

INTELIGÊNCIA ARTIFICIAL NO JUDICIÁRIO: DA OPACIDADE À EXPLICABILIDADE DAS DECISÕES JUDICIAIS

ARTIFICIAL INTELLIGENCE IN THE JUDICIARY: FROM OPACITY TO EXPLAINABILITY OF JUDICIAL DECISIONS

Daniel Nunes Pecego^A

<https://orcid.org/0000-0002-1231-1637>

Raphael Lobato Collet Janny Teixeira^B

<https://orcid.org/0009-0004-8969-1158>

^A Daniel Nunes Pêcego é Doutor, Mestre e Bacharel em Direito pela Universidade do Estado do Rio de Janeiro (UERJ). Especialista em Filosofia Medieval. Professor Adjunto da Faculdade de Direito da UERJ e Professor Colaborador do Programa de Pós-Graduação em Direito da UERJ. Coordenador do Curso de Graduação em Direito da UERJ e Coordenador Adjunto Administrativo Financeiro do Centro de Estudos e Pesquisas no Ensino do Direito da UERJ (CEPED).

^B Doutorando do Programa de Pós-Graduação em Direito da Universidade do Estado do Rio de Janeiro (UERJ). Mestre em Direito Constitucional pela Universidade Federal Fluminense (UFF); Pós-graduado em Propriedade Industrial pelo CEPED-UERJ. MBA em Gestão pela Fundação Dom Cabral. Graduado em Direito pela Universidade Presbiteriana Mackenzie (2002). Professor, advogado e consultor especializado em Direito Digital e Internet; Inteligência Artificial e questões ético-jurídicas relacionadas; Segurança da Informação e Proteção de Dados Pessoais; Direitos Autorais; Direito da Informática; Contratos de Tecnologia; Direito Regulatório em Telecomunicações.

Correspondência: dpecego@hotmail.com, raphalobato@gmail.com

DOI: <https://doi.org/10.12957/rfd.2024.87850>

Artigo recebido em 23/10/2024 e aceito para publicação em 28/10/2024.

Resumo: A adoção da Inteligência Artificial (IA) pelo setor público é recente, mas tem sido expandida pelo incremento do poder computacional e pela busca de redução de custos, dentre outros fatores. No Brasil, os órgãos do Judiciário são os que mais adotam a tecnologia em diversas aplicações de IA. Dado o grande número de processos, as ferramentas de IA são tidas como relevantes para que o Judiciário funcione de modo célere e eficaz. Apesar disso, surgem questões éticas relativas à adoção da IA em decisões judiciais, tais como as derivadas de vieses algoritmos e injustiças, o uso indevido de dados pessoais, além da adoção de algoritmos opacos de IA, em que não é possível alcançar explicabilidade de uma decisão. Este artigo identifica as possíveis causas da opacidade em algoritmos de IA utilizados em decisões judiciais, apresentando os fundamentos constitucionais, legais e regulatórios que

determinam a necessidade de explicabilidade da IA no Judiciário e propondo metodologias para alcançar explicabilidade em algoritmos opacos, apontando os desafios atuais para a implementação dessas metodologias. Apresentam-se definições de IA e de opacidade algorítmica, suas possíveis causas, bem como problemas identificados no processo de tomada de decisão judicial. Propõe-se um conceito de IA explicável e analisada a exigência legal e normativa de explicabilidade nas decisões judiciais no contexto jurídico brasileiro. São detalhados alguns métodos e ferramentas atuais que visam tornar os sistemas de IA mais transparentes e compreensíveis. São destacados os principais desafios, com indicações sobre os próximos passos para se obter uma maior explicabilidade nas decisões judiciais com uso de IA. Metodologicamente, a pesquisa foi bibliográfica, utilizando-se de fontes doutrinárias, legais e jurisprudenciais, priorizando-se publicações especializadas sobre a matéria.

Palavras-chave: Inteligência Artificial (IA); Inteligência Artificial explicável (xAI); Explicabilidade de decisões judiciais.

Abstract: The adoption of Artificial Intelligence (AI) by the public sector is recent, but it has been expanding due to the increase in computing power and the search for cost reduction, among other factors. In Brazil, the Judiciary is the one that most adopts the technology in various AI applications. Given the large number of cases, AI tools are considered relevant for the Judiciary to function quickly and effectively. Despite this, ethical issues arise regarding the adoption of AI in judicial decisions, such as those arising from algorithmic biases and injustices, the misuse of personal data, in addition to the adoption of opaque AI algorithms, in which it is not possible to achieve explainability of a decision. This article identifies the possible causes of opacity in AI algorithms used in judicial decisions, presenting the constitutional, legal and regulatory foundations that determine the need for explainability of AI in the Judiciary and proposing methodologies to achieve explainability in opaque algorithms, pointing out the current challenges for the implementation of these methodologies. The article presents definitions of AI and algorithmic opacity, their possible causes, and problems identified in the judicial decision-making process. The article proposes a concept of explainable AI and analyzes the legal and regulatory requirement for explainability in judicial decisions in the Brazilian legal context. Some current methods and

tools that aim to make AI systems more transparent and understandable are detailed. The main challenges are highlighted, with indications on the next steps to achieve greater explainability in judicial decisions using AI. Methodologically, the research was bibliographical, using doctrinal, legal, and jurisprudential sources, prioritizing specialized publications on the subject.

Keywords: Artificial Intelligence (AI); Explainable Artificial Intelligence (xAI); Explainability of judicial decisions.

1. INTRODUÇÃO

A adoção da Inteligência Artificial (IA) pelo setor público é relativamente recente, impulsionada pelo aumento do poder computacional, redução de custos e desenvolvimento de redes neurais, especialmente a partir de 2011 (OECD, 2019). De acordo com a pesquisa TIC Governo Eletrônico 2023, realizada pelo Cetic.br do NIC.br, sobre a adoção de tecnologias disruptivas no setor público brasileiro, a IA é a que mais cresceu (de 24%, em 2021, para 30%, em 2023) se comparado com Internet das Coisas (de 18% para 27%) e *Blockchain* (de 13% para 15%) (TIC..., 2023). A pesquisa apontou também que os órgãos do Judiciário são aqueles que mais adotam a tecnologia: em 2023, 68% das entidades ligadas ao Judiciário usavam aplicações de IA. O Judiciário também foi o Poder que mais promoveu capacitação em IA para os funcionários de TI (80%), quando comparado ao Legislativo (58%), Ministério Público (52%) e Executivo (25%) (TIC..., 2023).

De fato, inúmeras aplicações de IA vêm sendo utilizadas pelo Poder Judiciário. Como o Brasil lidera o ranking mundial em número de processos¹, as ferramentas de IA são vistas como importante recurso para que o Judiciário não seja imobilizado por tal quantidade de processos, viabilizando, assim, o seu andamento mais célere e uniforme.

Segundo estudo da Fundação Getúlio Vargas (FGV), sistemas de IA são utilizados para diversas finalidades pelo Judiciário (SALOMÃO, 2022). Nas atividades administrativas, auxiliam na administração de fóruns e tribunais por meio de *chatbots* e sistemas de

¹ Em 2022, o Brasil tinha cerca de 78 milhões de processos judiciais em andamento nos 91 tribunais, liderando o ranking mundial em número de processos (SALOMÃO, 2022).

monitoramento de visitantes. No que tange às atividades fins, esses sistemas contribuem para a prestação jurisdicional, como na transcrição de audiências e na classificação e triagem de processos, como fazem o ATHOS, do Superior Tribunal de Justiça (STJ), e o LARRY, do Tribunal de Justiça do Paraná.

O Supremo Tribunal Federal (STF), por sua vez, utiliza o sistema VICTOR, voltado para auxiliar os ministros na análise de admissibilidade recursal, indicando temas de repercussão geral. VICTOR consegue analisar se um recurso extraordinário se encaixa em um dos temas de repercussão em 5 segundos, enquanto o tempo médio de um servidor do STF é de 45 minutos (INGIZZA, 2023).

Essas tecnologias também são empregadas para avaliar processos passíveis de serem objeto de conciliação entre as partes (SALOMÃO, 2022). No entanto, a mensuração dos benefícios dessas ferramentas ainda é um desafio, sendo certo que futuras pesquisas serão necessárias para quantificar os impactos positivos da IA no Judiciário.

Em que pese os inquestionáveis benefícios decorrentes da adoção de IA pelo Judiciário, surgem inúmeros questionamentos de natureza ética relativos à adoção dessa tecnologia em decisões judiciais. Dentre outros problemas, como problemas relacionados a vieses algoritmos e injustiças (*bias*) e ao uso indevido de dados pessoais, questiona-se especialmente a adoção de algoritmos opacos de IA, que funcionam como uma caixa-preta (*black box*), onde não é possível conhecer o seu funcionamento interno ou obter uma explicação sobre como se chegou a uma predição, recomendação ou decisão específica (DEEKS, 2019).

Diante desse contexto, o presente artigo tem como objetivo identificar as possíveis causas da opacidade em algoritmos de IA utilizados em decisões judiciais. Além disso, visa apresentar os fundamentos constitucionais, legais e regulatórios que sustentam a necessidade de explicabilidade da IA no Judiciário brasileiro. Por fim, o artigo propõe metodologias para obter explicabilidade em algoritmos opacos e discute os desafios atuais para a implementação dessas metodologias.

Será apresentada uma definição de IA e da opacidade algorítmica, suas possíveis causas, bem como problemas identificados no processo de tomada de decisão judicial. Será proposto um conceito de IA explicável, também chamada de xAI (DEEKS, 2019), e analisada a exigência legal e normativa de explicabilidade nas decisões judiciais no contexto jurídico

brasileiro, avaliando como a legislação atual aborda a questão. Em seguida, serão detalhados alguns métodos e ferramentas atuais que visam tornar os sistemas de IA mais transparentes e compreensíveis. Por fim, serão destacados os principais desafios, oferecendo reflexões sobre os próximos passos para se obter uma maior explicabilidade nas decisões judiciais com uso de IA.

No que tange à metodologia adotada, empreendeu-se uma pesquisa bibliográfica, recorrendo às fontes doutrinária, legal e jurisprudencial, procurando-se identificar e priorizar publicações especializadas sobre o tema.

1. A PROBLEMÁTICA DA OPACIDADE ALGORÍTMICA EM DECISÕES JUDICIAIS

1.1 Definição de IA

Pesquisas sobre IA ganharam impulso após a Segunda Guerra Mundial com Alan Turing, considerado o pai da ciência da computação teórica e da IA. Turing abordou a questão da capacidade de computadores pensarem, propondo o “Teste de Turing” ou “Jogo da Imitação” para determinar a inteligência de uma máquina². A expressão "Inteligência Artificial" foi cunhada por John McCarthy et al. (2006), que a introduziram em uma conferência feita em 1956, no *Dartmouth College* (EUA).

Russell et al. (1995, p. 4) ressaltam que não há consenso entre cientistas e pesquisadores sobre a melhor definição de IA, cuja variação se dá em função de diferentes perspectivas, tais como processos de pensamento e raciocínio, comportamento e desempenho humano. A Organização para a Cooperação e Desenvolvimento Econômico (OECD, 2019), por sua vez, apresentou uma definição que vem sendo aceita como uma referência:

[...] um sistema de IA é um sistema baseado em máquina que pode, para um determinado conjunto de objetivos definidos pelo homem, fazer previsões, recomendações ou tomar decisões que influenciam ambientes reais ou virtuais. Os sistemas de IA são projetados para operar com vários níveis de autonomia.

² O teste envolve um interrogador que, ao obter respostas de um computador e um ser humano sem saber quem é quem, deve identificar qual é a máquina; se não conseguir, a máquina é considerada inteligente (TURING, 1950). Um exemplo moderno desse teste é o CAPTCHA, usado para diferenciar humanos de computadores na internet.

A IA pode ser entendida como a tecnologia capaz de executar tarefas que pressupõem o uso de inteligência, com um sistema computacional dotado de autoaprendizagem e que apresenta respostas de forma autônoma, sem uma predefinição específica dos seus programadores (TEIXEIRA, 2023, p. 17). Age não apenas automaticamente, mas de forma por vezes imprevisível em relação à sua programação inicial, sendo capaz de aprender por experiência, podendo alterar as instruções de programação iniciais e até criar novas, devido a técnicas de aprendizado de máquina (*machine learning*).

1.2 O que é opacidade algorítmica e suas possíveis causas

A opacidade algorítmica pode ser entendida como a impossibilidade de se obter uma explicação completa sobre funcionamento interno, recomendações, decisões e previsões de sistemas de IA “caixa preta” (*black box*), não sendo possível identificar como e por que os algoritmos alcançam seus resultados.

A opacidade em um sistema de IA pode decorrer, basicamente, de três causas: i) proteção à propriedade intelectual e do segredo de negócio, no caso de sistemas proprietários; ii) necessidade de possuir conhecimentos técnicos específicos para compreender o funcionamento do sistema; e, por fim, iii) aspectos técnicos relativos a modelos computacionais que possuem uma opacidade intrínseca à sua natureza.

No caso dos sistemas proprietários, as empresas desenvolvedoras de sistemas frequentemente mantêm o algoritmo de IA em segredo, não revelando o código-fonte, parâmetros de programação, seu funcionamento interno ou dados utilizados para treinamento da IA. As empresas alegam que é necessário proteger a sua propriedade intelectual e preservar o investimento realizado em pesquisa e desenvolvimento, impedindo, assim, que outras empresas copiem ou usem esses algoritmos sem permissão. Como resultado, os detalhes internos do algoritmo não são divulgados ao público, aos usuários, ou mesmo para agentes reguladores.

Quanto à necessidade de possuir conhecimentos técnicos específicos para compreensão de decisões automatizadas, é preciso ressaltar que algoritmos modernos, especialmente aqueles usados em *machine learning*, podem ser extremamente complexos. Eles podem envolver matemática avançada, estatística, ciência da computação e outras disciplinas técnicas que não são facilmente compreendidas sem uma elevada formação

especializada. Desta forma, mesmo que os detalhes de um algoritmo sejam divulgados ou acessíveis, entender como ele funciona pode ser extremamente difícil para quem não tenha o conhecimento técnico necessário. Essa barreira técnica contribui para a opacidade algorítmica, pois o funcionamento interno do sistema permanece "opaco" ou obscuro para a maioria das pessoas.

Quanto aos aspectos técnicos relativos a modelos computacionais que possuem uma opacidade intrínseca à sua natureza, ela é resultado dos ajustes contínuos que os algoritmos sofrem durante seu processo de aprendizado e operação, tornando quase impossível para os usuários e mesmos os seus desenvolvedores compreenderem completamente as decisões e previsões geradas (DEEKS, 2019). É preciso ressaltar, sem embargo, que a opacidade algorítmica neste caso não pode ser categorizada como uma falha de sistema em si, visto que existem limitações inerentes à complexidade de modelos de IA, como redes neurais profundas (*deep learning*). Alguns modelos podem ser mais explicáveis do que outros, mas sempre haverá um equilíbrio entre a transparência e o desempenho preditivo.

Modelos complexos, como *deep learning*, conseguem identificar relações e detalhes nos dados que modelos mais simples não conseguem, resultando em maior precisão. No entanto, essa complexidade torna esses modelos mais difíceis de interpretar. Por outro lado, modelos mais simples e altamente interpretáveis, como regressão linear ou árvores de decisão rasas, tomam decisões com base em regras claras e gerais. Em que pese esses modelos poderem ser considerados precisos, podem não capturar padrões específicos ou intrincados, como no caso de dados não estruturados, o que pode resultar em desempenho inferior se comparados a modelos de *deep learning* (MANN, 2023).

1.3 Como a opacidade algorítmica afeta o processo de tomada de decisão no Judiciário

A problemática da adoção de sistemas opacos pelo Poder Público desperta uma série de preocupações devido aos impactos em direitos fundamentais como liberdade, propriedade e privacidade. Como afirmou o ministro Gilmar Mendes, “vivemos na era das escolhas de Sofia automatizadas” (SUPREMO TRIBUNAL FEDERAL, 2020). Questiona-se, assim, a falta de transparência ou obscuridade de decisões automatizadas, não só para os indivíduos, como também para autoridades reguladoras e supervisoras (DONEDA et al., 2018).

Como pode-se observar no tópico anterior, a utilização de sistemas de IA *black box* apresenta desafios significativos devido à dificuldade em identificar como e por que esses algoritmos alcançam seus resultados, comprometendo a sua confiança e aceitabilidade. A opacidade dos algoritmos prejudica a satisfação da necessidade humana de respostas claras a perguntas como “Por quê?” e “Como você sabe?” (DEEKS, 2019).

Em contextos judiciais, conforme será mais desenvolvido no item seguinte, algoritmos opacos não atendem à exigência legal e constitucional de transparência e fundamentação das decisões. Além disso, prejudicam o exercício do contraditório e a ampla defesa, pois as partes envolvidas não têm informações suficientes para contestar ou compreender as decisões algorítmicas, restando inviabilizada a possibilidade de atacar decisões mediante recurso.

Além disso, sem uma compreensão clara de como os algoritmos operam, é difícil detectar e corrigir vieses, o que pode levar a decisões injustas ou discriminatórias. Conforme definido por Ntoutsis et al. (2020), o viés algoritmo é uma “*inclinação ou preconceito manifesto por uma decisão materializada por um sistema de IA que se apresenta como favorável ou contrária a uma pessoa ou grupo, especialmente de modo a traduzir uma injustiça*”. Algoritmos podem refletir preconceitos da sociedade, especialmente no tratamento de dados sensíveis³.

Dentre outros fatores, isso pode ocorrer porque os sistemas são desenvolvidos por pessoas que, voluntária ou involuntariamente, transmitem seus próprios vieses no código de programação, ou em razão das bases de dados utilizadas para treinar esses sistemas, que podem ser tendenciosas, o que pode comprometer a justiça nas decisões (DEEKS, 2019). Por esta razão, a falta de transparência na estrutura, conteúdo e testes dos algoritmos é uma preocupação significativa (DEEKS, 2019).

Por fim, há um risco considerável de que a discricionariedade judicial seja substituída pelo “viés de automação”. Este fenômeno ocorre quando as recomendações de algoritmos são aceitas sem a devido escrutínio ou crítica, sob alegação de se afastar particularismos em uma decisão. Não se olvida, por outro lado, que a capacidade de a IA identificar padrões proporciona decisões mais uniformes, objetivas e livres de subjetivismos. Contudo, a IA pode

³ Conforme definido no inciso II do Art. 5º da Lei 13.709/2018 (Lei Geral de Proteção de Dados), para os fins desta Lei, considera-se: “II - dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural”.

ser prejudicial em casos de conflito entre igualdade material e formal, onde decisões padronizadas podem resultar em injustiças por não considerar as peculiaridades do caso concreto (TEIXEIRA, 2023, p. 17).

Poderia haver, também, dificuldades para o magistrado que entenda que deve decidir de forma contrária à recomendação algorítmica em razão da falta de clareza em relação ao porquê daquela recomendação, trazendo um ônus argumentativo maior, ainda que, aos seus olhos, a orientação da IA seja injusta. Isso resultaria em uma evidente redução (ou até eliminação) da discricionariedade judicial.

De forma a ilustrar as questões acima apontadas, apresenta-se a seguir um caso prático sobre uso de IA no contexto da justiça criminal dos EUA. Juízes criminais são frequentemente instados a se manifestar sobre pedidos de liberdade de réus presos, analisando o cabimento da concessão de fiança, liberdade condicional ou suspensão da pena. Para análise desses pedidos, os juízes levam em consideração, dentre outros critérios, a periculosidade e a probabilidade de o requerente voltar a delinquir. Nesse contexto, algoritmos preditivos podem ser uma ferramenta valiosa para essas avaliações, reduzindo a dependência de intuições e preconceitos pessoais. Além disso, o uso desses algoritmos pode ajudar a diminuir as populações carcerárias sem afetar a segurança pública.

Ocorre que a adoção desse tipo de sistema vem sendo objeto de contestação, merecendo destaque o caso *“State v. Loomis”*, onde um réu contestou o uso pelo juiz de um algoritmo de sentenciamento chamado *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS) que o havia categorizado como representando um "alto risco de reincidência". O réu argumentou que o uso desse sistema de IA violou seus direitos processuais, em parte porque ele não foi capaz de avaliar a precisão do COMPAS (DEEKS, 2019).

A Suprema Corte do Estado do Wisconsin manteve a sentença, mas expressou cautela sobre o uso de algoritmos de sentenciamento opacos. A maioria dos juízes exigiu que os futuros relatórios de investigação antes da sentença contivessem avisos sobre as limitações do COMPAS, a fim de evitar possíveis violações do devido processo legal. Em sua opinião, a Juíza Shirley Abrahamson afirmou que “a falta de entendimento deste tribunal sobre o COMPAS foi um problema significativo no caso em questão” (DEEKS, 2019). Ela destacou que é indispensável que um tribunal registre e explique o uso de ferramentas baseadas em

evidências, incluindo suas limitações e pontos fortes, ao exercer discricionariedade na sentença. Mesmo o governo dos EUA também reconheceu que o uso de algoritmos de avaliação de risco não revelados pode levantar preocupações de devido processo legal, especialmente se o réu não tiver acesso às entradas fáticas (DEEKS, 2019).

Portanto, a adoção de sistemas opacos pelo Judiciário levanta preocupações significativas em função dos seus impactos em direitos fundamentais. Trata-se de uma preocupação real e atual, pois, conforme visto no caso “*State v. Loomis*”, o uso do algoritmo COMPAS foi contestado por sua opacidade, não provendo uma resposta clara e fundamentada sobre o direito à liberdade de réus presos.

2. FUNDAMENTOS E FERRAMENTAS DE xAI: BASES LEGAIS NO JUDICIÁRIO BRASILEIRO E MÉTODOS

2.1 Fundamento constitucional, legal e regulatório da IA explicável no Judiciário brasileiro

A explicabilidade da IA pode ser entendida como o conjunto de métodos e ferramentas para tornar os modelos de *machine learning* compreensíveis aos seres humanos, de sorte a explicar como esses modelos chegam às suas conclusões, permitindo, assim, que suas operações internas sejam interpretadas e compreendidas (DEEKS, 2019).

No contexto jurídico brasileiro, a xAI é respaldada pela necessidade de fundamentação das decisões judiciais por exigência constitucional e legal, razão pela qual magistrados que utilizem IA devem fundamentar suas decisões de forma humanamente inteligível, ou seja, passíveis de interpretação e compreensão humana.

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

A Lei nº 13.105/2015 (Código de Processo Civil; CPC) replica o texto constitucional no art. 11. Além disso, o art. 10 ressalta que o juiz não pode decidir com base em fundamento a respeito do qual não se tenha dado às partes oportunidade de se manifestar. Por

consequente, o juiz deve assegurar que as partes tenham a oportunidade de se pronunciar sobre os fundamentos derivados de análises feitas por IA. Isso implica em explicar de maneira compreensível como a IA chegou às suas recomendações e permitir que as partes questionem ou discutam essas conclusões.

O art. 489, por sua vez, define como elementos essenciais da sentença o relatório, os fundamentos e o dispositivo, e especifica no parágrafo 1º as situações em que uma decisão judicial não é considerada fundamentada.

Assim, a decisão que apenas reproduz ou parafraseia um ato normativo (I), que se utiliza de conceitos jurídicos indeterminados, como “razoabilidade” ou “boa-fé” (II), que faz mera invocação de precedentes ou enunciados de súmula (V), sem explicar ou demonstrar a pertinência ou relação daqueles fundamentos com o caso concreto, não é considerada suficientemente fundamentada. Isso porque a citação de uma norma, de um conceito jurídico ou invocação de precedentes deve ser acompanhada de uma análise contextual específica para que as partes compreendam sua aplicação. Portanto, a utilização de IA deve ser contextualizada, demonstrando de forma específica como as suas conclusões são pertinentes ao caso, detalhando como foram interpretados e aplicados.

Além disso, a fundamentação não pode utilizar motivos genéricos que poderiam justificar qualquer decisão (III), nem deixar de enfrentar todos os argumentos relevantes apresentados pelas partes (IV) ou deixar de seguir enunciado de súmula, jurisprudência ou precedente invocado pela parte, sem demonstrar a existência de distinção no caso em julgamento ou a superação do entendimento (VI). Desta forma, a decisão deve ser específica e completa para garantir a integridade do julgamento.

Em casos de colisão entre normas, o parágrafo 2º do art. 489 do CPC exige que o juiz justifique a ponderação efetuada, elucidando as razões que levaram à escolha de uma norma sobre a outra. O parágrafo 3º ressalta que a decisão deve ser interpretada em conformidade com o princípio da boa-fé, assegurando a coerência e a consistência na fundamentação. Desta forma, pelo princípio da boa-fé, a informação sobre o uso de IA pelo juiz deve ser transparente para as partes, incluindo descrever a metodologia utilizada pela IA, os dados de entrada considerados e o processo pelo qual a IA chegou à sua recomendação.

O Decreto-Lei nº 3.689/1941 (Código de Processo Penal; CPP) também estabelece requisitos para a fundamentação das decisões. O art. 381 dispõe que a sentença conterá a

indicação dos motivos de fato e de direito em que se funda a decisão, além de outros elementos essenciais. Essa disposição visa garantir que as decisões penais sejam bem fundamentadas e transparentes, permitindo uma compreensão clara dos motivos que levaram à condenação ou absolvição do acusado.

O Decreto-Lei n.º 4.657/1942 (Lei de Introdução às Normas do Direito Brasileiro; LINDB) também possui disposições específicas sobre decisões judiciais. O art. 20 dispõe que não se decidirá com base em valores jurídicos abstratos sem que sejam consideradas as consequências práticas da decisão. Além disso, o parágrafo único estipula que a motivação demonstrará a necessidade e a adequação da medida imposta ou da invalidação de ato, contrato, ajuste, processo ou norma administrativa, inclusive em face das possíveis alternativas. O art. 21, por sua vez, estipula que a decisão que decretar a invalidação de ato, contrato, ajuste, processo ou norma administrativa deverá indicar de modo expresso suas consequências jurídicas e administrativas, devendo indicar, quando for o caso, as condições para que a regularização ocorra de modo proporcional e equânime e sem prejuízo aos interesses gerais, não se podendo impor aos sujeitos atingidos ônus ou perdas que, em função das peculiaridades do caso, sejam anormais ou excessivos (parágrafo único).

A xAI também está intrinsecamente ligada ao princípio constitucional do devido processo legal, sendo condição *sine qua non* para a contestabilidade e revisibilidade das decisões judiciais. Este princípio assegura que as partes envolvidas em um processo administrativo ou judicial possam exercer plenamente o contraditório e a ampla defesa, conforme disposto no artigo 5º, LV da CRFB. O pleno exercício do contraditório e ampla defesa está condicionado à possibilidade de compreensão das razões de decidir do magistrado, o que somente será possível a partir de uma fundamentação clara, completa e concreta.

Do ponto de vista regulatório, a adoção de IA pelo Judiciário encontra-se regulamentada na Resolução 332, de 21 de agosto de 2020, do Conselho Nacional de Justiça (CNJ), que estabelece diretrizes para a ética, transparência e governança na utilização de IA no Poder Judiciário brasileiro. A Resolução destaca a necessidade de fornecer uma explicação satisfatória e passível de auditoria por autoridade humana quanto a qualquer proposta de decisão apresentada por modelos de IA, especialmente aquelas de natureza judicial (art. 8º, VI). A auditoria dos sistemas de IA, especialmente os opacos, requer uma abordagem que

inclua meios, condições e informações adequadas para a sua sindicabilidade. Isto significa que as decisões automatizadas devem ser passíveis de avaliação por auditores internos e externos, considerando-se os parâmetros, variáveis e bases de dados utilizados pelos desenvolvedores dos algoritmos.

Além disso, os sistemas computacionais que utilizem modelos de IA como ferramenta auxiliar para a elaboração de decisão judicial deverão permitir a supervisão do magistrado competente (art. 19, par. único). Esta previsão visa assegurar que magistrados, quando utilizem sistemas de IA, sejam capazes de justificar suas decisões de maneira compreensível e satisfatória.

Especificamente sobre decisões judiciais em matéria penal, o art. 23. da Resolução dispõe que, sobretudo com relação à sugestão de modelos de decisões preditivas, a utilização de IA não deve ser estimulada, salvo quando se tratar de utilização de soluções computacionais destinadas à automação e ao oferecimento de subsídios destinados ao cálculo de penas, prescrição, verificação de reincidência, mapeamentos, classificações e triagem dos autos para fins de gerenciamento de acervo (§ 1º).

Como se vê, a explicabilidade da IA nas decisões judiciais é uma decorrência de obrigação legal e constitucional de fundamentação, devendo os juízes e tribunais justificar suas decisões de maneira que permitam a sua compreensão, de forma clara, concreta e completa. Além disso, o uso de xAI auxiliará os juízes a entenderem melhor os parâmetros e funcionamento da IA, assegurando que os algoritmos utilizados são precisos e justos, evitando, assim, que erros significativos possam comprometer a integridade das decisões judiciais.

2.2. Métodos e ferramentas de xAI

Modelos de IA intrinsecamente explicáveis, embora tendam a ser menos complexos e precisos em suas previsões, oferecem uma vantagem significativa ao permitir uma compreensão clara de seu funcionamento interno. No entanto, muitos dos algoritmos mais avançados e eficazes utilizados atualmente não são intrinsecamente explicáveis, exigindo a adoção de abordagens específicas para promover a explicabilidade. Diante disso, apresentam-se a seguir alguns métodos xAI, tal como propostos por Deeks (2019), que, para algoritmos opacos, sugere ou uma abordagem exógena ou uma abordagem decomposicional.

As abordagens exógenas são aquelas que fornecem informações relevantes sobre o funcionamento externo do modelo sem requererem acesso ao seu *modus operandi* interno. Essa estratégia capacita indivíduos a entender e contestar as decisões algorítmicas que os afetam diretamente sem a necessidade de entender a lógica interna do modelo. Por exemplo, um indivíduo pode compreender que um determinado algoritmo classifica um conjunto de dados de uma maneira específica, sem necessariamente entender os cálculos internos que levaram a essa classificação.

Dentro das abordagens exógenas, podemos identificar abordagens centradas no modelo ou no sujeito. A abordagem centrada no modelo, conhecida como interpretabilidade global, envolve a explicação das intenções do criador do modelo, a família de modelos utilizada, os parâmetros especificados e as descrições dos dados de entrada, bem como o comportamento do modelo com novos dados. Outra variante dessa abordagem é a auditoria dos resultados do sistema para identificar vieses ou erros, assegurando que o modelo funcione de forma procedimentalmente regular.

A abordagem centrada no sujeito, ou interpretabilidade local, por sua vez, fornece informações sobre as características de indivíduos que receberam decisões similares. Essa abordagem pode utilizar contrafactuais, ajustando fatores de entrada para testar sua influência nas recomendações do algoritmo. Os contrafactuais são especialmente úteis pois permitem explorar diferentes “mundos possíveis próximos”, ajudando a entender como pequenas alterações nos dados de entrada poderiam mudar a decisão final do modelo.

Outra forma de buscar explicação de modelos não intrinsecamente explicáveis é adotar uma abordagem decomposicional. A abordagem decomposicional tenta explicar ou replicar o raciocínio interno do modelo.

Uma forma direta dessa abordagem é a revelação do código-fonte do modelo. Ocorre que o conhecimento do código fonte muitas vezes não é suficiente para se obter uma explicabilidade suficiente pois, conforme visto no tópico relativo às possíveis causas da opacidade, mesmo com a divulgação dos detalhes de um algoritmo, sua compreensão pode ser difícil para quem não possui o conhecimento técnico necessário. Além disso, mesmo programadores não conseguem extrair o motivo pelo qual algoritmos de *machine learning* fazem determinadas recomendações.

Outra forma de abordagem decomposicional é criar um “modelo substituto” que analisa pares de entrada e saída, sem necessidade de acesso aos pesos internos do modelo original. Por exemplo, uma árvore de decisão pode ser usada para refletir os cálculos de um modelo *black box*, permitindo compreender quais fatores influenciam suas decisões.

Abaixo, apresenta-se um quadro ilustrativo das diferentes abordagens de xAI:

Abordagem	Descrição
a) Modelos Intrinsecamente Explicáveis	Modelos menos complexos. Tendem a ser menos precisos.
b) Modelos Não Intrinsecamente Explicáveis	Modelos complexos que requerem abordagens adicionais para explicabilidade.
b.1. Abordagem Exógena	Fornecer informações sobre o funcionamento externo do modelo.
b.1.1 Centrada no Modelo	Explica as intenções do criador, parâmetros, dados de entrada e comportamento do modelo.
b.1.1.1 Auditoria de Resultados	Verifica vies ou erro no sistema.
b.1.2 Abordagem Centrada no Sujeito	Foca nas características de indivíduos que receberam decisões similares.
b.1.2.1 Contrafactuais	Ajusta fatores de entrada para testar influência nas recomendações.
b.2. Abordagem Decomposicional	Tenta explicar ou replicar o raciocínio interno do modelo.
b.2.1 Revelação do Código-Fonte	Divulga o código do modelo, mas pode ser insatisfatória devido à complexidade.
b.2.2 Modelo Substituto	Cria um modelo que analisa pares de entrada e saída para compreender decisões.

A Resolução 332/20 do CNJ não especifica, de forma explícita, uma metodologia relativamente à definição do método de xIA a ser adotado no Judiciário brasileiro. O art. 19 da Resolução dispõe que:

Art. 19. 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.

Parágrafo único. Os sistemas computacionais com atuação indicada no caput deste artigo deverão permitir a supervisão do magistrado competente.

Conforme estabelecido no referido artigo, a explicação dos passos que conduziram ao resultado produzido pela IA é o critério preponderante para definir a técnica utilizada. Pelo critério definido, parece que se opta por uma abordagem preferencialmente decomposicional, visto que se concentra em tentar explicar ou replicar o raciocínio interno do modelo, permitindo que o magistrado supervisione efetivamente o processo decisório.

3. DESAFIOS DA ADOÇÃO DA xAI

Há inúmeros desafios a serem enfrentados para a implementação de uma IA explicável no Judiciário brasileiro. O primeiro deles refere-se à definição de quais técnicas de xAI são consideradas satisfatórias para o atendimento da necessária fundamentação e explicabilidade das decisões judiciais.

Para integrar efetivamente a xAI no sistema judicial, é necessário considerar que as explicações devem ser adaptadas para diferentes audiências, desde advogados e juízes até as partes envolvidas e o público em geral. Cada grupo possui níveis distintos de familiaridade com conceitos tecnológicos e jurídicos, o que exige abordagens diferenciadas para garantir a compreensão adequada. Juízes podem preferir explicações centradas no modelo, que detalham como o algoritmo chega a determinadas conclusões. Réus, por outro lado, podem exigir explicações centradas no sujeito, que ajudem a entender como as decisões algorítmicas afetam especificamente seus casos. A execução de contrafactuais, onde cenários alternativos são simulados, pode beneficiar tanto juízes quanto réus, proporcionando uma visão mais abrangente das possíveis decisões (DEEKS, 2019).

É preciso levar em consideração que a abordagem exógena, que analisa o comportamento do modelo sem desmontá-lo, versus a abordagem decomposicional, que decompõe o modelo para entender suas partes constituintes, oferecem diferentes vantagens e desvantagens. Escolher a forma de xAI apropriada dependerá do contexto e do objetivo da explicação. O grau de explicação deve variar em função do seu uso, por exemplo, para elaboração de uma minuta de decisão, tal como ChatGPT⁴, onde o magistrado se utiliza da tecnologia para elaboração do relatório do caso ou para organizar os fundamentos a serem adotados, executando um trabalho, *mutatis mutandis*, de um assessor ou estagiário. Essa situação é diferente do caso de um algoritmo opaco que calcula o grau de periculosidade e fornece como resposta “alto risco de reincidência”, mas sem uma explicação adequada, como visto no caso do COMPASS.

Outro fator relevante a ser levado em consideração é o tempo necessário para compreender as explicações fornecidas. O tempo é um recurso valioso no ambiente judicial, e as explicações devem ser eficientes e rápidas, sem sacrificar a precisão e a profundidade das informações. Encontrar um equilíbrio entre a quantidade de detalhes fornecidos e a rapidez da compreensão é um desafio que os tribunais precisam enfrentar.

Portanto, propor um único método de xAI, tal como previsto na Resolução 332/20, que indica uma abordagem preferencialmente composicional, não parece ser a melhor forma de tratar a questão, pois não considera as diferentes situações e aplicações em que o Judiciário utiliza a IA. Neste aspecto, a Resolução deveria ser atualizada, de sorte a se sugerir o método de xAI aplicável em função do tipo de modelo de IA e do contexto em que a tecnologia é utilizada, sem perder de vista, em todos os casos, a necessidade de uma “explicação satisfatória”, compreensível, acessível e detalhada o suficiente para todas as partes envolvidas.

Outro desafio da xAI refere-se ao dilema explicabilidade *versus* acurácia do algoritmo. Apesar de a busca por explicabilidade trazer maior transparência para os modelos de IA, é preciso lembrar que, ao tentar-se transformar essas “caixas-pretas” em sistemas mais interpretáveis, existe o risco de sacrificar a sua acurácia e capacidade preditiva.

⁴ Recentemente, o Conselho Nacional de Justiça (CNJ) rejeitou um pedido para barrar o uso de ferramentas de IA, como o ChatGPT, no Judiciário para confeccionar atos processuais e fundamentar decisões. O CNJ afirmou que já existe regulamentação adequada sobre o tema. A decisão destacou a importância da supervisão humana nas decisões geradas por IA, garantindo a revisão e controle dos juízes (CONSELHO NACIONAL DE JUSTIÇA, 2024).

Também permanece o desafio relativo ao acesso a informações de sistemas desenvolvidos por empresas privadas. Conforme fora visto, empresas e desenvolvedores alegam que exigir transparência, como a revelação do código-fonte, parâmetros de programação ou dados utilizados para treinamento pode comprometer segredos comerciais, prejudicando sua vantagem competitiva. Os tribunais precisarão decidir como lidar com segredos comerciais ao exigir informações sobre algoritmos proprietários, incluindo revisões privadas ou revisões independentes sob acordo de confidencialidade, ou construir modelos substitutos que imitem o seu funcionamento sem comprometer os segredos comerciais. Diante disso, algumas jurisdições estão optando por adotar algoritmos que utilizam dados públicos e têm códigos-fonte disponíveis publicamente.

Entretanto, mesmo no caso de *softwares* públicos ou com acesso franqueado ao código fonte, a problemática da explicabilidade continua presente, pois, como visto, a complexidade desses algoritmos pode dificultar o entendimento de seu funcionamento, permanecendo hermético mesmo para os seus desenvolvedores, exigindo ferramentas adicionais de explicabilidade.

Cabe ressaltar, ainda, que construir sistemas explicáveis não é apenas tecnicamente desafiador, mas também financeiramente dispendioso. O desenvolvimento de ferramentas de xAI exige investimentos substanciais em tempo, mão de obra especializada e infraestrutura tecnológica (CONSELHO NACIONAL DE JUSTIÇA, 2024). Isso pode inviabilizar investimentos de pequenas e médias empresas, pois os altos custos necessários para investir na explicabilidade poderia limitar sua capacidade de competir com grandes corporações que dispõem de recursos para implementar essas tecnologias avançadas.

Em suma, enquanto a explicabilidade é indispensável para a adoção ética e justa da IA no Judiciário, os desafios associados não podem ser ignorados. A busca por um equilíbrio entre transparência e desempenho, inovação e proteção de propriedade intelectual, bem como a viabilidade econômica, é essencial para que a xAI se torne uma realidade viável e eficaz no contexto judicial.

CONCLUSÃO

A opacidade algorítmica é caracterizada pela impossibilidade de se obter uma explicação completa sobre o funcionamento interno, recomendações, decisões e previsões de

sistemas de IA descritos como “*black box*”. As causas dessa opacidade incluem: i) proteção à propriedade intelectual e ao segredo de negócio em sistemas proprietários, onde as empresas desenvolvedoras mantêm os algoritmos em segredo para proteger seu investimento; ii) a necessidade de conhecimentos técnicos específicos para compreender sistemas complexos de *machine learning*, que envolvem matemática avançada e ciência da computação; e iii) a opacidade intrínseca a modelos computacionais complexos, como redes neurais profundas, que sofrem ajustes contínuos durante o aprendizado e operação. Embora essa complexidade permita maior precisão ao identificar padrões complexos, acaba por tornar os modelos mais difíceis de interpretar. Por outro lado, modelos mais simples e interpretáveis, como regressões lineares, oferecem transparência, mas podem ser menos precisos ao lidar com dados não estruturados, resultando em desempenho inferior, se comparados aos modelos mais avançados.

A adoção de sistemas opacos pelo Judiciário suscita preocupações significativas devido aos seus impactos sobre direitos fundamentais como a liberdade, a propriedade e a privacidade. A falta de transparência em decisões automatizadas compromete a confiança e a aceitabilidade dos algoritmos, dificultando a compreensão e contestação das decisões pelos indivíduos e autoridades reguladoras. A opacidade também torna difícil detectar e corrigir vieses algorítmicos, levando a decisões injustas e discriminatórias. Também existe o risco do “viés de automação”, com aceitação acrítica das recomendações algorítmicas, reduzindo a discricionariedade judicial e resultando em decisões padronizadas que não consideram as peculiaridades de cada caso.

A explicabilidade da IA envolve métodos e ferramentas para tornar os modelos de *machine learning* transparentes aos seres humanos, permitindo a sua interpretação e compreensão. No contexto jurídico brasileiro, a xAI é indispensável para o atendimento da necessidade de fundamentação das decisões judiciais, devendo os juízes fornecerem uma explicação compreensível, detalhada e pertinente ao caso concreto, conforme exigido pelo artigo 93, IX da CRFB, arts. 10, 11 e 489 do CPC, art. 381 do CPP e arts. 20 e 21 da LINDB. A xAI também está vinculada ao princípio do devido processo legal (artigo 5º, LV da CRFB), visto que a adequada compreensão da decisão é condição indispensável para o exercício do contraditório e da ampla defesa.

Do ponto de vista regulatório, a adoção de IA pelo Judiciário brasileiro está regulamentada pela Resolução 332, de 21 de agosto de 2020, do CNJ, que estabelece diretrizes para a ética, transparência e governança na utilização de IA. A Resolução destaca a necessidade de fornecer explicações satisfatórias e auditáveis por autoridades humanas para qualquer decisão proposta por modelos de IA. A Resolução também exige que os sistemas de IA permitam a supervisão dos magistrados competentes. Especificamente em matéria penal, a Resolução dispõe que a utilização de modelos de decisões preditivas por IA não deve ser estimulada, exceto para automação e subsídios em cálculos de penas, verificação de reincidência e gerenciamento de acervo.

Modelos complexos e avançados de IA, como *deep learning*, não são explicáveis por si só, exigindo métodos específicos para obter a sua compreensão. As possíveis abordagens podem ser classificadas como exógenas ou decomposicionais. As exógenas permitem entender decisões algorítmicas sem acessar a lógica interna do modelo. Fornecem informações sobre o seu funcionamento externo e se dividem em centradas no modelo (interpretabilidade global) e no sujeito (interpretabilidade local). Já a abordagem decomposicional tenta replicar o raciocínio interno do modelo, seja pela revelação do código-fonte ou pela criação de “modelos substitutos”.

A implementação de IA explicável no Judiciário brasileiro enfrenta vários desafios, referentes ao equilíbrio entre transparência, desempenho, inovação, proteção de propriedade intelectual e viabilidade econômica. Primeiramente, é necessário definir técnicas de xAI que satisfaçam a fundamentação e explicabilidade das decisões judiciais, adaptando-as em função das diferentes audiências e aplicações. As abordagens exógenas ou decomposicional oferecem diferentes vantagens e desvantagens que devem ser sopesadas em função do contexto, sugerindo-se a forma adequada de xAI para cada situação. Remanesce, ainda, o dilema entre explicabilidade e acurácia do algoritmo, bem como o acesso a informações de sistemas proprietários *versus* a proteção de segredos comerciais. Por fim, deve-se levar em consideração o custo do desenvolvimento de sistemas explicáveis, que pode ser um obstáculo para pequenas e médias empresas.

REFERÊNCIAS

CONSELHO NACIONAL DE JUSTIÇA. PROCEDIMENTO DE CONTROLE ADMINISTRATIVO - 0000416-89.2023.2.00.0000. Brasília, 25 jun. 2024. Disponível em: https://www.conjur.com.br/wp-content/uploads/2024/07/9769C290CD7A61_0000416-89.2023.2.00.0000_5617.pdf. Acesso em: 26 jul. 2024.

DEEKS, A. THE JUDICIAL DEMAND FOR EXPLAINABLE ARTIFICIAL INTELLIGENCE. *Columbia Law Review*, v. 119, n. 7, p. 1829–1850, 2019.

DONEDA, D. et al. Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal. *Pensar, Fortaleza*, v. 23, n. 4, p. 1-17, dez. 2018.

INGIZZA, C. De 44 minutos a 5 segundos: como a Inteligência Artificial agiliza o Judiciário brasileiro. 2023. JOTA. Disponível em: <https://www.jota.info/justica/de-44-minutos-a-5-segundos-como-a-inteligencia-artificial-agiliza-o-judiciario-brasileiro-18092023>. Acesso em: 18 out. 2024.

MANN, H. Do All AI Systems Need to Be Explainable? *Stanford Social Innovation Review*. 2023. Disponível em: https://ssir.org/articles/entry/do_ai_systems_need_to_be_explainable. Acesso em: 26 jul. 2024.

MCCARTHY, J. et al. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Magazine*, v. 27, n. 4, p. 12-14, ago. 2006.

NTOUTSI, E. et al. Bias in Data-driven Artificial Intelligence systems—An Introductory Survey. *WIREs Data Mining and Knowledge Discovery*, v. 10, n. 3, 3 fev. 2020.

OECD. *Artificial Intelligence in Society*. Paris: OECD Publishing, 2019. Disponível em: https://www.oecd-ilibrary.org/science-and-technology/artificial-intelligence-in-society_eedfee77-en. Acesso em: 26 jul. 2024.

RUSSELL, S et al. *Artificial Intelligence: A Modern Approach*. New Jersey: Prentice Hall, 1995.

SALOMÃO, L. F. *Artificial intelligence: technology applied to conflict management within the brazilian judiciary*. repositorio.fgv.br, 2022.

SUPREMO TRIBUNAL FEDERAL. Ação Direta de Inconstitucionalidade nº 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.

TEIXEIRA, R. Regulação Ético-Jurídica de Inteligência Artificial na Administração Pública. Rio de Janeiro: Lumen Juris, 2023.

TIC Governo Eletrônico 2023 mostra que 91% das prefeituras disponibilizam ao menos um serviço online aos cidadãos. Disponível em: <https://cetic.br/pt/noticia/tic-governo-eletronico-2023-mostra-que-91-das-prefeituras-disponibilizam-ao-menos-um-servico-online-aos-cidadaos/>. Acesso em: 26 jul. 2024.

TURING, A. Computing Machinery and Intelligence. *Mind*, v. 59, n. 236, p. 433-460, out. 1950.