedade da informação. Como aponta Freitas (2012, p. 52), ela relaciona-se à inclusão social, envolvendo cidadania, democracia e desenvolvimento social, científico, econômico e ambiental. Assim, construir uma sociedade tecnológica democrática implica garantir que todos sejam não apenas usuários de tecnologia, mas sujeitos conscientes e participativos em seu uso e fiscalização.

Levando isso em consideração, a conscientização e o letramento digital são pilares indispensáveis para que a IA seja aplicada de forma ética, transparente e responsável no Poder Judiciário brasileiro. A consolidação de uma governança algorítmica confiável depende não apenas de regulações técnicas e institucionais, mas da formação crítica dos sujeitos que desenvolvem, operam e fiscalizam essas tecnologias. A ética da IA no Judiciário só se materializa quando há compreensão plena de suas funcionalidades, implicações e limitações – o que exige conhecimento técnico, sensibilidade social e responsabilidade pública.

A implementação da IA no sistema de justiça brasileiro é ainda recente e o país carece de orientações sistematizadas sobre como utilizá-la, para quais finalidades e com quais limites éticos. Isso gera preocupação, pois, como adverte Oliveira (2019, p. 15), a automação e a implementação de IA no Judiciário devem necessariamente considerar a falta de letramento tecnológico entre estudantes e profissionais do campo jurídico, sob pena de acarretar perda de eficiência e comprometimento da prestação dos serviços judiciais. Essa constatação reforça que a alfabetização digital no Direito não é um luxo acadêmico, mas uma condição de legitimidade e eficácia institucional.

Além disso, o letramento digital deve abranger não apenas os magistrados e demais servidores da justiça, mas também todos os cidadãos. A sociedade tem o direito de saber que sistemas de IA estão sendo utilizados pelo Judiciário, como operam, quais dados processam e de que modo influenciam as decisões judiciais. Essa transparência é condição para o exercício pleno da cidadania digital, entendida como a expansão, para o ambiente tecnológico, dos direitos de informação, participação e controle social sobre as ações do Estado.

Carvalho e Américo (2014, p. 77) complementam essa visão ao sustentar que o avanço tecnológico, por si só, é insuficiente: é preciso combiná-lo com educação, políticas públicas inclusivas, distribuição de renda justa e desenvolvimento de habilidades tecnológicas. A cidadania digital é, portanto, o exercício consciente de direitos e deveres no espaço digital, um conceito que ultrapassa o acesso técnico às ferramentas e envolve autonomia crítica e ética informacional.

Nessa perspectiva, Choi (2016, p. 9) define a cidadania digital em quatro dimensões interdependentes: a ética, relacionada ao uso consciente e responsável da tecnologia; a resistência crítica, que envolve o questionamento de estruturas de poder e desigualdade; a participação, que traduz o engajamento político, econômico e cultural nas redes; e a alfabetização digital, que diz respeito à capacidade de acessar, interpretar e manejar informações digitais. No Judiciário, essa alfabetização deve significar não apenas saber utilizar sistemas tecnológicos, mas compreender como eles impactam o exercício da justiça e quais riscos éticos e sociais podem emergir de seu uso automatizado.

Ademais, a Resolução n. 615/2025 do CNJ (Conselho Nacional de Justiça, 2025c) reconhece expressamente essa necessidade ao tratar do letramento e da capacitação continuada de magistrados e servidores (art. 2º, X e VIII). O art. 19, §3º, estabelece que os usuários de LLMs e IAGs devem “realizar capacitação e treinamentos específicos sobre melhores práticas, limitações, riscos e uso ético, responsável e eficiente” das ferramentas de IA, cabendo aos tribunais e suas escolas promover treinamentos continuados. O §4º determina que o Comitê Nacional de Inteligência Artificial do Judiciário elabore um manual de boas práticas em linguagem simples e o §5º impõe aos tribunais a tarefa de manter programas de capacitação permanente e atualização tecnológica. Tais dispositivos são expressão normativa da consciência de que sem letramento digital, não há ética digital nem justiça algorítmica efetiva.

Entretanto, como demonstrado no subcapítulo 4.1 desta tese, a pesquisa *O uso da Inteligência Artificial Generativa no Poder Judiciário Brasileiro* (Conselho Nacional de Justiça, 2024b) revelou que apenas 9,1% dos magistrados e 6,0% dos servidores responderam

ao questionário sobre o uso de IAs generativas, o que indica baixo engajamento e insegurança diante do tema. Embora cerca da metade dos respondentes tenha experimentado ferramentas como o ChatGPT, o uso cotidiano ainda é incipiente e pouco institucionalizado (Conselho Nacional de Justiça, 2024b, p. 49–60). Essa resistência decorre, em parte, da ausência de formação específica e de orientação institucional, o que reforça a urgência de políticas estruturadas de capacitação ética e técnica.

Nesse ponto, destaca-se também a advertência da UNESCO e do IRCAI (2024, p. 4): é essencial promover a conscientização pública e a educação sobre ética e vieses da IA, capacitando os usuários a interagir criticamente com as tecnologias e a exigir a proteção de seus direitos. Essa orientação internacional deve servir de base para o Comitê Nacional de Inteligência Artificial do Judiciário e para as escolas judiciais, que precisam desenvolver cursos e materiais voltados não apenas ao uso instrumental da IA, mas também à reflexão ética e crítica sobre seu impacto social.

Nesse sentido, propõe-se que o CNJ estabeleça programas nacionais de letramento digital em IA para todos os agentes envolvidos no sistema de justiça – magistrados, servidores, advogados, defensores, membros do Ministério Público e peritos –, com módulos diferenciados para cada público. Além disso, deve ser criado um canal público permanente de esclarecimento, de fácil acesso, destinado a responder dúvidas sobre os sistemas de IA em desenvolvimento ou em uso no Judiciário. Essa estrutura de suporte técnico e pedagógico é imprescindível para reduzir a assimetria de informação e ampliar a confiança institucional.

O letramento digital e a conscientização ética devem integrar de forma transversal o “Programa Justiça 4.0” (Conselho Nacional de Justiça, [20-?]), que visa aproximar o sistema judiciário da sociedade ao disponibilizar novas tecnologias e soluções baseadas em IA. O programa, ao automatizar atividades e otimizar o trabalho de magistrados e servidores, deve igualmente garantir que todos compreendam como essas ferramentas funcionam, para que servem e quais são seus limites éticos e jurídicos. Inserir o letramento digital como eixo estruturante do Justiça 4.0 significa reconhecer que a transformação digital só é legítima quando acompanhada de formação crítica e controle humano informado.

É nessa conjectura que se destaca também a importância da Resolução n. 654, de 4 de novembro de 2025, do CNJ que, de maneira inovadora, traz em seu art. 5º, IX, “a”, a responsabilidade digital e o uso de tecnologia, especificamente, o uso adequado de ferramentas digitais, incluindo a IA, como elemento a ser considerado para fins de avaliação de magistrados em vitaliciamento (Conselho Nacional de Justiça, 2025d).

Por fim, sob a perspectiva feminista, a conscientização e o letramento digital desempenham papel estratégico na prevenção e no combate a vieses de gênero e outras formas de discriminação. Ao compreenderem os mecanismos de treinamento e decisão algorítmica, magistrados e servidores tornam-se capazes de detectar distorções e corrigir desigualdades reproduzidas por sistemas automatizados. O letramento ético em IA, portanto, é não apenas uma ferramenta de eficiência administrativa, mas um instrumento de justiça social, indispensável para assegurar que o progresso tecnológico sirva à promoção da igualdade, da dignidade humana e da confiança pública no Poder Judiciário.

5.3.2 Bem-estar socioambiental quanto ao desenvolvimento e uso de sistemas inteligentes pelo Poder Judiciário

A dimensão socioambiental do desenvolvimento tecnológico é um dos eixos estruturantes da ética aplicada à IA. No contexto do Poder Judiciário, pensar a IA de forma sustentável significa não apenas assegurar que os sistemas sejam eficientes, transparentes e justos, mas também que contribuam para o bem-estar humano e ecológico, respeitando os limites planetários e as condições sociais que garantem uma vida digna para as gerações presentes e futuras.

O desenvolvimento e o uso da IA devem ir além da mera garantia negativa de não violação de direitos fundamentais, buscando a promoção ativa desses direitos em sua dimensão social, ambiental e democrática. Uma IA eticamente orientada deve operar como instrumento de solidariedade e equidade, fomentando soluções que melhorem a qualidade de vida, ampliem o acesso à justiça e reduzam desigualdades, sem causar degradação ambiental nem ampliar as assimetrias socioeconômicas.

Sob essa perspectiva, a Resolução n. 400/2021 do CNJ (Conselho Nacional de Justiça, 2021c), que institui a “Política de Sustentabilidade do Poder Judiciário”, oferece parâmetros para uma governança tecnológica responsável. Essa norma estabelece que todos os órgãos do Judiciário devem adotar práticas que reduzam impactos ambientais, promovam a gestão eficiente de recursos e estimulem a inovação sustentável.

Ao relacionar princípios de responsabilidade socioambiental, de mitigação dos impactos ambientais, de consumo consciente de energia e recursos naturais e de inclusão social à governança da IA, é possível afirmar que a sustentabilidade tecnológica não se limita à eficiência energética dos servidores ou à redução de resíduos eletrônicos, mas envolve também

o impacto social, ético e democrático de cada solução inteligente. O art. 2º, §5º da Resolução acima mencionada prevê expressamente o estímulo às ações de inovação relacionadas à tecnologia e sustentabilidade, diretriz que se aplica diretamente aos projetos de IA conduzidos pelos tribunais (Conselho Nacional de Justiça, 2021c):

Art. 2º Os órgãos do Poder Judiciário devem adotar modelos de gestão organizacional com processos estruturados que promovam a sustentabilidade, com base em ações ambientalmente corretas, economicamente viáveis, socialmente justas e inclusivas, culturalmente diversas e inovadoras, pautadas na integridade, em busca de um desenvolvimento nacional sustentável.

[...] § 5º As ações inovadoras devem promover a adoção de novas tecnologias, processos e métodos que contribuam para a eficiência, a sustentabilidade e a melhoria contínua na gestão pública, com foco na redução de impactos ambientais, na ampliação das ferramentas de acessibilidade, na otimização de recursos e na melhoria da qualidade dos serviços prestados.

Portanto, os sistemas de IA desenvolvidos ou utilizados pelo Poder Judiciário devem observar padrões de sustentabilidade socioambiental, considerando não apenas a eficiência técnica e o desempenho algorítmico, mas também o ciclo de vida completo da tecnologia: consumo energético de treinamento, infraestrutura computacional, descarte de equipamentos, origem e tratamento dos dados, além do impacto social de sua implementação. Tais preocupações são coerentes com o que estabelece a Resolução CNJ n. 615/2025, ao incluir entre os princípios estruturantes da governança algorítmica a responsabilidade institucional, a justiça, a inclusão e a não discriminação, que se estendem à dimensão ambiental (Conselho Nacional de Justiça, 2025b).

Em paralelo, as diretrizes internacionais de ética em IA reforçam esse dever. A UNESCO (2022, p. 19-21) define como um dos valores fundamentais para o desenvolvimento da IA de maneira ética a prosperidade ambiental e ecossistêmica, juntamente com os princípios da proporcionalidade e da sustentabilidade. De modo semelhante, a OCDE (OECD.AI Policy Observatory; GPAI Global Partnership, [20–]) propõe o princípio do “crescimento, sustentabilidade e bem-estar”, que orienta o desenvolvimento e o uso da IA para promover a prosperidade e gerar resultados positivos para as pessoas e para o planeta.

Tais princípios devem orientar o Judiciário brasileiro a adotar técnicas de IA ecologicamente corretas, energeticamente eficientes e socialmente justas, alinhadas aos Objetivos de Desenvolvimento Sustentável (ODS) da Agenda 2030 da ONU, especialmente os ODS 5 (igualdade de gênero), 9 (inovação e infraestrutura), 13 (ação climática) e 16 (paz, justiça e instituições eficazes) (Nações Unidas Brasil, [20-?]).

Além de seu impacto ambiental direto, a IA também produz efeitos intangíveis sobre o bem-estar social e psicológico. Embora possa ser utilizada para aprimorar habilidades

humanas e otimizar fluxos de trabalho, ela também pode gerar sobrecarga cognitiva, alienação, ansiedade e desumanização das interações institucionais, especialmente quando aplicada a processos decisórios automatizados. Assim, a sustentabilidade da IA envolve também o cuidado com a saúde mental e emocional dos usuários e jurisdicionados, assegurando que a automação não substitua o contato humano essencial à justiça.

O compromisso com o bem-estar socioambiental requer ainda que o Poder Judiciário incentive pesquisas em IA verde (*Green AI*), voltadas à eficiência energética de modelos e *data centers*, à reutilização de hardware, à priorização de algoritmos de baixo consumo e à compensação de emissões de carbono. Igualmente, recomenda-se que os tribunais elaborem relatórios de impacto socioambiental para novos sistemas de IA, avaliando não apenas o desempenho técnico, mas também seus efeitos ecológicos e sociais.

Em síntese, desenvolver uma IA ética no Judiciário implica reconhecer que sustentabilidade, justiça social e democracia são dimensões indissociáveis. A IA deve ser usada para beneficiar toda a humanidade, inclusive as futuras gerações, preservando processos democráticos, fortalecendo a deliberação humana e respeitando a diversidade de valores e modos de vida.

5.2.3 Manutenção da conscientização a respeito da disparidade de gênero e a luta constante pela igualdade

A incorporação da IA ao Poder Judiciário não pode ser dissociada da necessidade permanente de conscientização sobre as disparidades de gênero que atravessam as instituições e suas práticas. A desigualdade de gênero é estrutural e manifesta-se de múltiplas formas. Assim, qualquer esforço de modernização tecnológica deve caminhar junto à promoção de uma cultura institucional comprometida com a igualdade substantiva.

Deve-se manter uma consciência crítica, demandando que o Poder Judiciário adote uma postura proativa de vigilância e aperfeiçoamento constante de suas práticas institucionais. Isso inclui a criação de espaços de formação e diálogo permanentes sobre gênero e diversidade, o incentivo à pesquisa interdisciplinar sobre desigualdades no campo jurídico e tecnológico e o fortalecimento de políticas internas de promoção da equidade. Mais do que adotar diretrizes pontuais, é necessário consolidar uma cultura organizacional igualitária, que valorize a pluralidade de vozes e experiências na formulação e na execução de projetos de IA.

A luta pela igualdade no âmbito judicial deve ser compreendida como processo dinâmico e cumulativo. Ela se expressa tanto na representatividade das mulheres nos espaços de decisão e de produção tecnológica quanto na incorporação de perspectivas feministas e interseccionais nos parâmetros de governança da IA. A igualdade, nesse contexto, não se resume à presença numérica, mas envolve a efetiva participação qualificada de mulheres e minorias nos comitês técnicos, nas equipes de desenvolvimento e nas instâncias de auditoria e supervisão dos sistemas inteligentes.

Além disso, a conscientização sobre disparidade de gênero deve estar integrada à educação ética e tecnológica de magistrados, servidores e desenvolvedores, estimulando uma compreensão crítica sobre como decisões automatizadas podem impactar de maneira desigual diferentes grupos sociais. O reconhecimento dessas assimetrias é o primeiro passo para a construção de soluções que respeitem a diversidade e promovam a justiça substantiva.

Nesse sentido, o Poder Judiciário deve assumir o papel de agente transformador, utilizando a IA não apenas como ferramenta de eficiência administrativa, mas como instrumento de equidade. Isso requer o desenvolvimento de políticas institucionais que articulem inovação tecnológica, direitos humanos e igualdade de gênero, garantindo que os benefícios da automação sejam distribuídos de forma justa e inclusiva.

A manutenção da conscientização e a luta constante pela igualdade são dimensões complementares da governança ética da IA. O primeiro elemento – a conscientização – assegura que as desigualdades sejam reconhecidas e analisadas criticamente; o segundo – a luta pela igualdade – transforma essa consciência em ação institucional e política. Apenas por meio dessa articulação contínua será possível construir um Judiciário tecnologicamente avançado, mas também socialmente justo, plural e democrático.

5.2.4 Quadro resumo das garantias políticas para uma inteligência artificial ética e feminista no Poder Judiciário

Quadro 11. Garantias políticas para uma IA ética, inclusiva e alinhada à sustentabilidade socioambiental, protegendo direitos humanos fundamentais.

GARANTIAS POLÍTICAS PARA UMA IA ÉTICA NO JUDICIÁRIO	DESCRIÇÃO CONCEITUAL E TÉCNICA
Conscientização e letramento tecnológico	Capacitar profissionais e usuários para interagir com sistemas de IA de forma crítica e consciente. Garantir educação continuada sobre funcionalidades, limitações, riscos e vieses,

	materiais em linguagem simples, trilhas formativas por perfil (magistratura, servidores, advocacia, Ministério Público e Defensoria Pública) e canal público de esclarecimento sobre sistemas em desenvolvimento/uso. A conscientização e o letramento tecnológico fortalecem o controle humano informado, a transparência e a proteção de direitos.
Bem-estar socioambiental quanto ao desenvolvimento e uso de sistemas inteligentes pelo Poder Judiciário	Orientar o desenvolvimento e o uso da IA por critérios de sustentabilidade, mitigando impactos ambientais e promovendo benefícios sociais mensuráveis.
Manutenção da conscientização a respeito da disparidade de gênero e a luta constante pela igualdade	Institucionalizar uma vigilância permanente sobre desigualdades de gênero na cultura organizacional e nos sistemas de IA: formação continuada, espaços de diálogo, indicadores de gênero nos monitoramentos, participação qualificada de mulheres e grupos sub-representados em comitês técnicos e auditorias. Visa prevenir a naturalização de vieses nos dados/modelos e garantir que a automação avance a igualdade substantiva, não a reproduza.

Fonte: a autora.

As garantias políticas representam a dimensão social e transformadora da ética da IA no Poder Judiciário, voltando-se à construção de uma cultura institucional capaz de compreender criticamente as implicações da tecnologia e de orientá-la para a promoção dos direitos humanos, da equidade e da sustentabilidade. Elas complementam as garantias técnicas e regulatórias ao introduzir o elemento humano e pedagógico na governança da IA, assegurando que magistrados, servidores e demais operadores do Direito estejam aptos a interagir com os sistemas de forma consciente e informada.

A conscientização e o letramento tecnológico configuram instrumentos indispensáveis para democratizar o conhecimento sobre o funcionamento dos algoritmos, suas limitações e riscos. Ao promover a educação digital contínua e o acesso público à informação sobre os sistemas utilizados, essas medidas fortalecem o controle humano informado, a transparência e a confiança social, prevenindo o uso acrítico da tecnologia e garantindo o exercício efetivo do contraditório e da ampla defesa frente às decisões automatizadas.

No contexto desta tese, as garantias políticas também operam como mecanismos de sustentabilidade e justiça social, vinculando a inovação tecnológica aos princípios

constitucionais da dignidade humana, da igualdade de gênero e da proteção ambiental. A incorporação de critérios de bem-estar socioambiental orienta o desenvolvimento da IA judicial para a mitigação de impactos ecológicos e a maximização de benefícios sociais mensuráveis, alinhando-se aos objetivos de desenvolvimento sustentável (ODS) da Agenda 2030 da ONU. Paralelamente, a manutenção da conscientização sobre a disparidade de gênero assume caráter permanente e institucional, promovendo a formação continuada, a inclusão de mulheres e grupos sub-representados em comitês técnicos e o monitoramento sistemático de indicadores de gênero. Essas práticas impedem a naturalização de estereótipos nos dados e modelos algorítmicos e asseguram que a automação judicial avance a igualdade substantiva e não a desigualdade. Assim, as garantias políticas consolidam o compromisso do poder judiciário com uma ia ética, inclusiva e sustentável, que combina eficiência tecnológica com responsabilidade democrática e justiça de gênero.

CAPÍTULO 6 – PROTOCOLOS PARA O CUMPRIMENTO DE GARANTIAS ÉTICAS PARA O DESENVOLVIMENTO E O USO DA INTELIGÊNCIA ARTIFICIAL NO PODER JUDICIÁRIO

As garantias éticas propostas no Capítulo 5 constituem um marco normativo e técnico essencial para orientar o desenvolvimento e a utilização de sistemas de IA em contextos judiciais. Contudo, a efetivação dessas garantias requer instrumentos operacionais que permitam traduzir princípios abstratos em práticas verificáveis e rotinas institucionais.

Neste capítulo, são apresentados dois protocolos complementares, concebidos como instrumentos exemplificativos de cumprimento das garantias discutidas anteriormente. O primeiro, voltado a desenvolvedores e equipes técnicas, organiza as etapas e parâmetros necessários para assegurar que as soluções de IA respeitem as normas éticas e os direitos fundamentais. O segundo, destinado a magistrados, assessorias e demais usuários, oferece diretrizes para o uso consciente e fundamentado dessas ferramentas, reforçando a centralidade do juízo humano e o compromisso com a igualdade de gênero e com a transparência institucional.

Ambos os protocolos (Apêndice A e Apêndice B) foram estruturados com auxílio do ChatGPT, observando-se o “Guia orientador: Reflexão sobre o uso da inteligência artificial na pesquisa acadêmica” (UNIVERSIDADE ESTADUAL DE PONTA GROSSA, [2025?]). Os documentos possuem formato de *checklists* e perguntas orientadoras, de modo a permitir avaliações objetivas de conformidade e auditoria contínua. Além de operacionalizar as garantias éticas, esses instrumentos estimulam a cultura da responsabilidade compartilhada entre agentes técnicos e jurídicos, contribuindo para consolidar uma governança democrática da IA judicial.

6.1 Protocolo Ético para Desenvolvedores e Equipes Técnicas dos Sistemas Inteligentes Judiciais

A dimensão técnica da ética em IA corresponde ao nível mais estrutural das garantias: é nela que se definem os parâmetros que determinarão o comportamento, a robustez e os potenciais vieses dos sistemas. O Protocolo Ético para Desenvolvedores e Equipes Técnicas dos Sistemas Inteligentes Judiciais, cujo modelo está incluído no Apêndice A, propõe um conjunto de etapas práticas que visam assegurar que todo o ciclo de vida de uma IA judicial –

desde a coleta de dados até a implantação e monitoramento – seja conduzido sob os princípios da justiça algorítmica, transparência, explicabilidade e responsabilidade.

Mais do que um instrumento técnico, esse protocolo representa uma forma de governança ética aplicada à tecnologia pública. Ele permite que equipes de desenvolvimento, universidades, fornecedores e órgãos contratantes trabalhem sob critérios verificáveis, documentados e auditáveis, reforçando a legitimidade institucional das soluções algorítmicas utilizadas pelo Poder Judiciário.

6.1.1 Planejamento ético e análise de impacto

A primeira etapa refere-se à avaliação prévia dos impactos éticos e sociais do sistema, por meio da elaboração de um Relatório de Impacto Algorítmico (RIA). Esse documento deve mapear potenciais riscos à privacidade, à igualdade de tratamento e à autonomia dos usuários, além de propor medidas mitigadoras e mecanismos de supervisão contínua.

Essa fase deve envolver a constituição de um comitê multidisciplinar – com representação de gênero, raça e áreas diversas – para revisar periodicamente os objetivos e as hipóteses de uso do sistema. O planejamento ético também deve considerar a possibilidade de exclusão digital, avaliando se populações vulneráveis (como mulheres em situação de violência, pessoas com deficiência ou residentes em áreas rurais) poderão ser indiretamente afetadas pelo uso da tecnologia.

6.1.2 Curadoria e governança de dados

A governança de dados é um eixo central da justiça algorítmica. O protocolo propõe que os conjuntos de dados utilizados sejam representativos da diversidade social e territorial, evitando distorções regionais e replicações de desigualdades históricas. É necessário documentar todo o *data lineage*, isto é, a origem, a transformação e o destino dos dados, garantindo rastreabilidade e auditabilidade.

Também é fundamental a remoção de atributos correlacionados a gênero, raça ou classe que não tenham relação causal com o mérito jurídico das decisões. Por exemplo, campos como “estado civil” ou “ocupação” podem funcionar como *proxies* de gênero, influenciando indevidamente os resultados.

Testes de viés devem ser realizados já nessa etapa, observando a distribuição estatística das variáveis sensíveis (sexo, raça, território, idade), para identificar eventuais padrões de discriminação antes da modelagem.

6.1.3 Desenvolvimento, testes e validação ética

Durante o desenvolvimento, o modelo deve ser avaliado segundo métricas de fairness amplamente reconhecidas: *Demographic Parity* (paridade demográfica), *Equal Opportunity* (igualdade de oportunidades), *Equalized Odds* (probabilidades igualadas) e *Predictive Parity* (paridade de previsão). Cada uma delas reflete uma dimensão distinta da justiça algorítmica – ora orientada à igualdade de resultados, ora à igualdade de acertos.

Esses testes devem ser complementados por auditorias independentes e por métodos de explicabilidade local, como LIME ou SHAP, que permitem identificar as variáveis mais determinantes para cada decisão individual. Essa prática aumenta a transparência e viabiliza o controle externo das decisões automatizadas.

Como exemplo, em um sistema de apoio à decisão sobre liberdade provisória, o modelo deve ser testado para garantir que a TPR e a FPR seja similar entre homens e mulheres. Caso haja discrepâncias significativas, o sistema deve ser revisto antes da implantação.

6.1.4 Implementação e monitoramento contínuo

A fase de implementação exige a manutenção de *logs* detalhados, versionamento do modelo e documentação pública acessível sobre seus objetivos, limites e bases de treinamento. Essa transparência técnica deve estar acompanhada de auditorias periódicas, realizadas preferencialmente por instituições externas ou pelo próprio CNJ.

Além disso, deve ser disponibilizado um canal de denúncia ética, no qual usuários ou partes interessadas possam relatar comportamentos indevidos, vieses percebidos ou decisões injustas produzidas pelo sistema. O monitoramento ético, portanto, deve ser permanente e responsivo, garantindo que o sistema se mantenha confiável ao longo do tempo.

6.1.5 Responsabilidade e prestação de contas

A responsabilização é o elemento final e mais decisivo do protocolo. Cada sistema de IA deve ter um responsável técnico identificado, com assinatura digital e registro público, de modo que eventuais falhas possam ser rastreadas. Devem existir planos de contingência para casos de erro ou dano ético, prevendo a suspensão imediata do uso e a revisão extraordinária do modelo.

Por fim, é recomendável que todas as interfaces e comunicações do sistema passem por revisão de acessibilidade e neutralidade de gênero, assegurando que o *design* e a linguagem não reforcem estereótipos ou exclusões simbólicas.

Assim, o protocolo para desenvolvedores traduz, na prática, os princípios da justiça algorítmica e da responsabilidade institucional, deslocando o debate ético do plano abstrato para o operacional.

6.2 Protocolo Ético para Magistrados, Assessorias e Demais Usuários dos Sistemas Inteligentes

Se o primeiro protocolo se ocupa da dimensão técnica da ética, o segundo volta-se à sua dimensão humana e institucional: o uso consciente das tecnologias por quem decide. Nesse contexto, o magistrado ocupa posição central, pois é o agente de validação final da IA – aquele que deve compreender o funcionamento do sistema, interpretar criticamente suas recomendações e, acima de tudo, fundamentar juridicamente suas decisões.

O Protocolo Ético para Magistrados e Demais Usuários dos Sistemas Inteligentes (Apêndice B) tem como objetivo garantir que o uso da IA no Judiciário preserve os princípios constitucionais da igualdade, da fundamentação, do contraditório e da dignidade da pessoa humana, além de promover a incorporação do julgamento com perspectiva de gênero.

6.2.1 Compreensão e consentimento informado

Nenhum magistrado deve utilizar uma ferramenta de IA sem conhecer minimamente suas finalidades, seus limites e sua lógica decisória. O protocolo propõe que todo sistema empregado apresente documentação simplificada, em linguagem jurídica clara, sobre sua arquitetura e variáveis relevantes. O magistrado, por sua vez, deve declarar nos autos que teve

ciência de tais informações e que a decisão permanece fruto de sua deliberação racional e humana.

6.2.2 Uso responsável e transparente

A transparência é condição da legitimidade judicial. O uso da IA deve ser explicitamente mencionado na fundamentação da decisão, indicando como o parecer automatizado influenciou (ou não) o entendimento do julgador. Os relatórios gerados devem ser anexados aos autos e disponibilizados às partes, permitindo sua contestação. Assim, assegura-se que a tecnologia não opere em opacidade, mas dentro da lógica do devido processo legal.

6.2.3 Avaliação de equidade e impacto

O magistrado e sua equipe devem observar possíveis discrepâncias de tratamento entre grupos sociais, analisando se as recomendações algorítmicas tendem a favorecer ou penalizar mais determinados perfis (por exemplo, réus homens em detrimento de mulheres). Se tais padrões forem identificados, o caso deve ser comunicado ao comitê técnico de IA do tribunal, que deverá reavaliar o modelo e seus critérios.

Além disso, o tribunal deve manter indicadores públicos de equidade algorítmica, permitindo a comparação de desempenho entre grupos de gênero, raça e origem regional, conforme boas práticas internacionais.

6.2.4 Julgamento com perspectiva de gênero

Essa é a seção que materializa a garantia de equidade discutida no Capítulo 5. O protocolo recomenda que magistrados e servidores adotem, em todos os casos, uma leitura sensível às desigualdades estruturais entre homens e mulheres, especialmente em processos de violência doméstica, guarda, família ou execução penal.

A aplicação da IA não pode reproduzir ou reforçar estereótipos. O magistrado deve, portanto, interrogar criticamente os resultados apresentados, verificando se o modelo foi treinado com bases que contemplem decisões de diferentes gêneros e contextos.

Nos casos sensíveis, recomenda-se a revisão humana obrigatória por equipes especializadas ou por comissões de gênero do tribunal. A linguagem empregada nos relatórios algorítmicos também deve ser objeto de atenção: expressões discriminatórias, moralistas ou que associem condutas a papéis de gênero precisam ser identificadas e corrigidas.

6.2.5 Supervisão, capacitação e educação ética

Por fim, o protocolo prevê a institucionalização de comitês de ética em IA com participação feminina e interdisciplinar, além da oferta contínua de capacitações em letramento digital e justiça algorítmica. A formação ética é parte fundamental da governança tecnológica: sem compreender o funcionamento da IA, o magistrado perde sua capacidade crítica e, por consequência, sua autonomia.

CONSIDERAÇÕES FINAIS

A presente tese teve como objetivo investigar de que modo a IA, quando aplicada ao Poder Judiciário brasileiro, pode reproduzir ou intensificar desigualdades de gênero, e em que medida o Direito pode atuar para prevenir esses vieses e reorientar a tecnologia para promover a ética e a justiça social. A hipótese formulada – de que o Direito pode atuar como instrumento regulatório eficaz para prevenir e mitigar a reprodução de vieses de gênero pela inteligência artificial aplicada ao Poder Judiciário brasileiro, desde que articule simultaneamente garantias técnicas, garantias regulatórias e garantias políticas, assegurando que a tecnologia seja efetivamente reorientada para promover a ética e a justiça social – foi confirmada ao longo da pesquisa.

A análise teórica e documental mostrou que a tecnologia, ao ser incorporada ao sistema de justiça, reflete as assimetrias históricas de poder e gênero que marcam a sociedade brasileira. Nesse contexto, a IA não é neutra: ela reproduz os valores de quem a concebe, os dados que a alimentam e as instituições que a operam. Por outro lado, o estudo também evidenciou que o Direito, se aliado à técnica e orientado por princípios de equidade e direitos humanos, pode reconfigurar o uso da IA como ferramenta de transformação social, e não de reprodução de desigualdades.

A principal contribuição teórica desta tese reside na integração entre tecnofeminismo, governança algorítmica e teorias críticas do Direito, resultando em um marco conceitual que amplia a compreensão da IA enquanto fenômeno jurídico e social. Essa abordagem permitiu articular as dimensões ética, política e institucional da tecnologia, destacando o papel do Judiciário não apenas como aplicador da lei, mas como agente de regulação, fiscalização e exemplo normativo na construção de uma justiça digital inclusiva.

Metodologicamente, a pesquisa adotou uma abordagem interdisciplinar, combinando análise normativa, estudo comparado e revisão crítica de literatura nacional e internacional, o que possibilitou observar a convergência entre os debates sobre ética em IA e as agendas de direitos humanos e igualdade de gênero.

Do ponto de vista prático, foram formuladas diretrizes concretas para orientar o desenvolvimento e o uso ético da IA no Judiciário, como a incorporação obrigatória de recortes interseccionais em avaliações de impacto algorítmico, a criação de políticas de letramento digital com perspectiva de gênero e a promoção de uma governança participativa da tecnologia. Tais diretrizes dialogam com marcos institucionais recentes, como a Resolução n. 615/2025 do

CNJ (Conselho Nacional de Justiça, 2025c), e reforçam a necessidade de aprimorar seus dispositivos a fim de incluir salvaguardas específicas contra o viés de gênero. Além disso, o estudo evidenciou a urgência de que o “Protocolo para Julgamento com Perspectiva de Gênero” (Conselho Nacional de Justiça, 2021a) seja incorporado como referência ética e metodológica também no desenvolvimento de sistemas algorítmicos judiciais, garantindo que a automação não comprometa os avanços conquistados no campo dos direitos das mulheres.

As garantias técnicas, regulatórias e políticas analisadas no Capítulo 5 constituem o eixo normativo e prático que viabiliza a construção de uma IA ética no Poder Judiciário. As garantias técnicas – como *fairness*, acurácia, robusteza, explicabilidade, confiabilidade e governança de dados – operam como mecanismos de prevenção dos vieses de gênero desde a concepção e o treinamento dos modelos algorítmicos. Elas asseguram que os sistemas sejam tecnicamente sólidos, auditáveis e baseados em dados representativos, impedindo que distorções estatísticas ou sub-representações de grupos femininos sejam convertidas em decisões discriminatórias. A explicabilidade, em especial, traduz a ética em transparência operacional, permitindo que magistrados, partes e a sociedade compreendam o raciocínio algorítmico e exerçam controle democrático sobre ele.

As garantias regulatórias, por sua vez, estruturam o campo normativo e institucional dessa ética aplicada, transformando os princípios técnicos em obrigações jurídicas. A incorporação dos direitos humanos, o devido processo legal e a Resolução CNJ n. 615/2025 (Conselho Nacional de Justiça, 2025c) conformam um marco regulatório que obriga o Poder Judiciário a fiscalizar a equidade, a acurácia, a robustez e a não discriminação dos sistemas inteligentes. Essa regulação positiva garante que a governança algorítmica não seja apenas uma prática administrativa, mas um dever constitucional de transparência, motivação e respeito à igualdade. O “Protocolo para Julgamento com Perspectiva de Gênero” (Conselho Nacional de Justiça, 2021a), ao ser incorporado como referência de desenvolvimento e auditoria, traduz a ética feminista em diretriz regulatória, conectando o desempenho técnico ao compromisso com a equidade substantiva.

Por fim, as garantias políticas completam o tripé ético da IA judicial ao enfatizar a dimensão humana e institucional do controle tecnológico. Conscientização, letramento digital, bem-estar socioambiental e vigilância permanente sobre a desigualdade de gênero são condições políticas indispensáveis para que a ética se converta em cultura institucional. Essas garantias asseguram que o uso da IA seja acompanhado de senso crítico, transparência pública e corresponsabilidade social, evitando que a automação se transforme em instrumento de

desumanização. A ética, aqui, não é um apêndice do desenvolvimento tecnológico, mas seu fundamento político e epistêmico.

Como desdobramento dessas diretrizes, esta tese apresentou dois protocolos éticos – um voltado à dimensão técnica e outro à dimensão humana – como instrumentos complementares de implementação das garantias éticas discutidas no Capítulo 5. Juntos, eles delineiam um modelo de governança algorítmica do Poder Judiciário, no qual a tecnologia é compreendida não como substituta do humano, mas como ferramenta sob seu controle crítico.

O protocolo para desenvolvedores responde à necessidade de prevenção e responsabilidade técnica, assegurando que os sistemas sejam concebidos e monitorados de modo ético, transparente e auditável, em conformidade com princípios de justiça algorítmica, explicabilidade e responsabilidade institucional. Já o protocolo para magistrados e usuários reforça a dimensão interpretativa, a fundamentação das decisões, o julgamento com perspectiva de gênero e a autonomia decisória humana – pilares que sustentam a legitimidade democrática da justiça. Ambos constituem, portanto, exemplos de operacionalização das garantias éticas que permitem que os princípios se tornem práticas verificáveis no cotidiano institucional.

Esses instrumentos, ao transformarem princípios em procedimentos, constituem uma ponte entre o discurso ético e a prática institucional, permitindo que o uso da IA no Poder Judiciário brasileiro avance sob bases sólidas, auditáveis e comprometidas com os valores fundamentais da igualdade, da justiça e dos direitos humanos. Com isso, a pesquisa contribui não apenas para o debate teórico, mas também para o desenvolvimento de ferramentas práticas de regulação e autoavaliação ética das tecnologias judiciais, aproximando o ideal normativo da realidade institucional.

As implicações políticas e institucionais desta pesquisa indicam que a ética da IA no Judiciário depende de condições estruturais de conscientização, capacitação e participação democrática. É imprescindível investir em formação continuada de magistrados, servidores e técnicos, promovendo o letramento digital e a compreensão crítica sobre as implicações da automação.

A adoção da IA deve ser acompanhada de instrumentos de transparência, auditoria, prestação de contas e acesso público à informação, de modo a fortalecer a confiança social no sistema de justiça. Da mesma forma, a sustentabilidade socioambiental e o bem-estar coletivo devem orientar o desenvolvimento tecnológico, assegurando que a inovação ocorra dentro dos limites ecológicos e éticos compatíveis com os direitos humanos e os Objetivos de Desenvolvimento Sustentável (ODS) da Agenda 2030 da ONU (Nações Unidas Brasil, [20-?]).

A pesquisa, contudo, reconhece limitações decorrentes do caráter ainda restrito e opaco dos sistemas de IA em operação no Judiciário brasileiro, bem como da falta de acesso direto a seus algoritmos e bases de dados. Além disso, a natureza qualitativa e exploratória do estudo impossibilita generalizações estatísticas, embora proporcione uma leitura crítica e interpretativa dos fenômenos analisados. Essas limitações abrem caminho para estudos futuros mais empíricos, voltados à observação direta de modelos algorítmicos e seus efeitos discriminatórios, bem como comparações entre diferentes sistemas judiciais nacionais e internacionais.

Recomenda-se, portanto, a ampliação do debate interdisciplinar entre Direito, Ciência da Computação, Filosofia, Sociologia e Estudos de Gênero, a criação de projetos-piloto de avaliação ética de IA com participação da sociedade civil e o fortalecimento do papel do CNJ na regulação e fiscalização das tecnologias judiciais. Também se faz necessária a realização de estudos comparados e avaliações longitudinais sobre os impactos da IA na equidade de gênero e na democratização do acesso à justiça.

Em síntese, esta tese conclui que a IA só será legítima no Judiciário se for também ética, inclusiva e democrática. A tecnologia, quando orientada por princípios de equidade, sustentabilidade e direitos humanos, pode contribuir para um Judiciário mais eficiente e justo; contudo, sem vigilância crítica e compromisso político, corre o risco de se tornar novo vetor de desigualdade e opacidade institucional. As garantias técnicas, regulatórias e políticas aqui discutidas demonstram que a ética da IA é inseparável de sua arquitetura: é na forma como o sistema é projetado, auditado e governado que se decide se ele servirá à reprodução da desigualdade ou à promoção da justiça. O desafio, portanto, não é apenas técnico, mas civilizatório: trata-se de construir uma IA a serviço da justiça social e da dignidade humana, capaz de fortalecer, e não enfraquecer, o ideal constitucional de igualdade.

REFERÊNCIAS

- ABDOLLAHI, B.; NASRAOUI, O. Transparency in fair machine learning: The case of explainable recommender systems. In: ZHOU, J.; CHEN, F. (Org.). **Human and machine learning: Visible, explainable, trustworthy and transparent**. Cham: Springer, p. 21–35, 2018. DOI: https://doi.org/10.1007/978-3-319-90403-0_2. Disponível em: https://link.springer.com/chapter/10.1007/978-3-319-90403-0_2. Acesso em: 12 fev. 2026.
- ACCESSNOW.ORG. Human Rights in the Age of Artificial Intelligence. **Accessnow.org**, 2018. Disponível em: <https://www.accessnow.org/wp-content/uploads/2018/11/AI-and-Human-Rights.pdf>. Acesso em: 21 maio 2024.
- AIVODJI, U.; ARAI, H.; FORTINEAU, O.; GAMBS, S.; HARA, S.; TAPP, A. Fairwashing: The risk of rationalization. In: **Proceedings of the 36th International Conference on Machine Learning**, 161-170, 2019. Disponível em: <https://proceedings.mlr.press/v97/aivodji19a.html>. Acesso em: 23 out. 2025.
- ALMEIDA, B. T. A. de. **Democracia e respeito à lei**. Rio de Janeiro: Lumen Juris, 2015.
- ALMEIDA, M. C. Mulheres, vieses implícitos e o judiciário. **Jota**, mar. 2018. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/mulheres-vieses-implicitos-e-o-judiciario-12032018>. Acesso em: 29 maio 2018.
- ALVES, E. M. **Inteligência Artificial e Direito: uma análise sobre os impactos de novas tecnologias e o uso da inteligência artificial no judiciário brasileiro**. Orientador: André Gomes de Sousa Alves. Monografia (Curso de Graduação em Direito). 47f. Universidade Federal de Campina Grande, 2020.
- ALVES, M. A. S.; ANDRADE, O. M. de. Da “caixa-preta” à “caixa de vidro”: o uso da Explainable Artificial Intelligence (XAI) para reduzir a opacidade e enfrentar o enviesamento em modelos algorítmicos. **Revista de Direito Público**, Brasília, v. 18, n. 100, p. 349-373, out./dez. 2021.
- AMODEI, D.; OLAH, C.; STEINHARDT, J.; CRISTIANO, P.; SCHULMAN, J.; MANÉ, D. Concrete problems in AI safety. **arXiv:1606.06565 [cs.AI]**, jul. 2016. Disponível em: <https://arxiv.org/pdf/1606.06565>. Acesso em: 23 out. 2025.
- ANANNY, M.; CRAWFORD, K. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. **New Media and Society**, v. 20, n. 3, p. 973-989, 2018. DOI: <https://doi.org/10.1177/1461444816676645>. Disponível em: <https://journals.sagepub.com/doi/10.1177/1461444816676645>. Acesso em: 12 fev. 2026.
- ANDRADE, V. R. P. de. A soberania patriarcal: o sistema de justiça criminal no tratamento da violência sexual contra a mulher. **Revista Direito Público**, Brasília, n. 17, p. 52-75, jul./set. 2007.
- ANDRIGHI, F. N.; BIANCHI, J. Fl. Reflexão sobre os riscos do uso da inteligência artificial ao processo de tomada de decisões no Poder Judiciário. In: PINTO, H. A.; GUEDES, J. C.; CÉSAR, J. P. de C. (Coord.). **Inteligência artificial aplicada ao processo de tomada de decisões**. Belo Horizonte; São Paulo: D’Plácido, 2020.

ANGWIN, J.; LARSON, J.; MATTU, S.; KIRCHNER, L. Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. **ProPublica**, maio 2016.

APPIO, E. **Discrecionalidade política do Poder Judiciário**. Curitiba: Juruá, 2008.

AQUINO, E. M. L. Gênero e ciência no Brasil: contribuições para pensar a ação política na busca da equidade. In: BRASIL. Secretaria Especial de Políticas para as Mulheres. **Encontro nacional pensando gênero e ciência: núcleos e grupos de pesquisas**. Brasília, 2006.

ARAÚJO, H. B. **Qual o gênero do Supremo?** Diálogo institucional nas sabatinas para o STF, poder e profissionalismo. 2015. Trabalho de Conclusão de Curso (Graduação em Direito) – Escola de Formação da Sociedade Brasileira de Direito Público, São Paulo, 2015.

BADARÓ, S.; IBÁÑEZ, L. J.; AGÜERO, M. J. Sistemas expertos: fundamentos, metodologías y aplicaciones. **Ciencia y Tecnología**, v. 13, p. 349-364, 2013.

BADIA, R. L. de M.; GONZÁLEZ, P. M. **Inteligencia artificial**. Madrid: CSIC, 2017.

BADMAN, A.; KOSINSKI, M. O que é big data? **IBM**, 18 nov. 2024. Disponível em: <https://www.ibm.com/br-pt/think/topics/big-data>. Acesso em 11 fev. 2025.

BARBOSA, D. F. N.; SANTOS, T. F. de M.; VILAR, E. C. N. Inteligência artificial no judiciário brasileiro: rumo a uma justiça mais eficiente e acessível. **Revista Ibero-Americana de Humanidades, Ciências e Educação**, São Paulo, v. 9, n. 11, p. 2272–2296, nov. 2023. Disponível em: <https://periodicorease.pro.br/rease/article/view/12540>. Acesso em: 12 fev. 2026.

BARDIN, L. **Análise de conteúdo**. São Paulo: Edições 70, 2016.

BARFIELD, W. Liability for autonomous and artificially intelligent robots. **Paladyn**, v. 9, n. 1, p. 193–203, 2018.

BAROCAS, S.; SELBST, A. D.; RAGHAVAN, M. The hidden assumptions behind counterfactual explanations and principal reasons. In: **Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency**, p. 80-89, 2020. DOI: <https://doi.org/10.1145/3351095.3372830>. Disponível em: <https://dl.acm.org/doi/10.1145/3351095.3372830>. Acesso em: 12 fev. 2026.

BARON, J. **Thinking and Deciding**. 4. ed. Cambridge: Cambridge University Press, 2008.

BASS, D.; HUET, E. Researchers combat gender and racial bias in artificial intelligence. **Bloomberg**, pp. 1–9, dez. 2017. Disponível em: <https://www.bloomberg.com/news/articles/2017-12-04/researchers-combat-gender-and-racial-bias-in-artificial-intelligence>. Acesso em: 18 out. 2024.

BAUMAN, Z. **Modernidade líquida**. Trad. Plínio Dentzien. Rio de Janeiro: Zahar, 2021.

BAZERMAN, M. H.; MOORE, D. A. **Judgment in Managerial Decision Making**. 8a. ed. New Jersey: Wiley, 2013.

BELLMAN, R. **An Introduction to Artificial Intelligence: Can Computers Think?** San Francisco: Boyd & Fraser Pub. Co., c1978.

BELLO, A.; BLOWERS, T.; SCHNEEGANS, S.; STRAZA, T. To be smart, the digital revolution will need to be inclusive. In: UNESCO. **UNESCO Science Report**. Paris, 2021.

BERTONI, F. F.; AMARAL, M. E. A. Neurociência e decisão judicial: a influência das heurísticas e dos vieses. In: **9 Congresso Internacional de Ciências Criminais**, Porto Alegre, 2018.

BHATNAGAR, A.; GAJJAR, D. How is artificial intelligence affecting society? **UK Parliament**, maio 2024. DOI: <https://doi.org/10.58248/RR15>. Disponível em: <https://post.parliament.uk/how-is-artificial-intelligence-affecting-society/>. Acesso em: 12 fev. 2026.

BINNS, R. Fairness in Machine Learning: Lessons from Political Philosophy. **Journal of Machine Learning Research**, v. 81, p. 1-11, 2018. Disponível em: <https://export.arxiv.org/pdf/1712.03586v2>. Acesso em: 23 out. 2025.

BIONDI, P. P. P. **Inteligência artificial, ética e direito: desafios contemporâneos**. São Paulo: Thomson Reuters, 2020.

BOBBIO, N. **A era dos direitos**. Rio de Janeiro: Campus, 1992.

BOEING, D. H. A. **Ensinando um robô a julgar: pragmática, discricionariedade e vieses no uso de aprendizado de máquina no judiciário**. Orientador: Alexandre Moraes da Rosa. Trabalho de Conclusão de Curso (Graduação em Direito). Universidade Federal de Santa Catarina, 85 p., 2019.

BONELLI, M. da G. Profissionalismo, gênero e significados da diferença entre juízes e juízas estaduais e federais. **Contemporânea**, n. 1, p. 103-123, jan./jun. 2011.

BONAT, D.; VALE, L. M. B. do; PEREIRA, J. S. dos S. S. Inteligência artificial generativa e a fundamentação da decisão judicial. **Revista de Processo**, [s. l.], v. 346, p. 349-370, dez. 2023.

BOURDIEU, P. **A dominação masculina**. Trad. Maria Helena Kühner. 2a ed. Rio de Janeiro: Bertrand Brasil, 2002.

BRANDÃO, L. Inteligência artificial e discriminação contra mulheres: os dados e o sistema. **Iris – Instituto de referência em internet e sociedade**, 8 mar. 2021. Disponível em: <https://irisbh.com.br/inteligencia-artificial-e-discriminacao-contra-mulheres-os-dados-e-o-sistema/>. Acesso em: 21 nov. 2024.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 17 set. 2022.

BRASIL. **Decreto n. 678, de 6 de novembro de 1992**. Promulga a Convenção Americana sobre Direitos Humanos (Pacto de São José da Costa Rica), de 22 de novembro de 1969.

Disponível em: https://www.planalto.gov.br/ccivil_03/decreto/d0678.htm. Acesso em: 1 nov. 2025.

BRASIL. **Lei n. 13.105, de 16 de março de 2015**. Institui o Código de Processo Civil. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13105.htm. Acesso em: 1 nov. 2025.

BRASIL. **Lei n. 13.709, de 14 de agosto de 2018a**. Dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado; altera a Lei n.º 12.965, de 23 de abril de 2014 (Marco Civil da Internet); e dá outras providências. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 1 nov. 2025.

BRASIL. **Lei n. 6.015, de 31 de dezembro de 1973**. Dispõe sobre os registros públicos e dá outras providências. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l6015compilada.htm. Acesso em: 1 nov. 2025.

BRASIL. Senado Federal. **Projeto de Lei n. 2338, de 2023a**. Estabelece princípios, direitos e deveres para o desenvolvimento e a aplicação da inteligência artificial no Brasil. Autor: Senador Rodrigo Pacheco (PSD/MG). Brasília, DF, 2023. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>. Acesso em: 24 out. 2025.

BRASIL. Superior Tribunal de Justiça. Terceira Seção. Recurso Especial n. 1.675.874/MS. Ementa. Recurso Especial. Recurso submetido ao rito dos Repetitivos (art. 1.036 do CPC, c/c o art. 256, I, do RISTJ). Violência doméstica e familiar contra a mulher. Danos morais. Indenização mínima. Art. 397, IV, do CPP. Pedido necessário. Produção de prova específica dispensável. Dano in re ipsa. Fixação consoante prudente arbítrio do juízo. Recurso Especial provido. 1. O Superior Tribunal de Justiça - sob a influência dos princípios da dignidade da pessoa humana (CF, art. 1º, III), da igualdade (CF, art. 5º, I) e da vedação a qualquer discriminação atentatória dos direitos e das liberdades fundamentais (CF, art. 5º, XLI), e em razão da determinação de que "O Estado assegurará a assistência à família na pessoa de cada um dos que a integram, criando mecanismos para coibir a violência no âmbito de suas relações" (art. 226, § 8º) - tem avançado na maximização dos princípios e das regras do novo subsistema jurídico introduzido em nosso ordenamento com a Lei nº 11.340/2006, vencendo a timidez hermenêutica no reproche à violência doméstica e familiar contra a mulher, como deixam claro os verbetes sumulares n. 542, 588, 589 e 600. 2. Refutar, com veemência, a violência contra as mulheres implica defender sua liberdade (para amar, pensar, trabalhar, se expressar), criar mecanismos para seu fortalecimento, ampliar o raio de sua proteção jurídica e otimizar todos os instrumentos normativos que de algum modo compensem ou atenuem o sofrimento e os malefícios causados pela violência sofrida na condição de mulher. 3. A evolução legislativa ocorrida na última década em nosso sistema jurídico evidencia uma tendência, também verificada em âmbito internacional, a uma maior valorização e legitimação da vítima, particularmente a mulher, no processo penal. 4. Entre diversas outras inovações introduzidas no Código de Processo Penal com a reforma de 2008, nomeadamente com a Lei n. 11.719/2008, destaca-se a inclusão do inciso IV ao art. 387, que, consoante pacífica jurisprudência desta Corte Superior, contempla a viabilidade de indenização para as duas espécies de dano - o material e o moral -, desde que tenha havido a dedução de seu pedido na denúncia ou na queixa. 5. Mais robusta ainda há de ser tal compreensão quando se cuida de danos morais experimentados pela mulher vítima de violência doméstica. Em tal situação,

emerge a inarredável compreensão de que a fixação, na sentença condenatória, de indenização, a título de danos morais, para a vítima de violência doméstica, independe de indicação de um valor líquido e certo pelo postulante da reparação de danos, podendo o quantum ser fixado minimamente pelo Juiz sentenciante, de acordo com seu prudente arbítrio. 6. No âmbito da reparação dos danos morais - visto que, por óbvio, os danos materiais dependem de comprovação do prejuízo, como sói ocorrer em ações de similar natureza -, a Lei Maria da Penha, complementada pela reforma do Código de Processo Penal já mencionada, passou a permitir que o juízo único - o criminal - possa decidir sobre um montante que, relacionado à dor, ao sofrimento, à humilhação da vítima, de difícil mensuração, deriva da própria prática criminosa experimentada. 7. Não se mostra razoável, a esse fim, a exigência de instrução probatória acerca do dano psíquico, do grau de humilhação, da diminuição da autoestima etc., se a própria conduta criminosa empregada pelo agressor já está imbuída de desonra, descrédito e menosprezo à dignidade e ao valor da mulher como pessoa. 8. Também justifica a não exigência de produção de prova dos danos morais sofridos com a violência doméstica a necessidade de melhor concretizar, com o suporte processual já existente, o atendimento integral à mulher em situação de violência doméstica, de sorte a reduzir sua revitimização e as possibilidades de violência institucional, consubstanciadas em sucessivas oitivas e pleitos perante juízos diversos. 9. O que se há de exigir como prova, mediante o respeito ao devido processo penal, de que são expressão o contraditório e a ampla defesa, é a própria imputação criminosa - sob a regra, derivada da presunção de inocência, de que o onus probandi é integralmente do órgão de acusação -, porque, uma vez demonstrada a agressão à mulher, os danos psíquicos dela derivados são evidentes e nem têm mesmo como ser demonstrados. 10. Recurso especial provido para restabelecer a indenização mínima fixada em favor pelo Juízo de primeiro grau, a título de danos morais à vítima da violência doméstica. TESE: Nos casos de violência contra a mulher praticados no âmbito doméstico e familiar, é possível a fixação de valor mínimo indenizatório a título de dano moral, desde que haja pedido expresso da acusação ou da parte ofendida, ainda que não especificada a quantia, e independentemente de instrução probatória. (REsp n. 1.675.874/MS, relator Ministro Rogério Schietti Cruz, Terceira Seção, julgado em 28/2/2018, DJe de 8/3/2018.) Relator Ministro Rogério Schietti Cruz. Brasília, DF, 28 de fevereiro de 2018. Acórdão publicado em 08 de março de 2018b.

BRASIL. Superior Tribunal de Justiça. Terceira Seção. Recurso Especial n. 1.986.672/SC. Ementa. Penal. Recurso especial. Crime de estelionato. Fixação de valor indenizatório mínimo. Inclusão do nome da vítima em cadastros de inadimplentes. Dano moral in re ipsa. Desnecessidade de instrução probatória específica, no caso concreto. Exigência, porém, de pedido expresso e valor indicado na denúncia. Ausência de indicação, na peça acusatória, da quantia pretendida para a compensação da vítima. Recurso especial a que se dá provimento, para excluir a fixação do valor indenizatório mínimo. 1. A liquidação parcial do dano (material ou moral) na sentença condenatória, referida pelo art. 387, IV, do CPP, exige o atendimento a três requisitos cumulativos: (I) o pedido expresso na inicial; (II) a indicação do montante pretendido; e (III) a realização de instrução específica a fim de viabilizar ao réu o exercício da ampla defesa e do contraditório. Precedentes desta Quinta Turma. 2. A Quinta Turma, no julgamento do AgRg no REsp 2.029.732/MS em 22/8/ 2023, todavia, adotou interpretação idêntica à da Sexta Turma, no sentido de que é necessário incluir o pedido referente ao valor mínimo para reparação do dano moral na exordial acusatória, com a dispensa de instrução probatória específica. Esse julgamento não tratou da obrigatoriedade, na denúncia, de indicar o valor a ser determinado pelo juiz criminal. Porém, a conclusão foi a de que a indicação do valor pretendido é dispensável, seguindo a jurisprudência consolidada da Sexta Turma. 3. O dano moral decorrente do crime de estelionato que resultou na inclusão do

nome da vítima em cadastro de inadimplentes é presumido. Inteligência da Súmula 385/STJ. 4. Com efeito, a possibilidade de presunção do dano moral *in re ipsa*, à luz das específicas circunstâncias do caso concreto, dispensa a obrigatoriedade de instrução específica sobre o dano. No entanto, não afasta a exigência de formulação do pedido na denúncia, com indicação do montante pretendido. 5. A falta de uma indicação clara do valor mínimo necessário para a reparação do dano almejado viola o princípio do contraditório e o próprio sistema acusatório, por na prática exigir que o juiz defina ele próprio um valor, sem indicação das partes. Destarte, uma medida simples e eficaz consiste na inclusão do pedido na petição inicial acusatória, juntamente com a exigência de especificar o valor pretendido desde o momento da apresentação da denúncia ou queixa-crime. Essa abordagem reflete a tendência de aprimoramento do contraditório, tornando imperativa a sua inclusão no âmbito da denúncia. 6. Assim, a fixação de valor indenizatório mínimo por danos morais, nos termos do art. 387, IV, do CPP, exige que haja pedido expresso da acusação ou da parte ofendida, com a indicação do valor pretendido, nos termos do art. 3º do CPP c/c o art. 292, V, do CPC/2015. 7. Na peça acusatória (apresentada já na vigência do CPC/2015), apesar de haver o pedido expresso do valor mínimo para reparar o dano, não se encontra indicado o valor atribuído à reparação da vítima. Diante disso, considerando a violação do princípio da congruência, dos princípios constitucionais do contraditório e da ampla defesa e do sistema acusatório, deve-se excluir o valor mínimo de indenização por danos morais fixado. 8. O entendimento aqui firmado não se aplica aos casos de violência doméstica e familiar contra a mulher, que continuam regidos pela tese fixada no julgamento do tema repetitivo 983/STJ. 9. Recurso especial provido para excluir a fixação do valor indenizatório mínimo. (REsp n. 1.986.672/SC, relator Ministro Ribeiro Dantas, Terceira Seção, julgado em 8/11/2023, DJe de 21/11/2023.) Relator Ministro Ribeiro Dantas. Brasília, DF, 08 de novembro de 2023. Acórdão publicado em 21 de novembro de 2023b.

BRASIL. Supremo Tribunal Federal. Ação Direta de Inconstitucionalidade n. 6387. Ementa. Medida Cautelar em Ação Direta de Inconstitucionalidade. Referendo. Medida Provisória Nº 954/2020. Emergência de Saúde Pública de Importância Internacional Decorrente do Novo Coronavírus (covid-19). Compartilhamento de Dados dos usuários do serviço telefônico fixo comutado e do serviço móvel pessoal, pelas empresas prestadoras, com o Instituto Brasileiro de Geografia e Estatística. Fumus Boni Juris. Periculum in Mora. Deferimento. Recorrente: Conselho Federal da Ordem dos Advogados do Brasil - CFOAB. Intimado: Presidente da República. Relator: min. Rosa Weber. Julgamento em 6 e 7.5.2020. DJE nº 137, divulgado em 02/06/2020.

BRAY, F. Gender and Technology. **The Annual Review of Anthropology**, vol. 36, pp. 37-53, out. 2007.

BREAZEL, C. Toward sociable robots. **Robotics and Autonomous Systems**, v. 42 n. 3-4, pp. 167-175, 2003.

BRITTO, M. C. de S.; CRUZ, F. B. da. O espaço da mulher na ciência e na tecnologia: ciência e tecnologia feminista. **Coisas do Gênero**, v. 9, n. 1, p. 335-356, jan./jun. 2023.

BRITTO, M. C. de S.; PREUSS, L. T.; CRUZ, F. B. da. Políticas sociais de inclusão e cidadania digital: o Programa de Extensão Universidade Aberta para a Terceira Idade. **Revista Conexão UEPG**, Ponta Grossa, Paraná - Brasil. v. 19, e2322018, p. 01-15, 2023.

BRUNETTA, C. M.; FERRAZ, T. S.; ALENCAR, A. C. de. Entre alucinações e inovações: o desafio da implementação responsável da IA Generativa no Sistema de Justiça. **Revista**

Opinião Jurídica (Fortaleza), Fortaleza, v. 23, n. 44, p. 1–26, 2025. DOI: 10.12662/2447-6641oj.v23i44.p1-26.2025. Disponível em: <https://periodicos.unichristus.edu.br/opiniaojuridica/article/view/5913>. Acesso em: 18 dez. 2025.

BRUSTOLIN, A.; SAMPAIO, S. C. C.; CASTILHO, E. W. V. de. “Eu acreditava que ele ia mudar”: de vítima de violência doméstica a autora de homicídio por omissão imprópria. In: SEVERI, F. C. (Org.). **Reescrivendo decisões judiciais em perspectivas feministas**. Ribeirão Preto: Universidade de São Paulo, Núcleo de Pesquisa e Extensão “O Direito Achado na Rua”, 2022.

BUOLAMWINI, J.; GEBRU, T. Gender shades: Intersectional accuracy disparities in commercial gender classification. **Proceedings of the 1st Conference on Fairness, Accountability and Transparency**. Proceedings of Machine Learning Research, vol. 81, pp. 77-91, 2018.

BUTLER, J. **Gender Trouble: Feminism and the Subversion of Identity** (Notations). New York: Routledge, 1990.

CABALLERO TRENADO, L. Reseña: BARRIO, A. M. (Dir.). Derecho de los robots. **Revista de Derecho Privado**, v. 37, p. 367-371, jun. 2019. <https://dx.doi.org/10.18601/01234366.n37.15>.

CABALLERO, B. I.; MEIER, I. F.; LOPES, J. G.; JUNQUILHO, T. A.; MENEZES DE AZEVÊDO, U. **Relatório de pesquisa: mapeando riscos da IA no Poder Judiciário brasileiro**. Org. Tainá Aguiar Junquilha e Laura Schertel Ferreira Mendes. Revisão: Ian Ferrare Meier e Tainá Aguiar Junquilha. Brasília: Laboratório de Governança e Regulação de Inteligência Artificial (LIA) do Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP), 2024.

CADEMARTORI, S. U.; GOMES, N. C. A teoria da interpretação jurídica de Kelsen: uma crítica a partir da obra de Friedrich Muller. **Sequência**, n. 57, p. 95-114, dez. 2008.

CALISKAN, A.; BRYSON, J.J.; NARAYANAN, A. Semantics derived automatically from language corpora contain human-like biases. **Science**, v. 356, n. 6334, pp. 183–186, 2017. <https://doi.org/10.1126/science.aal4230>

CAMBRIDGE DICTIONARY. Bias. Disponível em: https://dictionary.cambridge.org/dictionary/english-portuguese/bias#google_vignette. Acesso em: 21 maio 2024.

CAPPELLETTI, M.; GARTH, B. **Acesso à justiça**. Tradução de Ellen Gracie Northfleet. Porto Alegre: Sérgio Antonio Fabris Editor, 1988.

CÁRDENAS, Erick Rincón; MOLANO, Valeria Martinez. Un estudio sobre la posibilidad de aplicar la inteligencia artificial en las decisiones judiciales. **Revista Direito FGV**, v. 17, n. 1, e2101, 2021.

CARTA DE BRASÍLIA PELA IGUALDADE DE GÊNERO NO PODER JUDICIÁRIO DE 2023. 2ª ed. Escola de Formação e Aperfeiçoamento de Magistrados; Superior Tribunal de Justiça; Conselho Nacional de Justiça, 2023. Disponível em: <https://www.cnj.jus.br/wp->

content/uploads/2023/12/carta-de-brasilia-pela-igualdade-de-genero-no-poder-judiciario-de-2023.pdf. Acesso em: 05 nov. 2024.

CARTA DE BRASÍLIA PELA IGUALDADE DE GÊNERO NO PODER JUDICIÁRIO. Conselho Nacional de Justiça; Escola de Formação e Aperfeiçoamento de Magistrados; Associação Nacional dos Magistrados da Justiça do Trabalho; Associação dos Juizes Federais do Brasil; Associação dos Magistrados Brasileiros, 2022. Disponível em: <https://www.cnj.jus.br/wp-content/uploads/2022/10/carta-de-brasilia-mulheres-na-justica-3-3-2023.pdf>. Acesso em: 05 nov. 2024.

CARVALHO, A. M. G. de; AMÉRICO, M. T. Inclusão e Cidadania Digital no Brasil: a (des) articulação das políticas públicas. **Redes.com**, n. 9, 2014. Disponível em: <https://repositorio.unesp.br/bitstream/handle/11449/135513/ISSN2255-5919-2014-01-69-84.pdf?sequence=1&isAllowed=y>. Acesso em: 30 nov. 2022.

CARVALHO, L. B. de; BIONDI, P. P. P. **Mulheres no sistema de justiça:** representatividade e desafios. Brasília: CNJ, 2021.

CASTELLS, M. **A sociedade em rede**. Trad. Roneide Venancio Majer. Vol. 1 (A era da informação: economia, sociedade e cultura). 6a ed. São Paulo: Paz e Terra, 2011.

CEPEJ-AIAB – EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE, ADVISORY BOARD ON ARTIFICIAL INTELLIGENCE. **1st AIAB Report on the Use of Artificial Intelligence (AI) In the Judiciary Based on the Information Contained in the Resource Centre on Cyberjustice and AI**. Strasburg, 2025. Disponível em: <https://rm.coe.int/cepej-aiab-2024-4rev5-en-first-aiab-report-2788-0938-9324-v-1/1680b49def>. Acesso em: 27 jul. 2025.

CEPEJ; COUNCIL OF EUROPE. **Carta Europeia de Ética para o Uso da Inteligência Artificial em Sistemas Judiciais e seu Ambiente**. Estrasburgo, 2018. Disponível em: <https://rm.coe.int/carta-etica-traduzida-para-portugues-revista/168093b7e0>. Acesso em: 28 nov. 2025.

CERNADAS, E.; CALVO-IGLESIAS, E. Gender perspective in Artificial Intelligence (AI). In: **TEEM'20 Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality**, Salamanca, pp. 173-176, jan. 2021. Disponível em: Acesso em: 13 fev. 2025.

CHAN, I. (Ed.). Por que o machismo cria barreiras para as mulheres na tecnologia. **Progamaria**, [20–]. Disponível em: <https://www.programaria.org/especiais/mulheres-tecnologia/#:~:text=A%20associa%C3%A7%C3%A3o%20entre%20tecnologia%20e,em%20si%E2%80%9D%2C%20finaliza%20B%C3%A1rbara>. Acesso em: 13 fev. 2025.

CHOI, M. A Concept Analysis of Digital Citizenship for Democratic Citizenship Education in the Internet Age. **Theory & Research in Social Education**, pp. 1-43, 2016. Disponível em: <http://global-awareness.org/resources/misc/Choi2016ConceptAnalysisDigitalCitizenship.pdf>. Acesso em: 25 out. 2025.

CITRON, D. K.; PASQUALE, F. The Scored Society: Due Process for Automated Predictions. **Washington Law Review**, v. 89, 2014.

CLÈVE, C. M. Poder Judiciário: autonomia e justiça. **R. Inf. Legisl. Brasília**, a. 30, n. 117, jan./mar. 1993.

COCKBURN, C. **Women's Movements Against War: An International Perspective**. 2006. Disponível em: <http://www.cynthiacockburn.org/BlogwomenVwarInternational.pdf>. Acesso em: 09 set. 2022.

COMISSÃO AJUFE MULHERES. **Nota Técnica: AJUFE mulheres 01/2017: resultados da pesquisa para se conhecer o perfil das associadas da AJUFE**. Brasília: AJUFE Mulheres, 2017.

COMISSÃO EUROPEIA. **Comunicação da Comissão: Inteligência artificial para a Europa**, de 25 de abril de 2018. Bruxelas, 2018. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:52018DC0237>. Acesso em: 18 abr. 2025.

COMISSÃO EUROPEIA. **Shaping Europe's digital future: Draft Ethics guidelines for trustworthy AI**. Comissão Europeia, 2018. Disponível em: <https://digital-strategy.ec.europa.eu/pt/node/2236>. Acesso em: 18 set. 2024.

CONNELL, R.; PEARSE, R. **Gênero: uma perspectiva global**. Moschkovich, Marília. São Paulo: nVersos, 2015.

CONSELHO NACIONAL DE JUSTIÇA. Justiça 4.0. **CNJ - Conselho Nacional de Justiça**, [20-?]. Disponível em: <https://www.cnj.jus.br/tecnologia-da-informacao-e-comunicacao/justica-4-0/>. Acesso em: 25 out. 2025.

CONSELHO NACIONAL DE JUSTIÇA. **Justiça em números 2024**. Conselho Nacional de Justiça. Brasília: CNJ, 2024a. Disponível em: <https://www.cnj.jus.br/wp-content/uploads/2024/05/justica-em-numeros-2024.pdf>. Acesso em: 04 ago. 2025.

CONSELHO NACIONAL DE JUSTIÇA. **Justiça em números 2025**. Conselho Nacional de Justiça. Brasília: CNJ, 2025a. Disponível em: <https://www.cnj.jus.br/wp-content/uploads/2025/10/justica-em-numeros-2025.pdf>. Acesso em: 20 out. 2025.

CONSELHO NACIONAL DE JUSTIÇA. Módulo de Produtividade Mensal. **CNJ**, [20-?]. Disponível em: <https://www.cnj.jus.br/pesquisas-judiciarias/modulo-de-produtividade-mensal/>. Acesso em 20 out. 2025.

CONSELHO NACIONAL DE JUSTIÇA. **O uso da inteligência artificial generativa no Poder Judiciário brasileiro: relatório de pesquisa**. Brasília: CNJ, 2024b. Disponível em: <https://www.cnj.jus.br/wp-content/uploads/2024/09/cnj-relatorio-de-pesquisa-iag-pj.pdf>. Acesso em: 30 jul. 2025.

CONSELHO NACIONAL DE JUSTIÇA. Painel de Projetos de Inteligência Artificial em 2024, **CNJ**, 22 abr. 2025b. Disponível em: <https://paineisanalytics.cnj.jus.br/single/?appid=51977be5-96d0-4362-98ff-ed3eb3337781&sheet=maXvpqE&theme=horizon&lang=pt-BR&opt=ctxmenu,currsel&select=Ramo,&select=Tribunal,&select=projetoIA2024Tribunal&select=nomeProjeto>. Acesso em: 28 nov. 2025.

CONSELHO NACIONAL DE JUSTIÇA. **Participação Feminina na Magistratura:** Atualizações 2023. Brasília: Conselho Nacional de Justiça, 2023a.

CONSELHO NACIONAL DE JUSTIÇA. **Portaria n. 271 de 04/12/2020.** Regulamenta o uso de Inteligência Artificial no âmbito do Poder Judiciário. DJe/CNJ nº 389/2020, de 9/12/2020, 2020a. E republicada no DJe/CNJ nº 393/2020, de 14/12/2020.

CONSELHO NACIONAL DE JUSTIÇA. **Protocolo para julgamento com perspectiva de gênero.** Brasília: Conselho Nacional de Justiça; Escola Nacional de Formação e Aperfeiçoamento de Magistrados, 2021a.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 255 de 04/09/2018.** Institui a Política Nacional de Incentivo à Participação Institucional Feminina no Poder Judiciário, 2018a. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/2670>. Acesso em: 04 set. 2024.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 270 de 11/12/2018.** Dispõe sobre o uso do nome social pelas pessoas trans, travestis e transexuais usuárias dos serviços judiciários, membros, servidores, estagiários e trabalhadores terceirizados dos tribunais brasileiros, 2018b. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/2779>. Acesso em: 08 nov. 2024.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 284 de 05/06/2019.** Institui o Formulário Nacional de Avaliação de Risco para a prevenção e o enfrentamento de crimes e demais atos praticados no contexto de violência doméstica e familiar contra a mulher, 2019. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/2925>. Acesso em: 08 nov. 2024.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 332 de 21/08/2020.** Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências. DJe/CNJ, nº 274, de 25/08/2020, 2020b. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3429>. Acesso em: 23 out. 2025.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 348 de 13/10/2020.** Estabelece diretrizes e procedimentos a serem observados pelo Poder Judiciário, no âmbito criminal, com relação ao tratamento da população lésbica, gay, bissexual, transexual, travesti ou intersexo que seja custodiada, acusada, ré, condenada, privada de liberdade, em cumprimento de alternativas penais ou monitorada eletronicamente, 2020c. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3519>. Acesso em: 08 nov. 2024.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 376 de 02/03/2021.** Dispõe sobre o emprego obrigatório da flexão de gênero para nomear profissão ou demais designações na comunicação social e institucional do Poder Judiciário nacional, 2021b. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3765>. Acesso em: 08 nov. 2024.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 400 de 16/06/2021.** Dispõe sobre a política de sustentabilidade no âmbito do Poder Judiciário, 2021c. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3986>. Acesso em: 25 out. 2025.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 492 de 17/03/2023.** Estabelece, para adoção de Perspectiva de Gênero nos julgamentos em todo o Poder Judiciário, as diretrizes do protocolo aprovado pelo Grupo de Trabalho constituído pela Portaria CNJ n. 27/2021, institui obrigatoriedade de capacitação de magistrados e magistradas, relacionada a direitos humanos,

gênero, raça e etnia, em perspectiva interseccional, e cria o Comitê de Acompanhamento e Capacitação sobre Julgamento com Perspectiva de Gênero no Poder Judiciário e o Comitê de Incentivo à Participação Institucional Feminina no Poder Judiciário, 2023b. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/4986>. Acesso em: 08 nov. 2024.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 525 de 27/09/2023**. Altera a Resolução CNJ n. 106/2010, dispondo sobre ação afirmativa de gênero, para acesso das magistradas aos tribunais de 2º grau., 2023c. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/5277>. Acesso em: 04 ago. 2025.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 532 de 16/11/2023**. Determina aos tribunais e magistrados(as) o dever de zelar pelo combate a qualquer forma de discriminação à orientação sexual e à identidade de gênero, ficando vedadas, nos processos de habilitação de pretendentes e nos de adoção de crianças e adolescentes, guarda e tutela, manifestações contrárias aos pedidos pelo fundamento de se tratar de casal ou família monoparental, homoafetivo ou transgênero, e dá outras providências, 2023d. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/5342>. Acesso em: 08 nov. 2024.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 540 de 18/12/2023**. Altera a Resolução CNJ nº 255/2018 e dispõe sobre paridade de gênero, com perspectiva interseccional de raça e etnia, em atividades administrativas e jurisdicionais no âmbito do Poder Judiciário, 2023e. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/5391#:~:text=Altera%20a%20Resolu%C3%A7%C3%A3o%20CNJ%20n%C2%BA,no%20%C3%A2mbito%20do%20Poder%20Judici%C3%A1rio.&text=DJe%2FCNJ%20n.,2%2D4>. Acesso em: 08 nov. 2024.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 542 de 18/12/2023**. Cria o Fórum Nacional de Enfrentamento à Violência Contra a Mulher (Fonavim), com o objetivo de aprimorar a atuação do Poder Judiciário no enfrentamento à violência contra a mulher, 2023f. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/5393>. Acesso em: 08 nov. 2024.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 615 de 11/03/2025**. Estabelece diretrizes desenvolvimento, para utilização o e governança de soluções desenvolvidas com recursos de inteligência artificial no Poder Judiciário, 2025c. Disponível em: <https://atos.cnj.jus.br/files/original1555302025031467d4517244566.pdf>. Acesso em: 18 abr. 2025.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 654 de 04/11/2025**. Dispõe sobre o processo de vitaliciamento dos magistrados de 1º grau, 2025d. Disponível em: <https://atos.cnj.jus.br/files/original13492120251107690df8e187279.pdf>. Acesso em: 30 nov. 2025.

CONSELHO NACIONAL DE JUSTIÇA. **Uso da inteligência artificial no Poder Judiciário**: pesquisa nacional com tribunais. Brasília: CNJ, 2023g. Disponível em: https://bibliotecadigital.cnj.jus.br/jspui/bitstream/123456789/858/1/Pesquisa%20uso%20da%20inteligencia%20artificial%20IA%20no%20poder%20judici%C3%A1rio_2023.pdf. Acesso em: 30 jul. 2025.

CORRÊA, B. K. N. **Direito e tecnologia em perspectiva amefricana**: autonomia, algoritmos e vieses raciais. Tese de Doutorado, Pontifícia Universidade Católica do Rio de Janeiro.

Orientadora: Maria Celina Bodin de Moraes; Co-orientadora: Caitlin Sampaio Mulholland. 2021.

CORRÊA, G. F. A fixação do valor mínimo de reparação no estupro de vulnerável contra vítima do gênero feminino. **Revista CNJ**, Brasília, v. 8, n. 2, p. 147-167, jul./dez. 2024. Disponível em: <https://revistacnj.cnj.jus.br>. Acesso em: 20 out. 2025.

CORREIA, A. L. de S. Cibertecnologia, ciberfeminismo e tecnofeminismo: um novo olhar sobre a filosofia da tecnologia. **Controvérsia**, v. 16, n. 2, p. 3-14, maio-ago. 2020.

CÔRTEZ, A. de M. Discriminação judicial por identidade de gênero: um diagnóstico. **Rev. Direito e Práx.**, v. 10, n. 1, jan./mar. 2019. Disponível em: <https://www.scielo.br/j/rdp/a/hgvnGQFKcqVSjfn6gNqPwcT/?lang=pt>. Acesso em: 21 out. 2025. DOI: <https://doi.org/10.1590/2179-8966/2018/30750>. Disponível em: <https://www.scielo.br/j/rdp/a/hgvnGQFKcqVSjfn6gNqPwcT/?lang=pt>. Acesso em: 12 fev. 2026.

COSTA, D. C. **O viés do algoritmo e a discriminação por motivos relacionados à sexualidade**. 2020. 175f. Dissertação (Mestrado) – Universidade Federal da Bahia, Faculdade de Direito, Salvador, 2020.

CRUZ, F. B. da; BRITTO, M. C. de S.; MOREIRA, G. M.; BRITTO JUNIOR, A. de S. Robôs substituem juízes? O estado da arte da inteligência artificial no judiciário brasileiro. **Revista Antinomias**, v. 3, n. 1, jan./jun., 2022.

CRUZ, P. B. L.; SOUZA, P. V. N. C. S. de. Inteligência artificial no Poder Judiciário brasileiro: ameaças do robô-juiz aos direitos individuais e coletivos. **Revista de Estudos Interdisciplinares**, v. 7, n. 1, jan./fev. 2025.

CUI, Y. **Artificial Intelligence and Judicial Modernization**. Trad. Cao Yan; Liu Yan. Shangai: Springer, 2020.

D'ALMEIDA, A. C. (Coord.). **The Future of AI in the Brazilian Judicial System: Mapping, Integration, and Governance**. SIPA, [202?]. Disponível em: <https://itsrio.org/wp-content/uploads/2020/06/SIPA-Capstone-The-Future-of-AI-in-the-Brazilian-Judicial-System-1.pdf>. Acesso em: 27 jul. 2025.

DALLARI, D. de A. **O poder dos juízes**. 3ª ed. São Paulo: Saraiva, 2010.

DASTIN, J. Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women. **Reuters**, out. 2018. Disponível em: <https://www.reuters.com/article/usamazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showedbias-against-women-idUSKCN1MK08G/>. Acesso em: 21 nov. 2024.

DAZINGER, S.; LEVAV, J.; AVNAIM-PESSO, L. Extraneous factors in judicial decisions. **PNAS**, v. 108, n. 17, abr. 2011.

DEEKS, A. The Judicial Demand for Explainable Artificial Intelligence. **Columbia Law Review**, v. 119, n. 7, p. 1829-1850, 2019.

DHABLIYA, D.; DARI, S. S.; DHABLIA, A.; AKHILA, N.; KACHHORIA, R.; KHETANI, Addressing bias in machine learning algorithms: promoting fairness and ethical design. **E3S Web of Conferences**, v. 491, n. 02040, 2024. DOI:

<https://doi.org/10.1051/e3sconf/202449102040>. Disponível em: https://www.e3s-conferences.org/articles/e3sconf/abs/2024/21/e3sconf_icecs2024_02040/e3sconf_icecs2024_02040.html. Acesso em: 12 fev. 2026.

DICICCIO, C.; HSU, B.; YU, Y.; NANDY, P.; BASU, K. Detection and mitigation of algorithmic bias via predictive parity. In: **ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY (FAccT '23)**, 2023, Chicago, IL, USA. Proceedings... New York: Association for Computing Machinery, 2023. p. 1801-1816. DOI: <https://doi.org/10.1145/3593013.3594117>. Disponível em: <https://dl.acm.org/doi/10.1145/3593013.3594117>. Acesso em: 12 fev. 2026.

DODGE, J.; LIAO, Q. V.; ZHANG, Y.; BELLAMY, R. K. E.; DUGAN, C. Explaining models: An empirical study of how explanations impact fairness judgment. In: **Proceedings of the 24th International Conference on Intelligent User Interfaces**, p. 275-285, 2019. DOI: <https://doi.org/10.1145/3301275.3302310>. Disponível em: <https://dl.acm.org/doi/10.1145/3301275.3302310>. Acesso em: 12 fev. 2026.

DONEDA, D. C. M.; MENDES, L. S.; SOUZA, C. A. P. de; ANDRADE, N. N. G de. Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal. **Pensar**, Fortaleza, v. 23, n. 4, p. 1-17, dez. 2018.

DONEDA, D.; ALMEIDA, V. A. F. O que é governança de algoritmos? **PoliTICS**, out. 2016. Disponível em: <https://politics.org.br/pt-br/privacidade-e-vigilancia-news/o-que-e-governanca-de-algoritmos>. Acesso em: 15 ago. 2025.

DOSHI-VELEZ, F.; KIM, B. Towards a Rigorous Science of Interpretable Machine Learning. **arXiv preprint arXiv:1702.08608**, mar. 2017. Disponível em: <https://arxiv.org/abs/1702.08608>. Acesso em: 23 out. 2025.

DUARTE, A. M; BELLOC, I; BURIGO, J. Os corpos que governam o Brasil: do sujeito à sujeita universal. **Portal Catarinas**, 6 out. 2021, 12h56.

DYKSTRA, M. M. de M.; BRITTO, M. C. de S.; LEINER, G. Justiça 4.0: um contributo à otimização do acesso à justiça. In: PINTO, D. J. A.; SILVA, M. V. V. da. **Direito, governança e novas tecnologias III**. CONPEDI – XXIX Congresso Nacional do CONPEDI Balneário Camboriú – SC, Florianópolis: CONPEDI, 2022.

EQUIPO DE EXPERTOS EN CIENCIA Y TECNOLOGÍA. Ciencia y Tecnología: ¿Podemos confiar en las decisiones sugeridas por algoritmos de inteligencia artificial? **Universidad Internacional de Valencia**, 3 ago. 2018. Disponível em: <https://www.universidadviu.com/es/actualidad/nuestros-expertos/podemos-confiar-en-las-decisiones-sugeridas-por-algoritmos-de>. Acesso em: 18 maio 2024.

ESTÉBANEZ, M. E. La mujer y la ciencia iberoamericana. In: BIELLI, A. *et al.* **Proyecto iberoamericano de ciencia, tecnología y género - GENTEC - UNESCO** S. l.: OEI: Unesco, 2004. p. 1-39. Disponível em: http://www.comunicacion.amc.edu.mx/comunicacion/docs/Reporte_Final%20GENTEC.pdf. Acesso em: 12 set. 2022.

EYGHEN, H. V. Cognitive Bias: Phylogenesis or Ontogenesis? **Fortiers Psychology**, v. 13, jul. 2022. DOI: <https://doi.org/10.3389/fpsyg.2022.892829>. Disponível em: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.892829/full>. Acesso em: 12 fev. 2026.

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. L. F. de. **Inteligência Artificial**: Uma abordagem de aprendizagem de máquina. Rio de Janeiro: LTC, 2011.

FAIRLEARN. Fairness in Machine Learning: Common Fairness Metrics. Fairlearn, [20–]. Disponível em: https://fairlearn.org/main/user_guide/assessment/common_fairness_metrics.html. Acesso em: 22 out. 2025.

FERNANDES, L.; BERNARDES, M. N. Racismo institucional e as intervenções contra Janaína Quirino: desumanização de mulheres negras, sentidos de maternagem e as intervenções estatais. In: SEVERI, F. C. (Org.). **Reescrivendo decisões judiciais em perspectivas feministas**. Ribeirão Preto: Universidade de São Paulo, Núcleo de Pesquisa e Extensão “O Direito Achado na Rua”, 2022.

FISKE, S. T. Controlling other people: The impact of power on stereotyping. **American Psychologist**, v. 48, pp. 621-628, 1993.

FIT INSTITUTO DE TECNOLOGIA. Fairness em IA: o que significa e por que é importante. **Medium**, mar. 2025. Disponível em: <https://medium.com/@fitinstitutodetecnologia/fairness-em-ia-o-que-significa-e-por-que-é-importante-6920a5535b63>. Acesso em: 22 out. 2025.

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

FLORES, J. H. **A reinvenção dos direitos humanos**. Trad. Carlos Roberto Diogo Garcia. Florianópolis: Fundação Boiteux, 2009.

FONG, T.; NOURBAKHSH, I.; DAUTENHAHN, K. A Survey of Socially Interactive Robots. **Robotics and Autonomous Systems**, v. 42, n. 3-4, pp. 143-166, 2003.

FONSECA, A. P. **A disrupção do Direito pela inteligência artificial e os seus reflexos no ativismo judicial**. Tese de Doutorado – Orientador: Prof. Dr. Carlos Eduardo Dieder Reverbel, Universidade Federal do Rio Grande do Sul. Porto Alegre, 2022.

FORBES TECHNOLOGY COUNCIL. AI and Fairness Metrics: How to Quantify and Mitigate Bias in AI Systems. **Forbes Councils**, jul. 2024. Disponível em: <https://councils.forbes.com/blog/ai-and-fairness-metrics>. Acesso em: 22 out. 2025.

FORGAS-COLL, S.; HUERTAS-GARCIA, R.; ANDRIELLA, A.; ALENYÀ, G. The effects of gender and personality of robot assistants on customers’ acceptance of their service. **Service Business**, v. 16, pp. 359-389, 2022.

FOUCAULT, M. O sujeito e o poder. In: DREYFUS, H. L.; RABINOW, P. (Orgs.), Michel Foucault. **Uma trajetória filosófica**: Para além do estruturalismo e da hermenêutica. Rio de Janeiro: Forense Universitária, 1995.

FOUCAULT, M. **Vigiar e punir: nascimento da prisão**. 31. ed. Petrópolis: Vozes, 2006.

FOUCAULT, M. **História da sexualidade I: A vontade de saber**. 7a ed. Rio de Janeiro: Graal, 1990.

FOULDS, J., ISLAM, R., KEYA, K. N. AND PAN, S. An Intersectional Definition of Fairness. **2020 IEEE 36th International Conference on Data Engineering (ICDE)**, 2020.

FRAGALE FILHO, R.; MOREIRA, R. S.; SCIAMMARELLA, A. P. de O. Magistratura e gênero: um olhar sobre as mulheres nas cúpulas do judiciário brasileiro. **E-cadernos CES**, n. 24, p. 57-77, 2015.

FRANK, Jerome. **Law & the Modern Mind**. London/New York: Routledge, 2017.

FRANQUEIRA, B. D. **Como a inteligência artificial reforça a discriminação de gênero no ambiente de trabalho**. 2019. 59f. Trabalho de conclusão de curso – Fundação Getúlio Vargas – FGV, Rio de Janeiro - RJ, 2019.

FREITAS, C. O. de A. **Direito e não-coisas**. Rio de Janeiro: Lumen Juris, 2025.

FREITAS, C. O. de A. Redes Sociais: Sociedade Tecnológica e Inclusão Digital. In: WACHOWICZ, M. (Org.). **Direito da Sociedade da Informação & Propriedade Intelectual**. Curitiba: Juruá, 2012.

FRIEDMAN, B.; NISSENBAUM, H. Bias in Computer Systems. **ACM Transactions on Information Systems**, v. 14, n. 3, pp. 330-347, jul. 1996.

FRIEDRICH, D. B.; PHILIPPI, J. M. Inclusão digital e blockchain como instrumentos para o desenvolvimento econômico. **International Journal of Digital Law**, Belo Horizonte, a. 1, n. 1, p. 97-115, jan./abr. 2020.

GABRIEL, A.; PORTO, F. R.; ARAÚJO, V. S. de. Justiça 4.0 e a Inteligência Artificial: pragmática eficiência. **Revista de Análise Econômica do Direito**, [s. l.], v. 9, 2025.

GARCÍA, S. M. Ética e inteligencia artificial. **Cuadernos de la Cátedra CaixaBank de Responsabilidad Social Corporativa**, n. 42, set. 2019.

GARCÍA-ULL, F.; MELERO-LÁZARO, M. Gender stereotypes in AI-generated images. **Profesional de la información**, v. 32, n. 5, e320505, 2023.

GARG, P.; VILLASENOR, J.; FOGGO, V. Fairness Metrics: A Comparative Analysis. In: **2020 IEEE International Conference on Big Data (Big Data)**. IEEE, 2020. p. 3662-3666. DOI: 10.1109/BigData50022.2020.9378025. Disponível em: <https://ieeexplore.ieee.org/document/9378025>. Acesso em: 12 fev. 2026.

GARGARELLA, R. Del reino de los jueces al reino de los políticos. **Jueces para la democracia**, n. 28, pp. 24-28, 1997.

Gender Equality and Women's Rights 25 Years after Beijing. Discussion Paper no. 36. Ago, 2020.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GILLESPIE, T. Algorithm. In: PETERS, B. **Digital Keywords**. Princeton: Princeton University Press, 2016, p. 18-30.

GILPIN, L. H.; BAU, D.; YUAN, B. Z.; BAJWA, A.; SPECTER, M.; KAGAL, L. Explaining Explanations: An Overview of Interpretability of Machine Learning. **IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)**, 2018. DOI: 10.1109/DSAA.2018.00018. Disponível em: <https://ieeexplore.ieee.org/document/8631448>. Acesso em: 12 fev. 2026.

GODOY, A. S. Introdução à pesquisa qualitativa e suas possibilidades. **Revista de Administração de Empresas**, São Paulo, v. 35, n. 2, p. 57-63, mar./abr. 1995.

GOLDSTEIN, W. M.; HOGARTH, R. M. Judgement and decision research: Some historical context. **Research on Judgment and Decision Making: Currents, Connections, and Controversies**, p. 1-65, 1997.

GOMES, J. C. A.; NOGUEIRA, R.; ARGUELHES, D. W. Gênero e comportamento judicial no Supremo Tribunal Federal: os ministros confiam menos em relatoras mulheres? **Revista Brasileira de Políticas Públicas**, Brasília, v. 8, n. 2, p. 854-876, 2018. DOI: <https://doi.org/10.5102/rbpp.v8i2.5326>. Disponível em: <https://www.publicacoes.uniceub.br/RBPP/article/view/5326>. Acesso em: 12 fev. 2026.

GONZÁLES, A. S.; RAMPINO, L. A design perspective on how to tackle gender biases when developing AI-driven systems. **AI and Ethics**, jan. 2024. DOI: <https://doi.org/10.1007/s43681-023-00386-2>. Disponível em: <https://link.springer.com/article/10.1007/s43681-023-00386-2>. Acesso em: 12 fev. 2026.

GOODWIN, S. A.; GUBIN, A.; FISKE, S. T.; YZERBYT, V. T. Power can bias subordinates by default and by design. **Group Processes & Intergroup Relations**, v. 3, pp. 227-256, 2000.

GUNNING, D.; AHA, D. DARPA's Explainable Artificial Intelligence (XAI) Program. **AI Magazine**, v. 40, n. 2, 44-58, 2019. DOI: <https://doi.org/10.1609/aimag.v40i2.2850>. Disponível em: <https://onlinelibrary.wiley.com/doi/epdf/10.1609/aimag.v40i2.2850>. Acesso em: 12 fev. 2026.

HAI AHMAND, N.; STIGHOLT, L.; PENZENSTADLER, B. Ai Systems' Negative Social Impacts and Their Potential Factors. **SSRN**, mar. 2022. Disponível em: <https://ssrn.com/abstract=4047871>. Acesso em: 12 fev. 2025.

HALL, S. **A identidade cultural na pós-modernidade**. Trad. por Tomaz Tadeu da Silva, Guacira Lopes Louro. 11ª ed. 1ª reimp. Rio de Janeiro: DP&A Editora, 2011.

HARAWAY, D. A Cyborg Manifesto: science, technology, and socialist-feminism in the late 20th century. In: Weiss, J.; Nolan, J.; Hunsinger, J.; Trifonas, P. (Eds.) **The International handbook of virtual learning environments**. Dordrecht: Springer, 2006.

HARAWAY, D. **A manifesto for Cyborgs**: science, technology, and socialist feminism in the 1980s. *Aust Fem Stud.*, v. 2, n. 4, pp. 1-42, 1987. DOI: Disponível em: <https://doi.org/10.1080/08164649.1987.9961538>.

<https://www.tandfonline.com/doi/abs/10.1080/08164649.1987.9961538>. Acesso em: 12 fev. 2026.

HARAWAY, D. J. **Ciencia, cyborgs y mujeres: La reinvención de la natureza**. Trad. Manuel Talens. Ediciones Cátedra, Universitat de València, Instituto de la Mujer, 1991.

HARAWAY, D. **When species meet**. Minneapolis: University of Minnesota Press, 2008.

HARDT, M.; PRICE, E.; SREBRO, N. Equality of opportunity in supervised learning. **30th Conference on Neural Information Processing Systems (NIPS 2016)**, Barcelona, 2016.

HASELTON, M. G., NETTLE, D., & MURRAY, D. R. The evolution of cognitive bias. In: D. M. BUSS (Ed.). **The handbook of evolutionary psychology: Integrations**. 2ª ed. Nova Jersey: John Wiley & Sons, Inc., p. 968-987, 2016.

HEIKKILÄ, M. The viral AI avatar app Lensa undressed me – without my consent: My avatars were cartoonishly pornified, while my male colleagues got to be astronauts, explorers, and inventors. **MIT Technology Review**, 12 dez. 2022. Disponível em: <https://www.technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me-without-my-consent>. Acesso em: 21 nov. 2024.

HEIKKILÄ, M. These new tools let you see for yourself how biased AI image models are: Bias and stereotyping are still huge problems for systems like DALL-E 2 and Stable Diffusion, despite companies' attempts to fix it. **MIT Technology Review**, 22 mar. 2023. Disponível em: <https://www.technologyreview.com/2023/03/22/1070167/these-news-tool-let-you-see-for-yourself-how-biased-ai-image-models-are/>. Acesso em: 21 nov. 2024.

HEMPEL, J. Melinda Gates and Fei-Fei Li Want to Liberate AI from “Guys With

HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE. **Draft Ethics Guidelines for Trustworthy AI**. European Commission: Brussels, 2018. Disponível em: <https://digital-strategy.ec.europa.eu/en/library/draft-ethics-guidelines-trustworthy-ai>. Acesso em: 1 nov. 2025.

HOFFMANN-RIEM, W. BIG DATA E INTELIGÊNCIA ARTIFICIAL: desafios para o Direito. **REI - Revista Estudos Institucionais**, v. 6, n. 2, p. 431-506, 2020.

HOLANDA, M. C. **Inteligência artificial no Poder Judiciário: revolução tecnológica, governança, gestão de riscos e desenvolvimento responsável nos sistemas de justiça**. Dissertação (Mestrado). Orientação: Prof. Dr. Hugo de Brito Machado Segundo, Centro Universitário Christus - Unichristus, Mestrado em Direito, Fortaleza, 2025.

Hoodies”. **WIRED**, 4 maio 2017. Disponível em: <https://www.wired.com/2017/05/melinda-gates-and-fei-fei-li-want-to-liberate-ai-from-guys-with-hoodies/>. Acesso em: 21 nov. 2024.

HOOKER, S. Moving beyond “algorithmic bias is a data problem”. **Patterns**, v. 2, n. 4, p. 100241, 2021. DOI: <https://doi.org/10.1016/j.patter.2021.100241>. Disponível em: [https://www.cell.com/patterns/fulltext/S2666-3899\(21\)00061-1?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS2666389921000611%3Fshowall%3Dtrue](https://www.cell.com/patterns/fulltext/S2666-3899(21)00061-1?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS2666389921000611%3Fshowall%3Dtrue). Acesso em: 12 fev. 2026.

HOOKS, B. **Feminism Theory**: from the margin to the center. Routledge: New York, 2015.

HORTA, R. de L. e. **Direito e tomada de decisão**: elementos para uma Teoria da Decisão Jurídica. 2019. 188 f. Tese (Doutorado em Direito) – Universidade de Brasília, Brasília, 2019.

HUMAN RIGHTS COUNCIL. Annual report of the United Nations High Commissioner for Human Rights and reports of the Office of the High Commissioner and the Secretary-General. **A/HRC/48/31**. 48 Session. 13 set. 2021. Disponível em: <https://docs.un.org/en/A/HRC/48/31>. Acesso em: 12 fev. 2026.

HUTCHINSON, B.; MITCHELL, M. 50 Years of Test (Un)fairness: Lessons for Machine Learning. **Proceedings of the Conference on Fairness, Accountability, and Transparency**, ACM, 29 jan. 2019. DOI: <https://doi.org/10.1145/3287560.3287600>. Disponível em: <https://dl.acm.org/doi/10.1145/3287560.3287600>. Acesso em: 12 fev. 2026.

IBGE – Instituto Brasileiro de Geografia e Estatística. Estatísticas de Gênero - Indicadores sociais das mulheres no Brasil: Tabelas – Indicadores sociais das mulheres no Brasil – 3ª Edição, **IBGE**, [2013?]. Disponível em: <https://www.ibge.gov.br/estatisticas/multidominio/genero/20163-estatisticas-de-genero-indicadores-sociais-das-mulheres-no-brasil.html>. Acesso em 21 abr. 2025.

IBM. O que é uma rede neural? **IBM**, [20-?]. Disponível em: <https://www.ibm.com/br-pt/topics/neural-networks>. Acesso em: 11 fev. 2025.

IGUAL, L.; SEGUÍ, S. **Introduction to Data Science**: A Python Approach to Concepts, Techniques and Applications. Springer: Switzerland, 2017.

INSTITUTE OF ELECTRICAL AND ELECTRONIC ENGINEERS (IEEE). **Ethically Aligned Design**: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. Version 2. 2017. Disponível em: https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf. Acesso em: 24 out. 2025.

INTITUTO BRASILEITO DE GEOGRAFIA E ESTATÍSTICA. Estatísticas de gênero: indicadores sociais das mulheres no Brasil. **Estudos e Pesquisas - Informação Demográfica e Socioeconômica**, n. 38, 2018.

JACOBSTEIN, N. apud WARD, Mark. Does Rampante AI Threaten Humanity? **BBC News**, 2 dez. 2014. Tech. Disponível em: <https://www.bbc.com/news/technology-30293863>. Acesso em: 28 maio 2021.

JUNQUILHO, T. A. **Inteligência artificial no direito**: limites éticos. São Paulo: Juspodivm, 2022.

JUSTIÇA FEDERAL 2ª REGIÃO. Ferramenta de IA ApoIA já está disponível na Plataforma Digital do Poder Judiciário Brasileiro. **JF2 – JUSTIÇA FEDERAL 2ª REGIÃO**, 5 maio 2025. Disponível em: <https://www.trf2.jus.br/jf2/aviso-jf2/2025/ferramenta-de-ia-apoia-ja-esta-disponivel-na-plataforma-digital-do-poder/>. Acesso em: 2 ago. 2025.

KAHNEMAN, D. **Rápido e devagar**. Rio de Janeiro: Objetiva, 2012.

KAHNEMAN, D.; TVERSKY, A. The simulation heuristic. In: KAHNEMAN, D.; TVERSKY, A.; SLOVIC, P. (Org.). **Judgment under uncertainty: heuristics and biases**. New York: Cambridge University Press, 1982.

KAHNEMAN, D.; TVERSKY, A. Thinking, fast and slow. New York: Farrar Straus Giroux, 2013.

KELLER, E. F. Gender and Science. In: HARDING, S.; HINTIKKA, M. B. (Eds.). **Discovering Reality: Feminist Perspectives on Epistemology, Metaphysics, Methodology, and Philosophy of Science**. Kluwer Academic Publishers: New York, Boston, Dordrecht, London, Moscow, 2004.

KELSEN, H. **Teoria pura do Direito**. 6ª ed. São Paulo: Martins Fontes, 1998.

KUMAR, S.; CHOUDHURY, S. Gender and feminist considerations in artificial intelligence from a developing-world perspective, with India as a case study. **Humanities and social sciences communications**, v. 9, n. 31, 2022.

KWON, J.; YUN, H. AI Chatbot Shut Down After Learning to Talk Like a Racist Asshole. **Vice World News**, 12 jan. 2021. Disponível em: <https://www.vice.com/en/article/ai-chatbot-shut-down-after-learning-to-talk-like-a-racist-asshole/>. Acesso em: 21 nov. 2024.

LAGES, F. M.; CAÇÃO, T. R. A mulher no poder judiciário brasileiro: desafios e conquistas. **Revista de Direito Brasileira**, v. 26, n. 10, p. 106-121, 2020. Disponível em: <https://www.indexlaw.org/index.php/rdb/article/view/5291>. Acesso em: 20 out. 2025.

LAMARTINE, C. Interseccionalidade e ciberfeminismo: Construindo laços transfronteiriços nos Feminismos Contemporâneos. **Contracorrente: Revista do Programa de Pós-Graduação Interdisciplinar em Ciências Humanas**, n. 20, p. 98-109, dez. 2023.

LAMBRECHT, A.; TUCKER, C. Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. **Management Science**, v. 65, n. 7, 2019. DOI: <https://doi.org/10.1287/mnsc.2018.3093>. Disponível em: <https://pubsonline.informs.org/doi/10.1287/mnsc.2018.3093>. Acesso em: 12 fev. 2026.

LARSON, Jeff; MATTU, Surya; KIRCHNER, Lauren; ANGWIN, Julia. How We Analyzed the COMPAS Recidivism Algorithm. **Propublica**, maio 2016. Disponível em: <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. Acesso em: 23 out. 2019.

LATOUR, B. **Jamais fomos modernos: ensaio de antropologia simétrica**. Trad. Carlos Irineu da Costa. Rio de Janeiro: 34 Licenciatura S/C Ltda, 1994.

LEAVY, S. Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning. In: **Proceedings of the International Conference on Software Engineering**. IEEE Computer Society, pp. 14-16, 2018.

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

LEINER, G. S.; DYKSTRA, M. M. de M.; BRITTO, M. C. de S.; CRUZ, F. B. da C. A otimização do acesso à justiça no Brasil sob a perspectiva da teoria crítica dos direitos humanos de Herrera Flores. **Emancipação**, Ponta Grossa - PR, Brasil., v. 25, p. 1–14, 2025. DOI: 10.5212/Emancipacao.v.25.2524207.034. Disponível em: <https://revistas.uepg.br/index.php/emancipacao/article/view/24207>. Acesso em: 23 out. 2025.

LEWIS, N. D. **Learning from Data Made Easy with R: A Gentle Introduction for Data Science**. N.D. Lewis, 2016.

LIMA, A. B. de M.; OLIVEIRA, G. H. de. Acesso à justiça e o impacto de novas tecnologias na sua efetivação. **Revista de Cidadania e Acesso a Justiça**, v. 5, n. 1, p. 69-87, jan./jun. 2019.

LIMA, F. A. de. Depoimentos e Vivências de Mulheres nos Cursos de Computação da Universidade Tecnológica Federal do Paraná. **Informática na Educação: teoria & prática**, Porto Alegre, v. 19, n. 3, p. 161-177, set./dez. 2016.

LIMA, M. P. As mulheres na Ciência da Computação. **Estudos Feministas**, Florianópolis, v. 21 n. 3, 496, set./dez. 2013.

LIPTON, Z. C. The Mythos of Model Interpretability. **Communications of the ACM**, v. 61, n. 10, p. 36-43, 2018. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/3233231>. Acesso em: 23 out. 2025.

LUNDBERG, S. M.; LEE, S. A unified approach to interpreting model predictions. In: **Advances in Neural Information Processing Systems (NeurIPS 2017)**, 31., 2017, Long Beach, CA. Proceedings... Red Hook, NY: Curran Associates, 2017. Disponível em: https://papers.nips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf. Acesso em: 23 out. 2025.

LUNDEMIR, T. B. Inteligência artificial e aprendizado de máquina: estado atual e tendências. **Estudos Avançados**, v. 35, n. 101, p. 85-94, 2021.

MACHADO, H. F. de S. Algoritmos, regulação e governança: uma revisão de literatura. **Revista de Direito Setorial e Regulatório**, Brasília, v.4, n.1, p.39-62, maio 2018.

MACHADO, L. A.; NEGREIROS, E. C. de. Teoria da interseccionalidade e feminismo contemporâneo: uma análise crítica das complexas experiências das mulheres. **Revista Direito e Sexualidade**, Salvador, v. 5, n. 1, p. 142–162, 2024.

MARCONI, M. de A.; LAKATOS, E. M. **Técnicas de pesquisa**. 5ª ed. rev. ampl. São Paulo: Atlas, 2002.

MARON, D. F. Science Career Ads Are Disproportionately Seen by Men. **Scientific American**, 25 jul. 2018. Disponível em: <https://www.scientificamerican.com/article/science-career-ads-are-disproportionately-seen-by-men/#:~:text=Women%20see%20fewer%20advertisements%20about,the%20economics%20of%20ad%20sales>. Acesso em: 21 nov. 2024.

MATOS, A. S. de M. C.; MILÃO, D. A. P. Decisionismo e hermenêutica negativa: Carl Schmitt, Hans Kelsen e a afirmação do poder no ato interpretativo do direito. **Sequência**, n. 67, dez. 2013.

MCCARTHY, J.; MINSKY, M. L.; ROCHESTER, N.; SHANNON, C. E. **A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence**, ago. 1995.

Disponível em: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>. Acesso em: 20 maio 2024.

MEHRABI, N.; MORSTATTER, F.; SAXENA, N.; LERMAN, K.; GALSTYAN, A. A Survey on Bias and Fairness in Machine Learning. **ACM Computing Surveys**, v. 54, n. 6, 2021.

MELO, H. P. de; THOMÉ, D. **Mulheres e poder: histórias, ideias e indicadores**. Rio de Janeiro: FGV, 2018.

MENDONÇA, E.; BARROSO, L. R. Retrospectiva 2012: STF entre seus papéis contramajoritário e representativo. **Consultor Jurídico**, 3 de jan. 2013, 7h39. Disponível em: <http://www.conjur.com.br/2013-jan-03/retrospectiva-2012-stf-entre-papeis-contramajoritario-representativo>. Acesso em: 27 maio 2018.

MICHAELIS. Dicionário Brasileiro da Língua Portuguesa. Disponível em: <https://michaelis.uol.com.br/palavra/O3BL/decidir/>. Acesso em: 05 maio 2024.

MILLER, T. Explanation in Artificial Intelligence: Insights from the Social Sciences. **Artificial Intelligence**, v. 267, p. 1-38, 2019. DOI: <https://doi.org/10.1016/j.artint.2018.07.007>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0004370218305988>. Acesso em: 23 out. 2025.

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES. **Estratégia Brasileira de Inteligência Artificial - EBIA**, 2021. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-documento_referencia_4-979_2021.pdf. Acesso em: 17 jan. 2025.

MITCHELL, T. **Machine Learning**. McGraw-Hill, 1 ed. 1997.

MITTELSTADT, B. D.; ALLO, P.; TADDEO, M.; WACHTER, S.; FLORIDI, L. The ethics of algorithms: Mapping the debate. **Big Data & Society**, pp. 1-21, jul./dez. 2016.

MÓDULO DE PESSOAL E ESTRUTURA JUDICIÁRIA MENSAL DO PODER JUDICIÁRIO. Dados de pessoal do Poder Judiciário. **CNJ**, 20 out. 2025, 05:01:07. Disponível em: <https://justica-em-numeros.cnj.jus.br/painel-mpm-pessoal/>. Acesso em: 20 out. 2025.

MONTEIRO, R. L. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. 385f. Tese (Doutorado em Direito) – Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021.

MONTREAL DECLARATION FOR A RESPONSIBLE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE, 2018. Disponível em: <https://declarationmontreal->

iaresponsible.com/wp-content/uploads/2023/04/UdeM_Decl-IA-Resp_LA-Declaration-ENG_WEB_09-07-19.pdf. Acesso em: 18 maio 2024.

MONTREAL DECLARATION RESPONSIBLE AI. The Montréal Declaration for a Responsible Development of Artificial Intelligence, [2018?]. Disponível em: <https://montrealdeclaration-responsibleai.com/the-declaration/>. Acesso em: 18 maio 2024.

MOREIRA, J. A.; SILVA, R. M.; CARVALHO, M. E. P. Cenários Prospectivos: Uma Visão do Futuro da Presença Feminina em Cursos de Ciência da Computação de uma Instituição de Ensino Superior. In: Anais do XXVI workshop sobre educação em computação, 2018. DOI: <https://doi.org/10.5753/wei.2018.3519>. Disponível em: <https://sol.sbc.org.br/index.php/wei/article/view/3519>. Acesso em: 12 fev. 2026.

MOZETIĆ, V. A. Inteligência Artificial no Judiciário brasileiro: salvaguardas, riscos e novas fronteiras. **Revista Do Tribunal Regional Federal Da 3ª Região**, v. 36, n. 161, p. 17-48, jun. 2025. Disponível em: <https://www.revista.trf3.jus.br/index.php/rtrf3/article/view/721>. Acesso em: 15 ago. 2025.

MULHOLLAND, C. Responsabilidade civil e processos decisórios autônomos. In: **Inteligência artificial e direito**. 2ª Ed. São Paulo: Ed. Revista dos Tribunais, 2020, p. 327-350.

NACIONES UNIDAS; CEPAL. Cepalstat: Base de Datos y Publicaciones Estadísticas: Temas transversales / Género / Mujeres en la adopción de decisiones. **NACIONES UNIDAS; CEPAL**, [20-]. Disponível em: https://statistics.cepal.org/portal/cepalstat/dashboard.html?indicator_id=1347&area_id=543&lang=es. Acesso em: 21 abr. 2025.

NAÇÕES UNIDAS BRASIL. Objetivos de desenvolvimento sustentável: Sobre o nosso trabalho para alcançar os Objetivos de Desenvolvimento Sustentável no Brasil. **NAÇÕES UNIDAS BRASIL**, [20-?]. Disponível em: <https://brasil.un.org/pt-br/sdgs>. Acesso em: 26 out. 2025.

NICKERSON, Raymond S. Confirmation bias: A ubiquitous phenomenon in many guises. **Review of General Psychology**, v. 2, n. 2, 175-220, 1998. Disponível em: https://www.researchgate.net/publication/280685490_Confirmation_Bias_A_Ubiquitous_Phenomenon_in_Many_Guises. Acesso em: 14 maio 2021.

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

NORTHPOINTE. **COMPAS Risk & Need Assessment System**, 2012. Disponível em: http://www.northpointeinc.com/files/downloads/FAQ_Document.pdf. Acesso em: 23 out. 2019.

NUNES PECEGO, D.; LOBATO COLLET JANNY TEIXEIRA, R. Inteligência artificial no judiciário: da opacidade à explicabilidade das decisões judiciais. **Revista da Faculdade de Direito da UERJ - RFD**, [S. l.], n. 43, p. 1-22, 2024. DOI: 10.12957/rfd.2024.87850. Disponível em: <https://www.e-publicacoes.uerj.br/rfduerj/article/view/87850>. Acesso em: 15 ago. 2025.

NUNES, D. Uma introdução sobre o uso dos modelos de inteligência artificial analíticas e generativas no direito processual. In: VADELL et al. (Orgs.). **O sistema processual de século XXI: novos desafios**. Londrina: Ed. Thoth, 2023.

NUNES, D.; LUD, N.; PEDRON, F. **Desconfiando da imparcialidade dos sujeitos processuais**. Salvador: Juspodivm, 2018.

NUNES, D.; MARQUES, A. L. P. **Inteligência Artificial e Direito Processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas**. Revista de Processo, v. 285/2018, p. 421-477, 2018.

NUNES, I; MOREIRA, A.; ARAÚJO, J. GIRE: Gender-Inclusive Requirements Engineering. **Data & Knowledge Engineering**, v. 143, 2023. DOI: <https://doi.org/10.1016/j.datak.2022.102108>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169023X22000994?via%3Dihub>. Acesso em: 12 fev. 2026.

OBERMEYER, Z.; POWERS, B.; VOGELI, C.; MULLAINATHAN, S. Dissecting racial bias in an algorithm used to manage the health of populations. **Science**, v. 366, i. 6464, pp. 447-453, 2019.

OECD.AI Policy Observatory; GPAI Global Partnership. OECD AI Principles overview: The OECD AI Principles promote use of AI that is innovative and trustworthy and that respects human rights and democratic values. Adopted in May 2019, they set standards for AI that are practical and flexible enough to stand the test of time. OECD. **OECD.AI Policy Observatory; GPAI Global Partnership**, [20-]. Disponível em: <https://oecd.ai/en/ai-principles>. Acesso em: 8 maio 2021.

OLIVEIRA, I. J. da S. **Direito, lógica e inteligência artificial: por quê, como e em que medida automatizar a solução judicial de conflitos no Brasil**. Orientador: Torquato da Silva Castro Júnior; Coorientador: Frederico Luiz Gonçalves de Freitas. Tese (Doutorado) – Universidade Federal de Pernambuco, 107f., 2019.

OLIVEIRA, M. T. V. da S. Inteligência artificial e o Poder Judiciário. **Consultor Jurídico**, abr., 2021. Opinião. Disponível em: <https://www.conjur.com.br/2021-abr-20/maria-teresa-oliveira-inteligenciaartificial-poder-judiciario>. Acesso em: 28 mar. 2021.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. **Declaração Universal dos Direitos Humanos**. Resolução 217 A (III), 10 de dezembro de 1948, da Assembleia Geral das Nações Unidas. Disponível em: <https://www.unicef.org/brazil/declaracao-universal-dos-direitos-humanos>. Acesso em: 20 set. 2022.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. **Pacto Internacional sobre Direitos Civis e Políticos**. Nova Iorque, 16 de dezembro de 1966. Entrou em vigor em 23 de março de 1976. Disponível em: https://dcjri.ministeriopublico.pt/sites/default/files/documentos/instrumentos/pacto_internacional_sobre_os_direitos_civis_e_politicos.pdf. Acesso em: 20 set. 2022.

OTONI, L. Pandemia leva Judiciário a acelerar adaptação tecnológica. **Conselho Nacional de Justiça**, 24 ago. 2020. Disponível em: <https://www.cnj.jus.br/pandemia-leva-judiciario-a-acelerar-adaptacao->

tecnologica/#:~:text=O%20novo%20normal%20imposto%20pela,10%20%E2%80%93%20u%20mais%20%E2%80%93%20anos. Acesso em: 22 nov. 2024.

PEER, Eyal; GAMLIEL, Eyal. Heuristics and Biases in Judicial Decisions. **Court Review**, v. 39, 2013. Disponível em: <https://heinonline.org/HOL/LandingPage?handle=hein.journals/ctrev49&div=20&id=&page=>. Acesso em: 21 out. 2019.

PEREIRA, D. Tribunais de todo o país já podem utilizar primeira IA generativa integrada à PDPJ-Br. **CNJ – Conselho Nacional de Justiça**, Brasília, 20 maio 2025. Disponível em: <https://www.cnj.jus.br/tribunais-de-todo-o-pais-ja-podem-utilizar-primeira-ia-generativa-integrada-a-pdpj-br/>. Acesso em: 2 ago. 2025.

PEREIRA, J. R. G.; OLIVEIRA, R. M. de. Hércules, Hermes e a pequena sereia: Uma reflexão sobre estereótipos de gênero, subrepresentação das mulheres nos tribunais e (i)legitimidade democrática do Poder Judiciário. **Revista Brasileira de Políticas Públicas**, Dossiê especial: indução de comportamentos (Neurolaw): Direito, psicologia e neurociência, v. 8, n. 2, ago. 2018.

PERROT, M. **As mulheres ou os silêncios da história**. Bauru, SP: [s.n.], 2005.

PICCOLI, A. M. Judiciário exponencial: premissas para acelerar o processo de inovação. In: FERNANDES, R. V. de C.; CARVALHO, A. G. P. de (Coord.). **Tecnologia jurídica & direito digital: II Congresso Internacional de Direito, Governo e Tecnologia – 2018**. Belo Horizonte: Fórum, 2018. p. 191-204. ISBN 978-85-450-0584-1.

PINTO, H. A. A utilização da inteligência artificial no processo de tomada de decisões: por uma necessária accountability. **RIL Brasília**, a. 57, n. 225, p. 43-60, jan./mar. 2020.

PIOVESAN, F. **Temas de direitos humanos**. 10 ed. rev. atual. ampl. São Paulo: Saraiva, 2017.

PODER JUDICIÁRIO; CONSELHO NACIONAL DE JUSTIÇA. **Diagnóstico da participação feminina no Poder Judiciário**, 2019. Disponível em: <https://www.cnj.jus.br/wp-content/uploads/2021/08/relatorio-participacaofeminina.pdf>. Acesso em: 05 nov. 2024.

PODER JUDICIÁRIO; CONSELHO NACIONAL DE JUSTIÇA. **Perfil sociodemográfico dos magistrados brasileiros 2018**, 2018. Disponível em: <https://www.cnj.jus.br/wp-content/uploads/conteudo/arquivo/2018/09/49b47a6cf9185359256c22766d5076eb.pdf>. Acesso em: 02 set. 2024.

POHL, R. **An analysis of Donna Haraway's A cyborg manifesto: science, technology, and socialist-feminism in the late twentieth century**. London: Routledge, 2018.

PORTO, F. R. **O impacto da utilização da inteligência artificial no executivo fiscal**. Estudo de caso do Tribunal de Justiça do Rio de Janeiro. In: FERNANDES, R. V. de C.; CARVALHO, A. G. P. de (Coords.). **Tecnologia Jurídica & Direito Digital - II Congresso Internacional de Direito, Governo e Tecnologia – 2018**. Belo Horizonte: Fórum, 2018.

POSNER, R. A. **Cómo deciden los jueces**. Trad. Victoria Roca Pérez. Madrid: Marcial Pons, 2011.

QUIRÓS, C. T., E. G. MORALES, R. R. PASTOR, A. F. CARMONA, M. S. IBÁÑEZ, AND U. M. HERRERA. **Women in the Digital Age**. Brussels: European Commission, 2018.

REUTERS. Amazon desiste de ferramenta secreta de recrutamento: Sistema automatizado mostrou viés contra as mulheres. **Forbes**, 10 out. 2018. Disponível em: <https://forbes.com.br/negocios/2018/10/amazon-desiste-de-ferramenta-secreta-de-recrutamento/>. Acesso em: 21 nov. 2024.

REYES, P. Servicios Legaltech. **Revista Iberoamericana de Derecho Informático** (Segunda época). Federación Iberoamericana de Asociaciones de Derecho e Informática, 2019.

REZENDE, M. C. de M. **Democratização do Poder Judiciário no Brasil**. São Paulo: Contracorrente, 2018.

RIBEIRO, D. G.; CASSOL, J. Inteligência Artificial e Direito: uma análise prospectiva dos sistemas inteligentes no processo judicial. In: JÚNIOR, A. P. G. (Org.). **Direito Processual em movimento**. Volume IX. Curitiba: CRV, 2021, p. 243-266.

RIBEIRO, M. T.; SINGH, S.; GUESTIN, 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 (KDD '16)**. Association for Computing Machinery, New York, NY, USA, pp. 1135-1144, 2016. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/2939672.2939778>. Acesso em: 26 set. 2021.

RIBEIRO, M. V. M. A importância da inteligência artificial no Poder Judiciário brasileiro. **Revista DELOS**, Curitiba, v. 17, n. 61, p. 1-17, 2024.

RIO GRANDE DO SUL. Tribunal de Justiça do Estado do Rio Grande do Sul. Apelação Criminal nº 70074065210.CNJ: 0170636-96.2017.8.21.7000. 1ª Câmara Criminal. Relator Des. Jayme Weingartner Neto. Data do Julgamento: 27 set. 2017. Data da Publicação: 26 nov. 2017. Porto Alegre, 2017.

RIO GRANDE DO SUL. Poder Judiciário do Rio Grande do Sul. Vara do Júri. 2017.

RODRIGUES, T. M.; SANTOS, K. E. G. dos. Acesso à justiça e tecnologia. **Revista em Tempo**, [S.l.], v. 20, n. 1, nov. 2020. ISSN 1984-7858. Disponível em: <https://revista.univem.edu.br/emtempo/article/view/3259>. Acesso em: 28 out. 2022.

ROMERO, A. El tecnofeminismo de Judy Wajcman: alcance teórico y político. In: VIDAL, C. M.; FALGUERA, J. L.; SAGÜILLO, J. M.; VERDEJO, V. M.; PEREIRA-FARIÑA, M. **VII Conference of the Spanish Society for Logic, Methodology and Philosophy of Science**, Universidade de Santiago de Compostela: Santiago de Compostela, jul. 2013.

ROSA, A. M. da. **Guia do Processo Penal Conforme a Teoria dos Jogos**. 5. ed. Florianópolis: Emais, 2019.

RUSSELL, S. J.; NORVIG, P. (Eds.) **Artificial Intelligence: A Modern Approach**. 3rd ed. New Jersey: Pearson Education, Inc., 2010.

SADEK, M. T. A. Efetividade de direitos e acesso à justiça. In: RENAULT, S.; BOTTINI, P, (Orgs.). **Reforma do Judiciário**. 1. ed. São Paulo: Saraiva, 2005.

SADIN, E. La inteligencia artificial: el superyó del siglo XXI. **Revista Nueva Sociedad**, n. 279, jan./fev. 2019.

SALOMÃO, L. F. (Coord). **Inteligência Artificial: tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário brasileiro**. 2^a ed. Rio de Janeiro: FGV Conhecimento, 2022.

SALOMÃO, L. F.; LEME, E. (Coords). **Inteligência Artificial: tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário brasileiro**. 4^a ed. Rio de Janeiro: FGV Conhecimento, 2025.

SAMEK, W., MÜLLER, KR. Towards Explainable Artificial Intelligence. In: SAMEK, W., MONTAVON, G., VEDALDI, A., HANSEN, L., MÜLLER, KR. (Eds) **Explainable AI: Interpreting, Explaining and Visualizing Deep Learning**. Lecture Notes in Computer Science, vol. 11700. Springer, Cham, 2019. DOI: https://doi.org/10.1007/978-3-030-28954-6_1. Disponível em: https://link.springer.com/chapter/10.1007/978-3-030-28954-6_1. Acesso em: 12 fev. 2026.

SAMMUT, C.; WEBB, G. I. (Ed.). **Encyclopedia of Machine Learning and Data Mining**. New York: Springer, 2017.

SANTOS, A. C. dos. O uso da inteligência artificial pela suprema corte brasileira: desafios e potencialidades. **Revista Foco**, [s. l.], v. 18, n. 4, p. 1-23, 2025.

SANTOS, G. A. dos; SEVERI, F. C.; CAMPOS, G. C.; MOYSES, J. F.; CARVALHO, P. O. de; JOSÉ, T. da S.; SILVEIRA, T. B. H. O itinerário da criminalização do aborto no Brasil: uma reescrita a partir de perspectivas feministas interseccionais. In: SEVERI, F. C. (Org.). **Reescrevendo decisões judiciais em perspectivas feministas**. Ribeirão Preto: Universidade de São Paulo, Núcleo de Pesquisa e Extensão “O Direito Achado na Rua”, 2022.

SÃO PAULO. Tribunal de Justiça do Estado de São Paulo. Ação de obrigação de fazer nº 1001521-57.2017.8.26.0360. Requerente: Ministério Público do Estado de São Paulo; Requeridos: Janaina Aparecida Quirino; Município de Mococa. 2017.

SCAVONE, L. Estudos de gênero: uma sociologia feminista? **Revista de Estudos Feministas**, v. 16, n. 1, p. 173-186, 2008.

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

SCOTT, J. Gênero, uma categoria útil para a análise histórica. **Revista Educação e Realidade**. Porto Alegre, v. 20, n. 2, p. 71-99, 1995.

SEGUNDO, H. de B. M. TRIBUTAÇÃO E INTELIGÊNCIA ARTIFICIAL. **Revista Jurídica Luso Brasileira**. Publicação do Centro de Investigação de Direito Privado da Faculdade de Direito da Universidade de Lisboa (CIDP), v. 6, n. 1, p. 57-77, 2020. Disponível

em: https://www.cidp.pt/revistas/rjlb/2020/1/2020_01_0057_0077.pdf. Acesso em: 15 jun. 2021.

SEMENOVA, E.; PEREVOSHCHIKOVA, E.; IVANOV, A.; EROFEEV, M. **Fairness meets machine learning**: searching for a better balance. Moscow: National Research University Higher School of Economics, 2019. (Basic Research Program Working Papers, Series: Law, WP BRP 93/LAW/2019). Disponível em: <https://ssrn.com/abstract=3502708>. Acesso em: 27 out. 2025.

SEN, A. Well-being, Agency and Freedom: The Dewey Lectures 1984. **The Journal of Philosophy**, v. 82, n. 4, pp. 169-221, 1985.

SERASA EXPERIAN. Menos de 1% das mulheres trabalham com tecnologia no Brasil, indica pesquisa da Serasa Experian. **Serasa Experian**, 07 mar. 2024. Disponível em: <https://www.serasaexperian.com.br/sala-de-imprensa/rh/menos-de-1-das-mulheres-trabalham-com-tecnologia-no-brasil-indica-pesquisa-da-serasa-experian/#:~:text=Serasa%20Experian%3A%20Mulheres%20em%20TI%20representam%2026%2C5%25&text=%22O%20n%C3%BAmero%20de%200%2C07,diversidade%20de%20g%C3%AAnero%20neste%20setor>. Acesso em: 20 nov. 2024.

SEVERI, F. C. (Org.). **Reescrevendo decisões judiciais em perspectivas feministas**. Ribeirão Preto: Universidade de São Paulo, Núcleo de Pesquisa e Extensão “O Direito Achado na Rua”, 2022.

SHRESTHA, S.; DAS, S. Exploring gender biases in ML and AI academic research through systematic literature review. **Frontiers in Artificial Intelligence**, v. 5, out. 2022.

SILVA, F. F. da; RIBEIRO, P. R. C. Trajetória de mulheres na ciência: “ser cientista” e “ser mulher”. **Ciênc. Educ.**, v. 20, n. 2, Bauru, p. 449-466, 2014.

SILVA, J. M. **Feminismo na atualidade**: a formação da quarta onda. Recife: Independently published, 2019.

SILVA, T. T. da. A produção social da identidade e da diferença. In: SILVA, T. T. *et al.* (Orgs.). **Identidade e diferença**: A perspectiva dos estudos culturais. 4. ed. Petrópolis: Vozes, p. 73-102, 2000.

SIMIONI, F.; CARLOS, P. P. de; SILVA, V. R. da. “Qual o bem maior a ser protegido, o sofrimento temporal da gestante ou a vida do nascituro?”: Aborto em razão de malformação fetal grave. In: SEVERI, F. C. (Org.). **Reescrevendo decisões judiciais em perspectivas feministas**. Ribeirão Preto: Universidade de São Paulo, Núcleo de Pesquisa e Extensão “O Direito Achado na Rua”, 2022.

SIMMEL G. *apud* HOMEY, K. The Flight from Womanhood. **Women and Analysis**, ed. J. Strouse (New York: Dell, 1975), 1926.

SIMON, H. A. A Behavioral Model of Rational Choice. **Quartely Journal of Economics**, vol. 69, no. 1, pp. 99-118, fev. 1955.

SIMONITE, T. AI Is the Future - But Where Are the Women? Just 12 percent of machine learning researchers are women - a worrying statistic for a field supposedly reshaping society.

Wired, Business, 17 ago. 2018 7h:00. Disponível em: <https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/>. Acesso em: 23 set. 2024.

SIQUEIRA, D. P.; MENDES JUNIOR, F.; SANTOS, M. F. dos. Poder judiciário na era digital: o impacto das novas tecnologias de informação e de comunicação no exercício da jurisdição. **Revista Internacional Consinter De Direito**, v. 9, n. 17, 249, 2023. DOI: <https://doi.org/10.19135/revista.consinter.00017.10>. Disponível em: <https://revistaconsinter.com/index.php/ojs/article/view/554>. Acesso em: 12 fev. 2026.

SIQUEIRA, D. P.; MORAIS, F. S. de; SANTOS, M. F. dos. Inteligência artificial e jurisdição: dever analítico de fundamentação e os limites da substituição dos humanos por algoritmos no campo da tomada de decisão judicial. **Sequência: Estudos Jurídicos e Políticos**, Florianópolis, v. 43, n. 91, p. 1-34, 2022. DOI: <https://doi.org/10.5007/2177-7055.2022.e90662>. Disponível em: <https://periodicos.ufsc.br/index.php/sequencia/article/view/90662>. Acesso em: 12 fev. 2026.

SKEEM, J. L., LOWENKAMP, C. T. Risk, race, and recidivism: Predictive bias and disparate impact. **Criminology**, v. 54, n. 4, p. 680-712, 2016.

STEED, R.; CALISKAN, A. Image Representations Learned with Unsupervised Pretraining Contain Human-Like Biases. **FAccT '21**, mar. 3-10, 2021, Virtual Event, Canada. *ArXiv preprint*. Disponível em: <https://arxiv.org/pdf/2010.15052>. Acesso em: 21 nov. 2024.

STRECK, L. O (pós-)positivismo e os propalados modelos de juiz (Hércules, Júpiter e Hermes) – dois decálogos necessários. **Revista de direitos e garantias fundamentais**, n. 7, p. 15-45, 2010.

STRYKER, C. *O que é uma rede neural recorrente (RNN)?* **IBM**, [20-?]. Disponível em: <https://www.ibm.com/br-pt/think/topics/recurrent-neural-networks>. Acesso em: 1 nov. 2025.

TAFARELO, B. **Discriminação de gênero no julgamento de crimes de estupro: a violência institucional no discurso judicial**. 2023. 164 f. Dissertação (Mestrado Profissional em Direito) – Escola Nacional de Formação e Aperfeiçoamento de Magistrados (ENFAM), Brasília, DF, 2023.

Tang, Z.; Zhang, Y.; Zhang, C. What Is and How to Achieve Fairness in Machine Learning? **ACM Computing Surveys**, 2023.

TAVARES, J. P. L. G. Vieses implícitos e técnicas de automação decisória: riscos e benefícios. **Civil Procedure Review**, v. 12, n. 1, jan./abr., 2021. p. 109. Disponível em: <https://www.revistaannep.com.br/index.php/radp/article/view/44/pdf>. Acesso em: 25 abr. 2021.

THALER, R. H. **Comportamento Inadequado: a Construção da Economia Comportamental**. Lisboa: Actual, 2016.

THALER, Richard H.; SUNSTEIN, Cass R. **Nudge: Improving decisions about health, wealth, and happiness**. London: Penguin Books, 2008.

THOMSON, T. J.; THOMAS, R. J. Ageism, sexism, classism and more: 7 examples of bias in AI-generated images. **The Conversation**, 10 jul. 2023. Disponível em:

<https://theconversation.com/ageism-sexism-classism-and-more-7-examples-of-bias-in-ai-generated-images-208748> Acesso em: 21 nov. 2024.

TURING, A. M. Computing Machinery and Intelligence. **Mind, A Quarterly Review of Psychology and Philosophy**, vol. LIX, n. 236, p. 433-460, out. 1950.

TVERSKY, A.; KAHNEMAN, D. Judgment under uncertainty: Heuristics and biases. **Science**, v. 185, n. 4157, p. 1124-1131, 1974.

TVERSKY, A.; KAHNEMAN, D. The framing of decisions and the psychology of choice. **Science**, v. 211, 1981.

UNESCO INSTITUTE FOR STATISTICS. Women in Science. **Fact Sheet**, n. 55, jun. 2019. Disponível em: <https://uis.unesco.org/sites/default/files/documents/fs55-women-in-science-2019-en.pdf>. Acesso em: 17 out. 2024.

UNESCO. **Priority Gender Equality Action Plan 2014-2021**. Paris: UNESCO, 2014.

UNESCO. **Recomendação sobre a Ética da Inteligência Artificial**. Paris: UNESCO, 2022. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000381137_por. Acesso em: 18 abr. 2025.

UNESCO; IRCAI. **Challenging systematic prejudices: an Investigation into Gender Bias in Large Language Models**. UNESCO; IRCAI: Paris, 2024.

UNESCO; OECD; IDB. **Os efeitos da inteligência artificial na vida profissional das mulheres**. Paris, New York: UNESCO; OECD; IDB, 2023.

UNITED NATIONS DEVELOPMENT PROGRAMME. **2023 Gender Social Norms Index (GSNI): Breaking down gender biases: Shifting social norms towards gender equality**. New York, 2023.

UNIVERSIDADE ESTADUAL DE PONTA GROSSA. **Guia orientador: Reflexão sobre o uso da inteligência artificial na pesquisa acadêmica**. Versão 1.0. Ponta Grossa: UEPG, [2025?]. Disponível em: https://www2.uepg.br/profei/wp-content/uploads/sites/54/2025/09/Guia_IA-1.pdf. Acesso em: 18 dez. 2025.

VILLANI, C. SCHOENAUER, M.; BONNET, Y.; BERTHET, C.; CORNUT, A.; LEVIN, F.; RONDEPIERRE, B. Donner uns sens à li'intelligence artificielle: pour une stratégie nationale et européenne. **Vie publique: Au couer du débat public**. p. 140-142, mar. 2018. Disponível em: <https://www.vie-publique.fr/rapport/37225-donner-un-sens-lintelligence-artificielle-pour-une-strategie-nation>. Acesso em: 21 ago. 2025.

VON ENDE, L. B.; OLIVEIRA, R. S. de. Desigualdade de gênero e tecnologia: entre mulheres e algoritmos. **Revista Brasileira de Iniciação Científica (RBIC)**, Itapetininga, v. 7, n. 6, p. 210-219, 2020

WAJCMAN, J. **Feminism confronts technology**. The Pennsylvania State University Press, United States, 1991.

WAJCMAN, J. Feminist Theories of Technology. **Cambridge Journal of Economics**, v. 34, n. 1, pp. 143-152, 2010.

WAJCMAN, J. **TechnoFeminism**. Cambridge, MA: Polity Press, 2004.

WAJCMAN, J.; FRADE, R. Feminism and Technology: an interview with Judy Wajcman by Renata Frade. In: FRADE, R.; VAIRINHOS, M. **Technofeminism: multi and transdisciplinary contemporary views on woman in technology**. UA Editora: Aveiro, 2023.

WAJCMAN, J.; YOUNG, E.; FITZMAURICE, A. **The Digital Revolution: Implications for**

WITZEL, D. G. **Práticas discursivas, redes de memória e identidades do feminino: entre princesas, bruxas e lobos no universo publicitário**. 2011. 217 f. Tese de Doutorado – UNESP, Araraquara - SP, 2011. Disponível em: <https://repositorio.unesp.br/handle/11449/103558>. Acesso em: 09 set. 2022.

WOJCIECHOWSKI, P. B.; ROSA, A. M. da. **Vieses da justiça: como as heurísticas e vieses operam nas decisões penais e a atuação contraintuitiva**. Florianópolis: Empório Modara, 2018.

WOLKART, E. N. **Análise econômica do processo civil: como promover a cooperação para enfrentar a tragédia da Justiça no processo civil brasileiro**. Tese (Doutorado). Universidade do Estado do Rio de Janeiro, Faculdade de Direito. Orientador: Paulo Cezar Pinheiro Carneiro. 801f. Universidade do Estado do Rio de Janeiro – UERJ, Rio de Janeiro, 2018.

WOODWARD, K. Identidade e diferença: uma introdução teórica e conceitual. In: SILVA, T. T. *et al.* (Orgs.). **Identidade e diferença: A perspectiva dos estudos culturais**. 4. ed. Petrópolis: Vozes, p. 7-72, 2000.

YARGER, L.; COBB PAYTON, F.; NEUPANE, B. Algorithmic equity in the hiring of underrepresented IT job candidates. **Online Information Review**, v. 44, n. 2, p. 383-395, 2019

YIN, R. K. **Estudo de caso: planejamento e métodos**. 2. ed. Porto Alegre: Bookman, 2001.

YOUNG, E.; WAJCMAN, J.; SPREJER, L. Where are the Women? Mapping the Gender Job Gap in AI. Policy Briefing: Full Report. **The Alan Turing Institute**, 2021.

ZAFFARONI, E. R. **Estructuras judiciales**. Buenos Aires: Ediar, 1994.

ZIOSI, M., WATSON, D.; FLORIDI, L. A Genealogical Approach to Algorithmic Bias. **Minds & Machines**, v. 34, n. 9, 2024. DOI: <https://doi.org/10.1007/s11023-024-09672-2>. Disponível em: <https://link.springer.com/article/10.1007/s11023-024-09672-2>. Acesso em: 12 fev. 2026.

ZOU, J.; SCHIEBINGER, L. AI can be sexist and racist – it’s time to make it fair. **Nature**, 559, 324-326, 2018.

APÊNDICE A – PROTOCOLO ÉTICO PARA DESENVOLVEDORES E EQUIPES TÉCNICAS DOS SISTEMAS INTELIGENTES JUDICIAIS

Objetivo: assegurar que as soluções de IA empregadas no Poder Judiciário sigam as garantias éticas – justiça algorítmica, transparência, explicabilidade, auditabilidade, responsabilidade e equidade de gênero.

O protocolo foi estruturado com o auxílio do ChatGPT com base no Capítulo 5 da tese e tem por base orientadora o “Guia orientador: Reflexão sobre o uso da inteligência artificial na pesquisa acadêmica” (UNIVERSIDADE ESTADUAL DE PONTA GROSSA, [2025?]).

Planejamento ético do sistema

Pergunta de Verificação	Exemplo de Avaliação	Evidências / Observações
O sistema impacta direitos fundamentais (ex.: liberdade, dignidade, igualdade)?	Exemplo: modelos que sugerem medidas cautelares ou risco de reincidência requerem validação ética.	☐ Sim ☐ Não
Foi elaborado um Relatório de Impacto Algorítmico (RIA)?	Verificar se o RIA inclui potenciais efeitos discriminatórios e medidas de mitigação.	☐ Sim ☐ Não
Há comitê multidisciplinar acompanhando o desenvolvimento?	Exigir ata e composição do comitê com diversidade de gênero e áreas de conhecimento.	☐ Sim ☐ Não
Foram considerados cenários de exclusão digital?	Avaliar se grupos com acesso precário à tecnologia são representados nos testes.	☐ Sim ☐ Não

Curadoria e governança de dados

Pergunta de Verificação	Exemplo de Avaliação	Evidências / Observações
As bases refletem diversidade populacional (raça, gênero, região, classe)?	Exemplo: incluir dados de diferentes tribunais e regiões para reduzir enviesamento.	☐ Sim ☐ Não
Foram identificadas e removidas variáveis correlacionadas a gênero?	Exemplo: remover atributos como nome, profissão ou estado civil.	☐ Sim ☐ Não
Existe documentação de data lineage com registro da origem dos dados?	Garantir rastreabilidade e documentação transparente.	☐ Sim ☐ Não

Foram realizados testes de vies nos dados?	Exemplo: analisar distribuição de sentenças por gênero ou raça.	☐ Sim ☐ Não
---	---	---

Desenvolvimento e testes

Pergunta de Verificação	Exemplo de Avaliação	Evidências / Observações
Foram aplicadas métricas de fairness (Demographic Parity, Equal Opportunity, Equalized Odds, Predictive Parity)?	Verificar se o modelo é equitativo entre grupos sensíveis.	☐ Sim ☐ Não
O modelo fornece explicações locais (LIME, SHAP)?	Avaliar se há clareza sobre variáveis que influenciam a decisão.	☐ Sim ☐ Não
Há auditoria independente das decisões automatizadas?	Solicitar relatórios de conformidade externos.	☐ Sim ☐ Não
Os resultados foram testados por amostragem cega?	Exemplo: avaliadores sem acesso a dados sensíveis verificam consistência dos resultados.	☐ Sim ☐ Não

Implementação e monitoramento

Pergunta de Verificação	Exemplo de Avaliação	Evidências / Observações
Há documentação pública explicando objetivos e limites do sistema?	Publicar cartilha técnica ou sumário executivo de transparência.	☐ Sim ☐ Não
O sistema gera logs rastreáveis e versionamento de modelos?	Permitir auditoria futura de decisões e versões do modelo.	☐ Sim ☐ Não
O modelo é reavaliado periodicamente quanto à performance e vies?	Revisões semestrais ou anuais, conforme Resolução do CNJ n. 615/2025.	☐ Sim ☐ Não
Existe canal de denúncia ética ativo?	Canal no site institucional para relato de vieses e falhas.	☐ Sim ☐ Não

Responsabilidade e Prestação de Contas

Pergunta de Verificação	Exemplo de Avaliação	Evidências / Observações
O responsável técnico está formalmente identificado?	Nome e função visíveis no sistema e em relatórios técnicos.	☐ Sim ☐ Não
Há plano de contingência para falhas éticas ou violações de direitos?	Suspensão imediata e revisão extraordinária prevista em norma interna.	☐ Sim ☐ Não

O sistema permite rastrear versão e parâmetros do modelo usado em cada decisão?	Garantir integridade e auditabilidade dos resultados.	☐ Sim ☐ Não
Foram testadas as condições de acessibilidade e neutralidade de gênero na interface?	Evitar linguagem ou <i>design</i> com estereótipos de gênero.	☐ Sim ☐ Não

APÊNDICE B – PROTOCOLO ÉTICO PARA MAGISTRADOS, ASSESSORIAS E DEMAIS USUÁRIOS DOS SISTEMAS INTELIGENTES

Objetivo: garantir que o uso de IA nas decisões judiciais preserve as garantias constitucionais de igualdade, fundamentação, devido processo e julgamento com perspectiva de gênero.

O protocolo foi estruturado com o auxílio do ChatGPT com base no Capítulo 5 da tese e tem por base orientadora o “Guia orientador: Reflexão sobre o uso da inteligência artificial na pesquisa acadêmica” (UNIVERSIDADE ESTADUAL DE PONTA GROSSA, [2025?]).

Compreensão e consentimento informado

Pergunta de Verificação	Exemplo de Avaliação	Evidências / Observações
O magistrado compreende a função e os limites da ferramenta de IA?	Exemplo: IA recomenda precedentes, mas não determina o resultado.	☐ Sim ☐ Não
O relatório da IA explica claramente os critérios utilizados?	Deve listar variáveis relevantes como tempo de pena ou tipo penal.	☐ Sim ☐ Não
Há aviso nos autos sobre o uso de IA?	Incluir declaração visível sobre uso de ferramentas automatizadas.	☐ Sim ☐ Não
A decisão final é humana e fundamentada?	O juiz deve justificar o uso ou rejeição do parecer da IA.	☐ Sim ☐ Não

Uso Responsável e transparente

Pergunta de Verificação	Exemplo de Avaliação	Evidências / Observações
O parecer da IA é anexado integralmente ao processo?	Deve ser acessível às partes e à defesa.	☐ Sim ☐ Não
A decisão judicial menciona o uso da IA e sua influência?	Evita opacidade e reforça <i>accountability</i> .	☐ Sim ☐ Não
Há consistência entre o parecer da IA e a jurisprudência aplicável?	Divergências devem ser justificadas em sentença.	☐ Sim ☐ Não
O tribunal oferece acesso aos logs e versões do sistema?	Garante transparência e possibilidade de auditoria.	☐ Sim ☐ Não

Avaliação de Equidade e Impacto

Pergunta de Verificação	Exemplo de Avaliação	Evidências / Observações
-------------------------	----------------------	--------------------------

Há diferença nas recomendações da IA entre gêneros, raças ou classes?	Exemplo: IA sugere liberdade provisória mais para homens do que mulheres.	☐ Sim ☐ Não
Foi aberto relatório de não conformidade se detectado viés?	Relatar formalmente e enviar ao comitê técnico.	☐ Sim ☐ Não
Existe feedback ao desenvolvedor ou CNJ?	Assegurar ciclo de correção contínua.	☐ Sim ☐ Não
Há monitoramento estatístico de equidade das decisões?	Comparar taxas de erro entre grupos.	☐ Sim ☐ Não

Julgamento com perspectiva de gênero

Pergunta de Verificação	Exemplo de Avaliação	Evidências / Observações
O magistrado considera efeitos diferenciados das decisões sobre homens e mulheres?	Exemplo: impacto da prisão preventiva em mães chefes de família.	☐ Sim ☐ Não
O sistema incorpora diretrizes de julgamento com perspectiva de gênero?	Checar se bases incluem casos de violência contra a mulher.	☐ Sim ☐ Não
O parecer da IA foi revisado quanto à linguagem inclusiva e não discriminatória?	Evitar termos que reforcem estereótipos.	☐ Sim ☐ Não
Há revisão humana obrigatória para casos sensíveis a gênero?	Revisão adicional por magistrada ou comitê de gênero.	☐ Sim ☐ Não

Supervisão e educação ética

Pergunta de Verificação	Exemplo de Avaliação	Evidências / Observações
O tribunal mantém comitê de ética em IA com participação feminina?	Promover diversidade e interdisciplinaridade.	☐ Sim ☐ Não
Há capacitação contínua sobre letramento digital e equidade?	Registrar frequência e certificação.	☐ Sim ☐ Não
A equipe tem acesso a guias explicativos de fairness e viés (CNJ/ENFAM)?	Assegurar base comum de entendimento ético.	☐ Sim ☐ Não
O sistema é revisado à luz de novos parâmetros éticos (OCDE, UNESCO, CNJ)?	Ajustar critérios de avaliação conforme evolução normativa.	☐ Sim ☐ Não