ENTRE ALGORITMOS E DIREITOS FUNDAMENTAIS, RISCOS DE VIESES ALGORÍTMICOS NA AUTOMAÇÃO DAS DECISÕES JUDICIAIS

BETWEEN ALGORITHMS AND FUNDAMENTAL RIGHTS: RISKS OF ALGORITHMIC BIAS IN THE AUTOMATION OF COURT DECISIONS

Ludmila da Silva Vianna[1]

RESUMO

O crescimento exponencial da litigiosidade no Brasil, com cerca de 80 milhões de processos em tramitação segundo o Relatório Justiça em Números (CNJ,2025), tem impulsionado o uso de sistemas de Inteligência Artificial (IA) no Poder Judiciário como ferramenta estratégica para enfrentar a morosidade processual e otimizar fluxos decisórios. As aplicações são para auxiliar na triagem, classificação e apoio redacional. Sistemas como Victor (STF), Sinapses (CNJ) e GAIA (TJRS) já vêm sendo aplicados para classificação de demandas, análise de precedentes e elaboração de minutas de decisões, promovendo ganhos significativos de eficiência. Contudo, a crescente automação das decisões judiciais suscita preocupações quanto à segurança jurídica e à proteção de direitos fundamentais, em especial diante dos riscos de vieses algorítmicos. O estudo analisa os riscos de vieses algorítmicos nessas aplicações e seus impactos sobre direitos fundamentais (igualdade, devido processo, motivação e publicidade). Com abordagem qualitativa, de natureza exploratória e descritiva, fundamentada em revisão bibliográfica, análise documental e levantamento jurisprudencial, o estudo (i) sistematiza fontes de viés ao longo do ciclo de vida dos sistemas, (ii) confronta tais riscos com o marco constitucional, a LGPD, o CPC e diretrizes do CNJ, e (iii) propõe um modelo de governança (framework) com documentação fichas técnicas de dados (*datasheets*) e fichas técnicas de modelos (*model cards*), avaliação de impacto, métricas de fairness, fundamentação qualificada, contraditório técnico, monitoramento e auditorias. Conclui-se que ganhos de eficiência são compatíveis com as garantias processuais quando há controle humano significativo, transparência e prestação de contas. Os resultados preliminares indicam que, embora a IA traga benefícios, é urgente a definição de diretrizes éticas, regulamentações robustas e mecanismos de auditoria algorítmica que assegurem a transparência e a equidade.

Palavras-chave: Inteligência Artificial; poder judiciário; vieses algorítmicos; direitos fundamentais; governança tecnológica.

[1] Pós Graduação, Fundação do Ministério Público do Rio Grande do Sul, Ludmilavianna@yahoo.com.br

ABSTRACT

Brazil's exponential growth in litigation with roughly 80 million pending cases, according to the Justiça em Números report (CNJ, 2025, has driven the adoption of Artificial Intelligence (AI) systems in the Judiciary as a strategic tool to address procedural delays and optimize decision-making workflows. Current applications support case screening, classification, and drafting assistance. Systems such as Victor (STF), Sinapses (CNJ), and GAIA (TJRS) are already being used for classifying claims, analyzing precedents, and preparing draft decisions, yielding significant efficiency gains. However, the increasing automation of judicial decision-making raises concerns about legal certainty and the protection of fundamental rights, particularly in view of the risks of algorithmic bias. This article analyzes the risks of algorithmic bias in these applications and their impacts on fundamental rights (equality, due process, reason-giving, and publicity). Using a qualitative, exploratory, and descriptive approach—grounded in literature review, document analysis, and a survey of case law—the study (i) systematizes sources of bias across the systems' lifecycle, (ii) confronts these risks with the constitutional framework, the LGPD, the CPC, and CNJ guidelines, and (iii) proposes a governance model featuring technical documentation (datasheets/model cards), impact assessment, fairness metrics, enhanced reason-giving, technical contestability, monitoring, and audits. We conclude that efficiency gains are compatible with procedural guarantees when there is meaningful human control, transparency, and accountability. Preliminary findings indicate that, while AI brings benefits, there is an urgent need for ethical guidelines, robust regulation, and algorithmic audit mechanisms to ensure transparency and fairness.

Keywords: Artificial Intelligence; Judiciary; Algorithmic Bias; Fundamental Rights; Technological Governance.

1. INTRODUÇÃO

O Poder Judiciário brasileiro enfrenta um cenário de complexidade crescente, marcado pela sobrecarga processual e pela necessidade premente de modernização da prestação jurisdicional. De acordo com o Relatório Justiça em números, extraído do Painel de Estatísticas do Poder Judiciário (CNJ, 2025), tramitam atualmente mais de 80 milhões de processos, sendo que, apenas em 2025, foram ajuizadas cerca de 19 milhões de novas demandas. Tal realidade impõe desafios estruturais ao sistema e exige a adoção de soluções inovadoras capazes de promover celeridade e efetividade decisória.

Nesse contexto, a Inteligência Artificial (IA) vem sendo incorporada aos fluxos internos dos tribunais para automatizar tarefas repetitivas, organizar dados processuais e auxiliar magistrados na formação de precedentes. Iniciativas como o Victor (Supremo Tribunal Federal), o Sinapses (Conselho Nacional de Justiça) e o GAIA (Tribunal de Justiça do Rio Grande do Sul) ilustram essa transformação, oferecendo ferramentas de triagem de processos, classificação de demandas repetitivas e apoio à elaboração de minutas. Apesar dos ganhos potenciais de eficiência e padronização, a expansão do apoio automatizado à decisão reaviva um problema central: os vieses algorítmicos e seus impactos sobre direitos fundamentais e sobre a segurança jurídica.

A hipótese que orienta a pesquisa sustenta que, embora a IA possa reduzir a morosidade e favorecer a coerência jurisprudencial, modelos opacos treinados com bases enviesadas tendem a reproduzir discriminações históricas e a gerar decisões desproporcionais ou injustas, comprometendo princípios constitucionais como isonomia, ampla defesa, contraditório, motivação e devido processo legal. Diante disso, coloca-se a seguinte **pergunta de pesquisa**: *em que medida a automação (parcial) de decisões judiciais, tal como hoje praticada no Brasil, é compatível com as garantias fundamentais e quais salvaguardas são necessárias para prevenir e mitigar vieses algorítmicos?* O objetivo geral deste artigo é analisar criticamente os riscos decorrentes dos vieses algorítmicos na automação de decisões judiciais e seus impactos sobre direitos fundamentais, propondo diretrizes de governança tecnológica, auditoria algorítmica e parâmetros ético-regulatórios para o uso seguro de IA no sistema de Justiça. Especificamente, busca-se: (i) sistematizar fontes e tipos de viés relevantes ao contexto judicial; (ii) confrontar tais riscos com a Constituição e com marcos normativos e jurisprudenciais aplicáveis; e (iii) propor medidas operacionais de transparência, prestação de contas e contestabilidade. Metodologicamente, adota-se abordagem qualitativa, de natureza exploratória e descritiva, fundada em revisão bibliográfica, análise documental e levantamento jurisprudencial. São examinadas normas nacionais, relatórios e documentos oficiais de CNJ, STF e TJRS, além de literatura acadêmica e estudos comparados internacionais. O escopo concentra-se em ferramentas de apoio à decisão e priorização/triagem, não abrangendo automação

plena de sentenças criminais ou civis, salvo quando empregadas como contraste conceitual.

2. FUNDAMENTOS TÉCNICOS DE VIÉS ALGORÍTMICO E SUA TRADUÇÃO PARA O VOCABULÁRIO JURÍDICO

2.1. O que é "viés" em sistemas de IA e seu surgimento

Chamamos de viés algorítmico a distorção sistemática produzida por sistemas de IA que afeta de modo desigual pessoas, grupos ou tipos de casos. Diferentemente do erro aleatório, o viés tem padrão: nasce de dados históricos incompletos ou seletivos, de escolhas de modelagem e também da forma como a ferramenta é usada.

Do ponto de vista jurídico, ele pode se manifestar como: Discriminação direta: uso explícito de atributo sensível (p. ex., raça), discriminação indireta: uso de proxies (CEP, escolaridade, tipo de pedido) que geram impacto desproporcional, ainda que o atributo sensível não seja utilizado.

Para fins práticos, o viés tende a emergir em três blocos do ciclo de vida da solução. Primeiro, nos dados deentrada, quando há bases desbalanceadas, que sub-representam determinadas regiões, varas ou assuntos, rótulos produzidos no passado segundo critérios inconsistentes e o emprego de variáveis que funcionam como *proxies* de atributos sensíveis.

Segundo, no modelo de processamento, quando os objetivos de otimização privilegiam o desempenho "médio" e desconsideram custos jurídicos de erro, por exemplo, rotular como "repetitiva" uma ação que não o é, quando escolhas de parametrização prejudicam subgrupos raros e quando se perde calibração, de modo que um mesmo *score* deixa de corresponder à mesma probabilidade entre grupos distintos.

Terceiro, no uso institucional (saída), quando se fixam limiares uniformes para contextos heterogêneos, quando se verifica tendência de sobreconfiança na automação (*automation bias*), quando se instauram ciclos de retroalimentação (*feedback loops*, em que decisões presentes moldam os dados futuros) e quando

ocorre deriva (*drift*), isto é, mudanças de teses ou precedentes que tornam o modelo desatualizado.

2.2. Como avaliar "justiça" do sistema sem jargões

Em vez de um arsenal de métricas, três perguntas bastam para orientar o juiz: A IA de apoio, trata de forma equilibrada? As taxas de acerto/erro são semelhantes entre grupos relevantes (por vara, região, tipo de parte)? Se um grupo erra bem mais, há indício de impacto desproporcional.

Fala a mesma língua para todos? O mesmo *score* tem o mesmo significado (calibração) entre grupos e ao longo do tempo? Se não, o score não é comparável e a decisão pode se tornar arbitrária.

Não exclui sistematicamente? O uso do sistema não empurra certos casos para o fim da fila nem os marca indevidamente como "repetitivos" ou "improcedentes"? Aqui importa verificar efeitos acumulados na duração do processo e no acesso a precedentes.

No processo, falsos positivos e falsos negativos têm custos diferentes. A escolha do limiar de uso deve refletir esses custos jurídicos (isonomia, contraditório, motivação), não apenas "acurácia média".

2.3. Erros, custos e assimetrias

Em ambientes judiciais, erros não possuem o mesmo peso. Falsos positivos na triagem (p. ex., classificar caso como "repetitivo" quando não é) podem cercear contraditório; falsos negativos podem aumentar morosidade. Em contextos criminais, a assimetria é ainda mais aguda, causando impactos sobre liberdade do réu e danos irreparáveis à sua imagem. Logo, a escolha de limiares e métricas deve refletir os riscos jurídicos dos erros, não apenas métricas técnicas (accuracy, AUC).

Dados: curadoria e documentação (origem, recorte, lacunas), balanceamento e correção de rótulos; manter atributos sensíveis apenas para auditar impacto.

Modelo: metas de desempenho estratificado (por vara/tema/grupo), restrições simples de equidade quando cabíveis, verificação de calibração por grupo.

Operação: revisão humana obrigatória para baixa confiança/alto impacto; registro de superações da recomendação algorítmica (*overrides*) e motivos; limiares revisáveis.
Governança: datasheet (dados), model card (modelo), logs, avaliação de impacto algorítmico antes da adoção, monitoramento de drift e relatório periódico.

2.4. Explicabilidade, Fundamentação e contestabilidade

Motivação (CPC, art. 489, §1º): explicar se, como e quanto ao resultado algorítmico influenciou o convencimento, e por que é juridicamente pertinente; enfrentar argumentos relevantes das partes.
Contraditório técnico (arts. 9º e 10 do CPC): franquear às partes informações proporcionais ao risco (finalidade, dados usados em linhas gerais, versão e desempenho/impacto); prever rito de contestação e revisão humana.
Publicidade e transparência: documentação mínima acessível (finalidade, limites, métricas resumidas, versões) preservando segredos legais e dados pessoais.
Proporcionalidade: demonstrar idoneidade, necessidade e proporcionalidade em sentido estrito do uso da ferramenta no caso concreto.
Proteção de dados: observar princípios de finalidade, necessidade, não discriminação e responsabilização ao longo de todo o ciclo

2.5. Tradução para categorias jurídicas

Discriminação direta vs. indireta: uso explícito de atributos sensíveis (direta) ou uso de proxies que produzem impacto desproporcional (indireta). Métricas como paridade demográfica e *equalized odds* auxiliam a detectar impacto; logs e documentação ajudam a apurar intencionalidade.
Teste estatístico de impacto: diferenças sistemáticas em taxas de erro e calibração por grupo podem sinalizar violação a princípios de igualdade e não discriminação.
Devido processo, contraditório e motivação: exigem rastreabilidade ao longo do ciclo dados–modelo–decisão, direito a razões compreensíveis e canais de revisão humana.
Publicidade e transparência: publicidade ativa de documentação técnica (limites, dados de treinamento, métricas de desempenho/equidade) e de relatórios periódicos de auditoria.

Proporcionalidade: a adoção de IA deve ser idônea (evidência de eficácia), necessária (não há meio menos restritivo) e proporcional em sentido estrito (benefícios superam riscos residuais sob salvaguardas).

Proteção de dados: princípios de finalidade, adequação, necessidade, qualidade, transparência, segurança e não discriminação devem orientar todo o ciclo (da coleta ao descarte). O uso de dados sensíveis requer base legal, minimização e controles reforçados.

Accountability é a prestação de contas e exige responsabilização institucional de quem decide, quem audita, quem corrige, com papéis definidos, registros-logs e trilhas de decisão.

2.6. Nota sobre o contexto brasileiro

Viés algorítmico é distorção sistemática que pode surgir nos dados, no modelo e no uso. Sua avaliação, para fins jurídicos, se resume a verificar equilíbrio, calibração e não exclusão sistemática, sempre considerando os custos jurídicos dos erros. A resposta institucional combina medidas técnicas simples (documentação, estratificação, revisão humana) e deveres processuais (motivação qualificada, contraditório, transparência, proteção de dados)

A tradução técnica-jurídica deve dialogar com garantias constitucionais (igualdade, devido processo, publicidade, motivação) e com os princípios de proteção de dados (finalidade, necessidade, transparência, prevenção, não discriminação e responsabilização).

No setor público, a adoção de IA demanda ainda observância a princípios da administração pública (legalidade, impessoalidade, moralidade, publicidade, eficiência), com ênfase em registro, motivação e revisão. Em suma: métricas e técnicas de equidade algorítmica só ganham validade jurídica quando acompanhadas de processos institucionais que assegurem contestabilidade, auditabilidade e prestação de contas.

3. MARCO CONSTITUCIONAL, LEGAL E JURISPRUDENCIAL APLICÁVEL

3.1. Parâmetros constitucionais estruturantes

A adoção de sistemas de IA no âmbito do Judiciário deve conformar-se aos parâmetros constitucionais que estruturam o processo decisório. Em primeiro lugar, os princípios da igualdade e da não discriminação (art. 5º, caput) vedam resultados que produzam impacto desproporcional entre grupos, ainda que ausente intenção discriminatória explícita, impondo ao desenho e ao uso das ferramentas controles capazes de prevenir efeitos indiretos e seletivos.

Em segundo lugar, o devido processo legal, o contraditório e a ampla defesa (arts. 5º, LIV e LV) exigem que as partes tenham acesso às razões da decisão, inclusive quando informadas por saídas algorítmicas, e possibilidade real de contestar os fundamentos técnicos, influenciando o resultado por meio de participação efetiva.

No mesmo sentido, os deveres de publicidade e de motivação das decisões (art. 93, IX, e art. 5º, LX) impõem transparência quanto aos critérios empregados e motivação suficiente, também quando houver apoio em ferramentas automatizadas, com a explicitação do peso atribuído ao resultado do sistema e de sua pertinência jurídica ao caso concreto. A cláusula da duração razoável do processo (art. 5º, LXXVIII) legitima a busca por eficiência, mas não autoriza a redução de garantias processuais: eficiência é meio, não fim, devendo ser compatibilizada com a integridade das salvaguardas.

Além disso, os princípios da Administração Pública (art. 37, caput), como legalidade, impessoalidade, moralidade, publicidade e eficiência, orientam o desenho institucional das soluções de IA, reclamando bases normativas claras, documentação e registros auditáveis, segregação de funções, supervisão e prestação de contas.

Por fim, a proteção de dados pessoais como direito fundamental (EC 115/2022) condiciona o tratamento de dados, inclusive sensíveis, em sistemas algorítmicos utilizados pelo poder público, impondo observância a finalidade, necessidade, proporcionalidade e gestão ao longo de todo o ciclo de vida da solução. Em conjunto, esses vetores configuram o marco constitucional de compatibilidade ao qual toda iniciativa de IA judicial deve se submeter.

3.2. LGPD e setor público: Os limites e deveres

A Lei Geral de Proteção de Dados (Lei nº 13.709/2018) estabelece, no art. 6º, princípios que devem orientar todo o ciclo de vida de soluções de IA no Judiciário como: finalidade, adequação, necessidade, livre acesso, qualidade dos dados, transparência, segurança, prevenção, não discriminação e responsabilização, impondo que o tratamento seja desenhado desde a origem para atender a propósitos determinados, com minimização de dados, garantias de integridade e mecanismos de explicabilidade proporcionais ao risco.

No setor público, a base legal para o tratamento pode decorrer da execução de políticas públicas (art. 7º, III, c/c art. 23) ou do cumprimento de obrigação legal (art. 7º, II), hipóteses que não dispensam a observância dos princípios mencionados nem a documentação das escolhas técnicas e jurídicas. O tratamento de dados sensíveis (art. 11) demanda salvaguardas reforçadas, devendo ser estritamente necessário, tecnicamente justificado e devidamente registrado quanto à finalidade, à governança e aos controles de acesso. Quanto a decisões automatizadas (art. 20), assegura-se ao titular o direito a informações claras sobre os critérios e procedimentos empregados e a revisão por pessoa natural quando a decisão for exclusivamente automatizada e afetar seus interesses.

No âmbito judicial, ainda que a decisão não seja exclusiva da máquina, esse regime normativo exige, ao menos, motivação reforçada explicitando se, como e em que medida o resultado algorítmico foi considerado, e a existência de canal de revisão apto a viabilizar o contraditório técnico e a correção de eventuais distorções.

3.3. Fundamentação qualificada e transparência (CPC e LAI)

O Código de Processo Civil densifica o dever constitucional de motivação e estrutura garantias que se aplicam também ao uso de IA no processo decisório. Pelo art. 489, §1º, veda-se a fundamentação genérica ou por remissão acrítica, impondo o enfrentamento dos argumentos relevantes e dos precedentes invocados; o art. 371 exige valoração racional das provas, com indicação dos motivos do convencimento; e os arts. 9º e 10 asseguram contraditório efetivo e proíbem a decisão-surpresa, garantindo às partes oportunidade de se manifestarem sobre métodos e saídas algorítmicas relevantes ao caso. O art. 6º (cooperação) recomenda desenho

procedimental que viabilize a participação das partes, inclusive na depuração de vieses e na definição de métricas de desempenho e de equidade, quando pertinentes. Em complemento, a Lei de Acesso à Informação (Lei 12.527/2011) reforça a publicidade ativa: soluções de IA devem ter finalidade, bases de dados, versões de modelos, métricas (de desempenho e de equidade) e limitações devidamente documentadas; exigir relatórios periódicos de desempenho e de auditoria; e assegurar acesso ampliado a partes e peritos, sob controle judicial. Tais deveres observam, contudo, a proteção de segredos industriais e de dados pessoais, de modo a compatibilizar transparência e auditabilidade com as salvaguardas legais aplicáveis.

3.4. Diretrizes do CNJ e orientações jurisprudenciais

A orientação do Conselho Nacional de Justiça (CNJ) conforma um eixo ético-operacional para o uso de IA no Judiciário, assentado em governança, transparência, auditabilidade, responsabilização institucional e revisão humana efetiva. Iniciativas como Victor, Sinapses e GAIA materializam esse vetor ao padronizar documentação técnica mínima como as fichas técnicas de dados e fichas técnicas de modelos. O registro de decisões assistidas, e rotinas de monitoramento e melhoria contínua, além de prever avaliação de impacto algorítmico e canais de contestação. Tais diretrizes convergem com a LGPD e com os princípios constitucionais, servindo de parâmetro operativo para implantação e supervisão dessas ferramentas.

Em paralelo, a jurisprudência constitucional e infraconstitucional tem densificado requisitos para o emprego de tecnologias no processo decisório: (i) fundamentação suficiente e contextual, vedadas fórmulas vazias, com explicitação — quando houver — da relevância jurídica do resultado algorítmico; (ii) motivação *per relationem* qualificada, admitida a remissão a pareceres, laudos ou relatórios (inclusive de IA) desde que o julgador demonstre compreensão crítica e adaptação ao caso concreto; (iii) contraditório técnico, assegurando acesso e impugnação dos elementos informacionais determinantes (dados, premissas, métricas e relatórios de desempenho), preservados segredos legalmente protegidos; e (iv) observância do princípio da proporcionalidade (idoneidade, necessidade e proporcionalidade em

sentido estrito), com sopesamento entre ganhos de eficiência e riscos residuais — como viés, opacidade e erro sistêmico.

Em conjunto, diretrizes administrativas e vetores jurisprudenciais delineiam um padrão mínimo de compatibilidade para a adoção de IA no Judiciário, no qual a tecnologia atua como meio auxiliar submetido a controle humano, transparência e prestação de contas.

3.5. Teste de compatibilidade para ferramentas de apoio decisório

Propõe-se um *checklist* normativo-operacional para aferir a conformidade de sistemas de IA usados no processo judicial.

Para aferir a conformidade de sistemas de apoio automatizado à decisão no âmbito do Judiciário, impõe-se, em primeiro lugar, a explicitação da finalidade e da base normativa que os legitimam (CF/88, CPC e LGPD), com a identificação clara do caso de uso. Em seguida, deve-se assegurar revisão humana, mediante a definição de papéis, de quem utiliza, quem revisa e quem audita, e a delimitação do escopo funcional, inclusive quanto ao que o sistema não faz.

A transparência requer documentação pública proporcional ao risco, contendo finalidade, origem e qualidade dos dados, versões do modelo, métricas de desempenho e de equidade, bem como limitações conhecidas. Antes da implementação, é indispensável a avaliação de impacto algorítmico, com análise prévia de vieses, apresentação de resultados e incorporação dos planos de mitigação ao processo decisório.

No plano decisório, exige-se motivação reforçada, explicitando como o resultado algorítmico foi considerado, por que é juridicamente relevante ao caso concreto e de que modo foram enfrentados eventuais argumentos contrários. Importante garantir contraditório e contestabilidade, por meio de protocolo que assegure às partes acesso às informações essenciais, prazos e procedimentos de revisão. A proteção de dados deve observar minimização, segurança, governança de dados sensíveis e registro de acessos, com indicação clara das bases legais aplicáveis.

Em operação, impõe-se monitoramento e auditoria contínuos, com manutenção de logs, acompanhamento de indicadores de *drift*, rotinas de recalibração e auditorias internas e externas em periodicidade definida. Nos casos de conflito entre métricas, requer-se justificativa jurídico-substancial para a métrica de equidade adotada, considerando os custos do erro e o bem jurídico tutelado.

A prestação de contas demanda a identificação da autoridade responsável por aprovar, supervisionar e, se necessário, descontinuar o uso do sistema.

3.6. Síntese

O marco constitucional (igualdade, devido processo, motivação e publicidade), a LGPD (princípios, direitos e revisão), o CPC (fundamentação qualificada e contraditório) e as diretrizes do CNJ convergem para um modelo no qual a IA é meio auxiliar, submetido a revisão humana, transparência e auditabilidade. A compatibilidade jurídica depende menos da "caixa-preta" tecnológica e mais de processos institucionais que assegurem motivação suficiente, participação das partes e governança. Na Seção 4, aplicaremos esses parâmetros a casos-tipo e mapearemos riscos e salvaguardas operacionais no contexto brasileiro.

4. MODELO DE GOVERNANÇA, AUDITORIA E PRESTAÇÃO DE CONTAS

4.1. Arquitetura de governança e gradação de risco

O modelo de governança traduz CF/88, CPC e LGPD em rotinas verificáveis orientadas por: (i) proteção de direitos fundamentais; (ii) eficiência com controle humano significativo; (iii) transparência e auditabilidade; (iv) responsabilização institucional; (v) melhoria contínua. Adota-se tipologia de risco para calibrar salvaguardas: baixo (extração de metadados, buscas internas, apoio informacional), médio (triagem temática, priorização, sugestão de precedentes, minutas com revisão humana) e alto (influência material no mérito ou em direitos sensíveis; uso determinante de dados sensíveis). A classificação considera impacto sobre direitos, grau de autonomia, reversibilidade, sensibilidade dos dados, escala e opacidade; as

salvaguardas (documentação, AIA, transparência, auditoria, revisão humana) aumentam conforme o risco.

4.2. Transparência, governança de dados e ciclo de prestação de contas

Institui-se capacitação contínua (fundamentos de IA, equidade logaritmica, leitura de métricas, prevenção do *automation bias*, proteção de dados, motivação qualificada com apoio de IA e gestão de incidentes). Em governança de dados, asseguram-se minimização e qualidade, gestão de dados sensíveis, pseudonimização quando possível, controle de acesso, rastreabilidade de versões/usos, retenção e descarte regrados e vedação de reuso não compatível. Para transparência proporcional ao risco, cria-se Portal de IA com catálogo de sistemas, finalidades, versões, model cards/datasheets sintéticos, métricas (desempenho/equidade), relatórios de auditoria e canal de impugnação, resguardando segredos industriais e dados pessoais sem perder auditabilidade.

O ciclo de prestação de contas prevê: AIA antes da adoção; em produção, monitoramento de *drift*, calibração de limiares, registro de *overrides* e protocolo de incidentes (congelamento de versão, notificação, análise de causa e correção); e, anualmente, auditoria interna e, para alto risco, auditoria externa, com revalidação de métricas e salvaguardas. Em síntese, sistemas de IA agregam valor legítimo quando operam com finalidade clara, documentação e métricas auditáveis, motivação e contraditório preservados e sob prestação de contas contínua.

4.3. Papéis e responsabilidades (RASCI)

Para assegurar governança clara e responsabilização efetiva, recomenda-se a adoção de uma matriz RASCI que distribua papéis e deveres ao longo do ciclo de vida da solução de IA. O patrocinador institucional (R) aprova o caso de uso e aloca os recursos necessários, garantindo o alinhamento estratégico. O owner do processo (A), gestor da unidade usuária, responde pela aderência ao modelo de governança e pela integração do sistema às rotinas jurisdicionais.

A equipe técnica (S), composta por ciência de dados e engenharia, é responsável pela gestão de dados, pelo desenvolvimento e versionamento do modelo e pela aferição de métricas de desempenho e de equidade.

As assessorias jurídica e de conformidade (S) verificam a compatibilidade com a Constituição, o CPC e a LGPD, inclusive quanto à base legal e aos deveres de transparência e motivação. O encarregado de dados/DPO (S) zela pelo cumprimento dos princípios de proteção de dados, com ênfase em minimização, segurança, governança de dados sensíveis e registro de acessos.

A auditoria interna e/ou externa (C) realiza avaliações ex ante e ex post, examinando documentação, logs, métricas estratificadas e evidências de mitigação de viés. Por fim, o gabinete/usuário final (I) exerce a revisão humana significativa, registrando eventuais superações de recomendação algorítmica e a motivação correspondente no processo. Essa segregação funcional fortalece a prestação de contas e evita conflitos de interesse, ao mesmo tempo em que assegura controles jurídicos e técnicos ao longo de todo o pipeline. *(R = Responsible; A = Accountable; S = Support; C = Consulted; I = Informed).*

5. CONCLUSÕES

Este estudo concentrou-se em usos de apoio à decisão como: triagem, priorização, pesquisa de precedentes, geração de minutas e extração/estruturação de dados, não abrangendo, em detalhe, sistemas criminais de avaliação de risco nem a automação plena de sentenças.

Reconhece-se, como limitação, a ausência de evidência empírica baseada em casos reais, o que restringe a mensuração de efeitos causais e de possíveis *trade-offs* entre desempenho e equidade.

Os resultados indicam que os ganhos de eficiência trazidos por soluções de IA somente se mostram juridicamente sustentáveis quando acompanhados de (i) governança que explicite finalidades, limites e papéis; (ii) avaliação e mitigação de vieses ao longo de todo o ciclo (dados–modelo–uso); (iii) motivação qualificada, capaz de traduzir resultados técnico-computacionais em razões juridicamente relevantes; e (iv) contraditório técnico efetivo. A conformidade com igualdade, devido processo,

publicidade/motivação e proteção de dados depende menos da arquitetura do modelo e mais da institucionalização de processos que assegurem transparência, contestabilidade e prestação de contas.

Em termos operacionais, recomenda-se: Antes da adoção: realizar Avaliação de Impacto Algorítmico (AIA) com métricas de desempenho e de equidade estratificadas; publicar datasheet do conjunto de dados e model card do sistema.

No momento da decisão: vedar decisões exclusivamente automatizadas com impacto relevante; exigir revisão e motivação reforçada (CPC, art. 489, §1º), explicitando se, como e quanto a saída algorítmica influenciou o convencimento.

Em produção: monitorar drift de dados e de modelo, calibrar limiares, registrar superações da recomendação algorítmica e suas razões; manter portal de transparência com documentação proporcional ao risco.

Em caso de falhas: acionar protocolo de incidentes com congelamento da versão, notificação, análise de causa-raiz e plano corretivo técnico-procedimental.

Ciclo anual: realizar auditoria interna periódica e auditoria externa para casos de alto risco; revisar a AIA e revalidar métricas e salvaguardas, com reporte público sintético.

Em síntese, a adoção responsável de IA no Judiciário requer um arranjo institucional que alinhe eficiência a garantias processuais, por meio de modelo de gestão robusto, avaliação contínua de vieses, motivação qualificada e mecanismos efetivos de participação e controle.

REFERÊNCIAS

BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil**. Brasília, DF: Senado Federal, 1988.

BRASIL. **Emenda Constitucional nº 115, de 10 de fevereiro de 2022**. Altera a Constituição Federal para incluir a proteção de dados pessoais entre os direitos e garantias fundamentais. Brasília, DF: Congresso Nacional, 2022.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. **Lei de Acesso à Informação**. Brasília, DF: Presidência da República, 2011.

BRASIL. Lei nº 13.105, de 16 de março de 2015. **Código de Processo Civil**. Brasília, DF: Presidência da República, 2015.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. **Lei Geral de Proteção de Dados** Pessoais. Brasília, DF: Presidência da República, 2018.

CONSELHO NACIONAL DE JUSTIÇA (Brasil). **Painel de Estatísticas do Poder Judiciário**. Brasília, DF: CNJ, 2025.

CONSELHO NACIONAL DE JUSTIÇA (Brasil). **Plataforma Sinapses: inteligência artificial no Poder Judiciário**. Brasília, DF: CNJ, 2025.

CONSELHO NACIONAL DE JUSTIÇA (Brasil). **Resolução nº 332, de 21 de agosto de 2020**. Estabelece diretrizes para o desenvolvimento e uso de inteligência artificial no Poder Judiciário. Brasília, DF: CNJ, 2020.

SUPREMO TRIBUNAL FEDERAL (Brasil). **Projeto VICTOR: inteligência artificial para apoio à análise de admissibilidade recursal**. Brasília, DF: STF, 2021.

TRIBUNAL DE JUSTIÇA DO ESTADO DO RIO GRANDE DO SUL (Brasil). **Plataforma GAIA**: soluções de inteligência artificial do TJRS. Porto Alegre: TJRS, 2025.