

A INTELIGÊNCIA ARTIFICIAL E CRITÉRIOS PARA O CONTROLE JURISDICIONAL: UMA ANÁLISE CASUÍSTICA DOS SISTEMAS UTILIZADOS PELO STF E STJ EM SEDE DE FILTRO RECURSAL

ARTIFICIAL INTELLIGENCE AND CRITERIA FOR JUDICIAL REVIEW: A CASE-BY-CASE ANALYSIS OF THE SYSTEMS USED BY THE STF AND STJ IN THE CONTEXT OF APPELLATE FILTERING

Roberto Tadao Magami Junior¹

Doutorando em Direito (USP, São Paulo/SP, Brasil)

ÁREA(S): controle jurisdicional; direito administrativo; direito constitucional; dogmática jurídica; inteligência artificial.

RESUMO: Considerando que metade dos tribunais brasileiros possuem projetos de inteligência artificial em andamento, e mecanismos de IA já são utilizados no juízo de admissibilidade de recursos no âmbito do STF e do STJ, inexistindo, até o presente momento, informações para avaliar a vinculação lógica entre problemas, ações e resultados destas soluções, qual é a dimensão do impacto dessas mudanças no cotidiano da advocacia, seja ela pública

ou privada, e na atuação do Ministério Público e da Defensoria Pública e, ao fim e ao cabo, nos usuários dos serviços públicos de índole jurisdicional e, acima de tudo, na segurança jurídica? Quais dados são utilizados para que esses novos filtros recursais sejam aplicados aos recursos interpostos? Como esses dados são processados? Qual é o desempenho geral dos sistemas que utilizam dados específicos para realizar essas tarefas? Como esses algoritmos foram construídos? O presente estudo visa propor *standards* relacionados ao *accountability* para o adequado controle jurisdicional da inteligência artificial até que sobrevenha legislação sobre o

¹ Advogado, Procurador Autárquico e Árbitro integrante da Câmara de Mediação e Conciliação da FGV-SP. Pós-Graduado e Mestre em Direito Administrativo pela Pontifícia Universidade Católica de São Paulo. Pós-Graduado em Arbitragem pela Faculdade de Direito Universidade de Lisboa. Pós-Graduado em Direito da Proteção de Dados pela Faculdade de Direito Universidade de Lisboa/Irish Computer Society. E-mail: roberto.tadao@tadaomagami.adv.br. Currículo: <http://lattes.cnpq.br/2820266304990892>. Orcid: <https://orcid.org/0009-0004-7940-3386>.

tema, por intermédio dos métodos observacional e comparativo enquanto meios técnicos para garantir a objetividade e a precisão na sua compatibilidade com os padrões internacionais de controle da IA, incluindo a *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, editada em outubro de 2023 pelos EUA, o Regulamento UE nº 2024/1689, que cria regras harmonizadas sobre IA no âmbito da União Europeia, além do Projeto de Lei nº 2.338/2023 do Senado Federal.

ABSTRACT: *Considering that half of the Brazilian courts have artificial intelligence projects in progress, and AI mechanisms are already used in the admissibility judgment of appeals within the scope of the STF and the STJ, and so far there is no information to assess the logical link between problems, actions and results of these solutions, what is the dimension of the impact of these changes on the daily life of the legal profession, whether public or private, and in the performance of the Public Prosecutor's Office and the Public Defender's Office and, at the end of the day, in the users of public services of a jurisdictional nature and, above all, in legal certainty? What data is used for these new appeal filters to be applied to the appeals filed? How is this data processed? What is the overall performance of systems that use specific data to perform these tasks? How were these algorithms built? The present study aims to propose standards related to accountability for the adequate jurisdictional control of Artificial Intelligence until national legislation on the subject is enacted, through observational and comparative methods as technical means to ensure objectivity and accuracy in its compatibility with international standards for AI control, including the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence issued in October 2023 by the USA, EU Regulation 2024/1689 that creates harmonized rules on AI within the European Union, in addition to the proposal of Bill n. 2.338/2023 of the Brazilian Federal Senate.*

PALAVRAS-CHAVE: controle jurisdicional; direito administrativo; direito constitucional; dogmática jurídica; inteligência artificial; interdisciplinariedade; serviços públicos.

KEYWORDS: *judicial review; administrative law; constitutional law; legal dogmatics; artificial intelligence; interdisciplinarity; public services.*

SUMÁRIO: Introdução; 1 A repercussão geral no STF e a inteligência artificial; 2 A arguição de relevância da questão infraconstitucional no STJ e a inteligência artificial; 3 Conceito de inteligência artificial; 4 Riscos no desenvolvimento da IA; 5 *Accountability* e a imprescindível disponibilização de informações dos sistemas de inteligência artificial; 6 A Resolução nº 615, de 11 de março de 2025, do Conselho Nacional de Justiça; 7 Controle formal da IA – Critérios para a *accountability* a partir

do direito comparado e do Projeto de Lei do Senado nº 2.338/2023; 8 Controle jurisdicional material ou quanto ao objeto; Conclusão; Referências.

SUMMARY: *Introduction; 1 The general repercussion in the Brazilian Supreme Court and artificial intelligence; 2 The Argument of relevance of the infra-constitutional issue in the Superior Court of Justice and artificial intelligence; 3 The Concept of artificial intelligence; 4 Risks in the development of AI; 5 Accountability and the essential availability of information from artificial intelligence systems; 6 Resolution n. 615, of march 11, 2025, of the National Council of Justice; 7 Formal control of the IA – criteria for accountability based on comparative law and Senate Bill n. 2.338/2023; 8 Material jurisdictional control; Conclusion; References.*

INTRODUÇÃO

Cerca de metade dos tribunais brasileiros possuem projetos de inteligência artificial (IA) operantes ou em desenvolvimento, segundo relatório produzido pelo Centro de Inovação, Administração e Pesquisa do Judiciário da Fundação Getúlio Vargas (CIAPJ/FGV)².

Atualmente, a IA é utilizada pelo Poder Judiciário desde a transcrição de audiências, a elaboração de sugestões de minutas para decisões judiciais e até a realização do juízo de admissibilidade de recursos e o cálculo da probabilidade de reversão das decisões³.

Estão em desenvolvimento 64 projetos de inteligência artificial em 47 tribunais, além da plataforma operada pelo Conselho Nacional de Justiça (CNJ)⁴.

Por essa razão, é de extrema importância compreendermos e discutirmos as formas de implementação, execução e controle em termos de *accountability* da inteligência artificial em si no âmbito da Administração Pública, na medida em que haverá enormes reflexos não somente na prestação de serviços públicos, mas também nas próprias políticas públicas nacionais, e, até o momento, *há*

² Disponível em: <https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/09032021-Inteligencia-artificial-esta-presente-em-metade-dos-tribunais-brasileiros--aponta-estudo-inedito.aspx>. Acesso em: 10 mar. 2021.

³ Disponível em: <https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/09032021-Inteligencia-artificial-esta-presente-em-metade-dos-tribunais-brasileiros--aponta-estudo-inedito.aspx>. Acesso em: 10 mar. 2021.

⁴ Disponível em: <https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/09032021-Inteligencia-artificial-esta-presente-em-metade-dos-tribunais-brasileiros--aponta-estudo-inedito.aspx>. Acesso em: 10 mar. 2021.

um vácuo na dogmática jurídica clássica, porquanto ainda ausente uma regulação do tema, em que pese a sua ampla utilização e a pendência de discussão no Congresso Nacional do PL 2.338/2023.

Por conseguinte, é essencial a reflexão sob a ótica dos Direitos Constitucional e Administrativo, na medida em que possuem interconexão⁵ com outros ramos do conhecimento, como a ciência política, a ciência da administração, a economia e obviamente com a inovação tecnológica desde a EC 85/2015⁶, que, por se tratar de norma de eficácia contida e aplicação direta, imediata, mas não integral, foi complementada pelo Marco Legal da Inovação⁷, pelo Marco Legal das Startups (Lei Complementar nº 182/2021)⁸ e pela Lei de licitações e contratações públicas (Lei nº 14.133/2021)⁹.

⁵ Há multidisciplinaridade quando duas ou mais disciplinas se justapõem, mas, no entanto, mantêm identidade e estrutura originais. Já na interdisciplinaridade, a interação entre as disciplinas assume caráter mais proativo, de modo a que o método de uma seja incorporado na outra. A interdisciplinaridade pressupõe interconexão. (SANTOS DE MENDONÇA, J. V.; BAPTISTA, P. F. Dogmática jurídica e interdisciplinaridade: caminhos para um diálogo útil. *Rev. Quaestio Iuris*, Rio de Janeiro, v. 15, n. 04, p. 2296-2310, 2022)

⁶ “Art. 200. Ao sistema único de saúde compete, além de outras atribuições, nos termos da lei: [...] V – incrementar, em sua área de atuação, o desenvolvimento científico e tecnológico e a inovação; [...] Art. 218. O Estado promoverá e incentivará o desenvolvimento científico, a pesquisa, a capacitação científica e tecnológica e a inovação. § 1º A pesquisa científica básica e tecnológica receberá tratamento prioritário do Estado, tendo em vista o bem público e o progresso da ciência, tecnologia e inovação. § 2º A pesquisa tecnológica voltar-se-á preponderantemente para a solução dos problemas brasileiros e para o desenvolvimento do sistema produtivo nacional e regional. § 3º O Estado apoiará a formação de recursos humanos nas áreas de ciência, pesquisa, tecnologia e inovação, inclusive por meio do apoio às atividades de extensão tecnológica, e concederá aos que delas se ocupem meios e condições especiais de trabalho.” (grifos nossos)

⁷ O Marco Legal da Ciência, Tecnologia e Inovação, instituído pela Lei nº 13.243/2016 e regulamentado pelo Decreto nº 9.283/2018, favorece o desenvolvimento do ambiente de inovação no Brasil.

⁸ “Da contratação de soluções inovadoras pelo Estado [...] Art. 12. As licitações e os contratos a que se refere este Capítulo têm por finalidade: I – resolver demandas públicas que exijam solução inovadora com emprego de tecnologia; e II – promover a inovação no setor produtivo por meio do uso do poder de compra do Estado.” (grifos nossos)

⁹ “Art. 11. O processo licitatório tem por objetivos: [...] IV – incentivar a inovação e o desenvolvimento nacional sustentável. [...] Art. 32. A modalidade diálogo competitivo é restrita a contratações em que a Administração: I – vise a contratar objeto que envolva as seguintes condições: a) inovação tecnológica ou técnica; [...] Art. 75. É dispensável a licitação: [...] IV – para contratação que tenha por objeto: [...] d) transferência de tecnologia ou licenciamento de direito de uso ou de exploração de criação protegida, nas contratações realizadas por instituição científica, tecnológica e de inovação (ICT) pública ou por agência de fomento, desde que demonstrada vantagem para a Administração; [...] XV – para contratação de instituição brasileira que tenha por finalidade estatutária apoiar, captar e executar atividades de ensino, pesquisa, extensão, desenvolvimento institucional, científico e tecnológico e estímulo à inovação, inclusive para gerir administrativa e financeiramente es-

Atualmente e a despeito de qualquer regulação, já estão em discussões no âmbito da Administração Pública¹⁰ sistemas de IA, tendo inclusive o Tribunal de Contas da União (TCU) realizado um levantamento sobre o tema¹¹

...sas atividades, ou para contratação de instituição dedicada à recuperação social da pessoa presa, desde que o contratado tenha inquestionável reputação ética e profissional e não tenha fins lucrativos; [...] Do Procedimento de Manifestação de Interesse. Art. 81. A Administração poderá solicitar à iniciativa privada, mediante procedimento aberto de manifestação de interesse a ser iniciado com a publicação de edital de chamamento público, a propositura e a realização de estudos, investigações, levantamentos e projetos de soluções inovadoras que contribuam com questões de relevância pública, na forma de regulamento. [...] § 4º O procedimento previsto no caput deste artigo poderá ser restrito a startups, assim considerados os microempreendedores individuais, as microempresas e as empresas de pequeno porte, de natureza emergente e com grande potencial, que se dediquem à pesquisa, ao desenvolvimento e à implementação de novos produtos ou serviços baseados em soluções tecnológicas inovadoras que possam causar alto impacto, exigida, na seleção definitiva da inovação, validação prévia fundamentada em métricas objetivas, de modo a demonstrar o atendimento das necessidades da Administração.” (grifos nossos)

¹⁰ Conforme Portaria MCTI nº 4.617/2021, caberá ao Ministério de Ciência, Tecnologia e Inovações, no âmbito de suas atribuições: I – criar instâncias e práticas de governança para priorizar, implantar, monitorar e atualizar as ações estratégicas estabelecidas na Estratégia Brasileira de Inteligência Artificial; II – coordenar e estabelecer as ações que possibilitem a implementação da Estratégia Brasileira de Inteligência Artificial; III – convidar instituições do setor público, privado e da academia para subsidiá-lo nas ações estratégicas definidas na Estratégia Brasileira de Inteligência Artificial; IV – elaborar relatórios de avaliação da implementação das ações estratégicas definidas na Estratégia Brasileira de Inteligência Artificial; V – dar publicidade aos relatórios de que trata o inciso III do *caput* deste artigo no sítio eletrônico deste Ministério; e VI – avaliar a periodicidade de atualização da Estratégia Brasileira de Inteligência Artificial. A Governança da EBIA é composta por: I – Ministério da Ciência, Tecnologia e Inovações – MCTI; II – Rede MCTI/EMBRAPII de Tecnologias e Inovação Digital; e III – Instituições Convidadas.

¹¹ O Tribunal de Contas da União (TCU) realizou, sob a relatoria do Ministro Aroldo Cedraz, levantamento com o objetivo de avaliar o estágio atual e perspectivas de utilização de inteligência artificial (IA) na Administração Pública Federal (APF), identificar os riscos associados, conhecer os impactos para o controle e avaliar a proposta para uma Estratégia Brasileira de Inteligência Artificial (EBIA). Mais de um terço (38%) das organizações federais estão no nível zero de maturidade em IA, ou seja, não utilizam e sequer planejam utilizar essa tecnologia exponencial. Um outro terço da APF (33,5%) se encontra no nível 1. Isso significa que já estão tendo conversas internas sobre a inteligência artificial, mas de modo ainda especulativo. Cerca de três em cada dez instituições da APF (28,5%) se localizam nos níveis 2, 3 ou 4 de maturidade em IA. Sendo a maior parte delas (17,1% do total), na fase de experimentação, com provas de conceito elaboradas ou já em fase piloto. Em torno de 8% das organizações federais estão na fase de estabilização, com os primeiros projetos de IA em produção. Apenas 3,4% do total já está no nível 4, expandindo para novos projetos de IA. O levantamento atual do Tribunal revelou lacunas que podem comprometer o alcance dos objetivos das organizações federais. Foi verificado o problema da ausência de objetivos específicos, realistas e mensuráveis de IA na APF. A fiscalização do TCU apontou faltar uma referência inicial para avaliar os resultados em IA no âmbito federal. Além de terem sido encontradas falhas no modelo de vinculação lógica entre problemas, ações e resultados das soluções em inteligência artificial. Disponível em: <https://portal.tcu.gov.br/imprensa/noticias/tcu-avalia-uso-de-inteligencia-artificial-pelo-governo-federal.htm>. Acesso em: 2 set. 2022.

e encontrado falhas no modelo de vinculação lógica entre problemas, ações e resultados das soluções em inteligência artificial.

Logo, a realidade socioeconômica se impõe ao Direito que deverá responder aos problemas com razoável grau de previsibilidade, estabilidade e justiça, na medida em que o Poder Judiciário não pode se valer do *non liquet*¹².

Assim, apenas como um modelo em que poderemos extrair respostas iniciais, trataremos neste artigo acerca da IA enquanto mecanismo utilizado para o juízo de admissibilidade no âmbito do Supremo Tribunal Federal (STF) e do Superior Tribunal de Justiça (STJ), mas não sem antes contextualizar os seus denominados filtros recursais: (a) a repercussão geral, (b) a arguição de relevância da legislação infraconstitucional e (c) como esta é utilizada para supostamente auxiliar na análise de inúmeros recursos.

Desde há muito se discute sobre as condições físicas e materiais dos tribunais superiores analisarem a quantidade de recursos, sem contar os *habeas corpus* e os mandados de segurança que são decididos monocraticamente por ministros.

1 A REPERCUSSÃO GERAL NO STF E A INTELIGÊNCIA ARTIFICIAL

A partir da EC 45/2004, o § 3º do art. 102 da Constituição passou a exigir que qualquer recurso interposto ao STF apresente a denominada repercussão geral, a fim de se permitir a admissibilidade do recurso.

Em 2006, foi editada a Lei nº 11.418, dando concretude ao mandamento constitucional e conceituando a repercussão geral como algo que ultrapassa os interesses subjetivos do processo, seja sob o ponto de vista econômico, político, social ou jurídico, além de atribuir uma presunção *jure et de jure* de que qualquer decisão contrária à sumula ou jurisprudência consolidada do STF possui repercussão geral implícita.

Em razão dos desenvolvimentos práticos do instituto no cotidiano do STF, ocorreram inúmeras alterações em seu Regimento Interno¹³ que foram

¹² Código de Processo Civil: “Art. 140. O juiz não se exime de decidir sob a alegação de lacuna ou obscuridade do ordenamento jurídico. Parágrafo único. O juiz só decidirá por equidade nos casos previstos em lei”.

¹³ Arts. 322 a 329 do Regimento Interno do STF.

dialogando sempre em torno das diretrizes previstas nos arts. 1.035¹⁴ e seguintes do Código de Processo Civil.

Consoante números do STF, até 27 de dezembro de 2023 foram proferidas 36.177 decisões, sendo que 95,4% (34.520) negaram provimento aos recursos interpostos.

Para obter essas informações, desde 2017 o STF faz uso da inteligência artificial denominada Victor¹⁵, auxiliando-o na análise dos recursos extraordinários (RE) que são recepcionados em meio eletrônico.

Em razão da heterogeneidade do País e da autonomia do Poder Judiciário prevista no art. 99 da CF, os autos processuais chegam em distintas configurações eletrônicas à Suprema Corte: em alguns casos, como arquivos no formato “pdf”, em formato de imagem, sem camada de texto puro que viabilize a leitura por máquina e, em outros, sem qualquer identificação ou indexação, isto é, sem que as peças estejam rotuladas, como sentença, apelação, acórdão, recurso extraordinário etc.

Para solucionar essa questão, desde 2020 o Victor realiza quatro atividades ao analisar um RE ou o agravo dirigido ao STF que permita a subida daquele: converte imagens em textos no processo digital (*ocerização*); separa e categoriza do começo ao fim os documentos (*splitter*); separa e classifica as peças processuais mais utilizadas nas atividades do STF (indexação); e, ao final, sinaliza se um tema definido em sede de repercussão geral, ou mais de um, se aplica aos autos daquele processo, para posterior validação ou confirmação durante a efetiva apreciação do caso concreto pelos ministros.

Em seguida, o classificador entra em ação para indexar ou identificar essas peças de acordo com a nomenclatura utilizada pelo STF.

¹⁴ Arts. 1.035 a 1.041 do Código de Processo Civil.

¹⁵ O nome do projeto é uma homenagem a Victor Nunes Leal (falecido), Ministro do Supremo Tribunal Federal de 1960 a 1969, autor da obra *Coronelismo, enxada e voto* e principal responsável pela sistematização da jurisprudência da Corte em súmula, o que facilitou a aplicação dos precedentes judiciais aos recursos e foi incorporado pela Resolução nº 800, de 17 de maio de 2023, do STF (STF. Projeto Victor avança em pesquisa e desenvolvimento para identificação dos temas de repercussão geral. Disponível em: <https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=471331&ori=1>. Acesso em: 19 ago. 2021).

Atualmente, todos os recursos extraordinários e agravos visando à subida do primeiro ao STF são autuados e analisados pela Secretaria de Gestão de Precedentes e decididos pelo ministro presidente.

Decidindo-se pelo enquadramento em tema de repercussão geral, o processo é devolvido à instância de origem para sua apreciação. Em caso negativo, e igualmente não se verificando outras questões processuais, como a tempestividade e oportunidade de aplicação de súmulas, o presidente determina a distribuição dos recursos aos demais ministros da Corte.

A título complementar, desde 2020, o STF utiliza uma ferramenta denominada RAFA 2030 (Redes Artificiais Focadas na Agenda 2030), desenvolvida para apoiar a classificação das ações de controle concentrado e recursos extraordinários com repercussão geral reconhecida na Corte, de acordo com os Objetivos de Desenvolvimento Sustentável da Agenda 2030 da ONU¹⁶.

O algoritmo do Victor foi treinado com arquitetura de rede neural convolucional (usada principalmente em imagens¹⁷) e memória de longo prazo (busca identificar termos distantes e suas relações).

Redes neurais convolucionais classificam imagens e reconhecem objetos por intermédio de um desempenho a partir de camadas. Cada neurônio em uma camada convolucional está conectado apenas a uma pequena região para detectar características como bordas, texturas e formas. Todos os neurônios estão interconectados, preparando a rede para a tarefa específica, como, por exemplo, a classificação.

Com cada camada, a rede neural convolucional aumenta a sua complexidade identificando partes maiores da imagem. À medida que os dados da imagem progridem por meio das camadas da rede neural convolucional, começa a reconhecer elementos ou formas maiores do objeto até que finalmente identifica o dado pretendido¹⁸.

¹⁶ Os 17 Objetivos de Desenvolvimento Sustentável e 169 metas estabelecidas pela ONU buscam concretizar os Direitos Humanos de todos e alcançar a igualdade de gênero e o empoderamento das mulheres, de forma integrada, por intermédio das três dimensões do desenvolvimento sustentável: econômica, social e ambiental.

¹⁷ A visualização da imagem não é igual à nossa enquanto seres humanos, mas sim volumes tridimensionais, ou seja, objetos tridimensionais que são medidos pelos seus *pixels*, em vez de estruturas planas que medimos por largura e altura.

¹⁸ LECUN, Y. et al. Gradient-based learning Applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, 1998.

As redes neurais convolucionais são apenas uma das 10 principais arquiteturas de redes de *deep learning*, comumente utilizadas no reconhecimento de objetos, faces e imagens, dentre outras: 1) *Redes Multilayer Perceptrons*; 2) *Redes Neurais Convolucionais*; 3) *Redes Neurais Recorrentes*; 4) *Long Short-Term Memory (LSTM)*; 5) *Redes de Hopfield*; 6) *Deep Belief Network*; 7) *Generative Adversarial Network*; 8) *Deep Neural Network Capsules*; 9) *Deep Auto-Encoders*; 10) *Máquinas de Boltzmann*¹⁹.

2 A ARGUIÇÃO DE RELEVÂNCIA DA QUESTÃO INFRACONSTITUCIONAL NO STJ E A INTELIGÊNCIA ARTIFICIAL

Noutro giro, a CF alterada pela EC 125/2022 passou a exigir, em seu art. 105, § 2º²⁰, que qualquer recurso especial (REsp) demonstre a relevância das questões de direito federal infraconstitucional discutidas nos autos do processo, a fim de que a admissão do recurso seja examinada pelo tribunal, o qual somente pode dele não conhecer pela manifestação de 2/3 dos membros do órgão competente para o julgamento.

O Poder constituinte derivado atribuiu uma presunção de relevância da questão – afastando a necessidade de uma preliminar no recurso especial – quando do julgamento das ações penais; de improbidade administrativa; cujo valor da causa ultrapasse 500 salários-mínimos; que possam gerar inelegibilidade; e quando o acórdão recorrido contrariar sua jurisprudência dominante, sem prejuízo da criação de novas hipóteses por meio de lei.

Segundo estatística apresentada pela Comissão Gestora de Precedentes e de Ações Coletivas (COGEPAC)²¹, no ano inteiro de 2023 perante o STJ apenas 4% dos AREsps foram providos, ao passo que 59,8% dos AREsps sequer foram conhecidos, o que representava mais de 122 mil agravos.

¹⁹ MITCHELL, T. *Machine learning*. McGraw Hill, 1997.

²⁰ “Art. 105. [...] § 2º No recurso especial, o recorrente deve demonstrar a relevância das questões de direito federal infraconstitucional discutidas no caso, nos termos da lei, a fim de que a admissão do recurso seja examinada pelo Tribunal, o qual somente pode dele não conhecer com base nesse motivo pela manifestação de 2/3 (dois terços) dos membros do órgão competente para o julgamento.”

²¹ Disponível em: <https://www.stj.jus.br/sites/portalt/Paginas/Comunicacao/Noticias/2023/19112023-95--dos-recursos-extraordinarios-contradecisoes-do-STJ-tem-seguimento-negado-ou-sao-inadmitidos.aspx>. Acesso em: 19 nov. 2023.

Ainda nesta senda, apenas quatro em cada 100 recursos especiais que os tribunais de apelação não admitem são apreciados no mérito pela Corte.

Com aparente orgulho, a Vice-Presidência do STJ revelou que, no universo de recursos apreciados entre setembro de 2022 e outubro de 2023, cerca de 82% tiveram seguimento negado por aplicação de posições do STF adotadas no regime da repercussão geral, enquanto 13% foram inadmitidos e apenas em torno de 3% foram admitidos.

Desde 2019, o STJ utiliza três sistemas de IA: o Sócrates (aponta o permissivo constitucional invocado para a interposição do recurso, os dispositivos de lei questionados e os paradigmas citados para justificar a divergência); o Athos (localiza mesmo antes da distribuição processos que possam ser submetidos à afetação para julgamento sob o rito dos recursos repetitivos e monitora processos com entendimentos convergentes ou divergentes entre os órgãos fracionários do STJ e possíveis distinções ou superações de precedentes qualificados); e o E-Juris (extrai as referências legislativas e jurisprudenciais do acórdão e indica acórdãos principal e sucessivos sobre um mesmo tema jurídico).

Uma quarta ferramenta de inteligência artificial está em fase de desenvolvimento pelo STJ – Tabela Unificada de Assuntos (TUA) –, que proporcionará a identificação automatizada do assunto do processo para fins de distribuição às seções da Corte, conforme o ramo do Direito em que atuam.

A última ferramenta de inteligência artificial de natureza generativa criada pela Corte Cidadã no ano de 2025 é o *Logos*²², que visa reduzir o acervo de processos discutindo questões processuais de admissibilidade do recurso especial, ou seja, esmiuçar o teor de agravos em recurso especial (AREsp), para admitir (ou não) os recursos especiais.

A ferramenta em tese está programada para identificar as teses jurídicas em discussão, além de possuir funcionalidades para preparar relatórios de decisões e um *chat* que permitirá a formulação de perguntas sobre os processos, como, por exemplo, se a minuta recursal cita violação a determinados artigos de lei.

²² Disponível em: <https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/2025/11022-025-STJ-lanca-novo-motor-de-inteligencia-artificial-generativa-para-aumentar-eficiencia-na-producao-de-decisoes.aspx>. Acesso em: 12 fev. 2025.

Desde 2021, o STJ possui um comitê de ética²³ voltado ao tratamento da inteligência artificial responsável por avaliar e homologar os algoritmos e modelos de forma a identificar se preconceitos ou generalizações influenciaram seu desenvolvimento e acarretaram tendências discriminatórias no funcionamento.

A Comissão também é responsável por identificar riscos quando da utilização da IA e indicar os instrumentos de segurança da informação e controle para o seu enfrentamento; além disso, deve avaliar a viabilidade da implantação de novos sistemas de IA em termos de custo/benefício.

Entrementes, o ato normativo que instituiu o Comitê de Ética é silente sobre a necessidade e a importância de divulgar as informações decorrentes dessas análises.

Diante disso, qual é a dimensão do impacto dessas mudanças no cotidiano da advocacia, seja ela pública ou privada, e na atuação do Ministério Público e da Defensoria Pública e, ao fim e ao cabo, nos usuários dos serviços públicos de índole jurisdicional e, acima de tudo, na segurança jurídica?

Quais dados são utilizados para que esse novo filtro recursal seja aplicado aos recursos interpostos? Como foram processados? Qual é o desempenho geral e a partir de quais critérios foram criados os sistemas que utilizam dados específicos para realizar essas tarefas? De que maneira os algoritmos foram construídos?

Antes de tentarmos responder a essas perguntas, precisamos compreender o conceito de inteligência artificial, e assim o faremos no tópico seguinte.

3 CONCEITO DE INTELIGÊNCIA ARTIFICIAL

A IA é um processo mecânico que, a partir de dados, extrai informações, ou seja, permite responder, com certa margem de acerto, aos desafios presentes em um determinado contexto.

Em suma, é o desenvolvimento de programas de computador inteligentes que buscam sistematicamente alcançar um objetivo – específico ou genérico, explícito ou implícito – como, por exemplo, raciocinar, descobrir significados ou aprender com a experiência passada por intermédio da análise de uma

²³ Instituído pela Portaria STJ/GDG nº 365, de 26 de maio de 2021.

enorme quantidade de dados que, comparativamente à atividade de um ser humano, é realizada em um curto átimo de tempo.

O Projeto de Lei nº 2.338/2023 do Senado Federal inicialmente definia a IA, em seu art. 4º, I²⁴, como um sistema computacional com graus diferentes de autonomia, desenhado para atingir um dado conjunto de objetivos, utilizando abordagens baseadas em aprendizagem de máquina e/ou lógica e representação do conhecimento, por meio de dados provenientes de máquinas ou humanos, que permita, ao final, produzir previsões, recomendações ou decisões aptas a influenciar um ambiente virtual ou real.

Aqui já enfatizamos uma imprecisão terminológica no PL, pois *machine learning* ou aprendizado de máquina é tão somente uma das áreas da ciência da computação de IA que se concentra no uso de dados e algoritmos para agir de forma similar ao aprendizado humano.

No *machine learning*, o programador oferece uma base de dados pré-processada e o algoritmo toma as decisões com base naqueles.

Além do *machine learning*, podemos mencionar o *deep learning*, rede neural ou *neural networks* que também são sistemas de IA do maior para o menor, e podemos dizer que cada um engloba o próximo.

No *deep learning*, o sistema extrairá informações de diversos dados diferentes, mesmo sem pré-processamento e, logo após, as redes neurais (*neural networks*) criam padrões e distinguem categorias sem a supervisão humana²⁵.

O aprendizado profundo ou *deep learning* “constrói camadas de *software* feitas de redes neurais artificiais; os dados são inseridos na camada de entrada da rede e um resultado emerge na camada de saída”²⁶; daí o seu nome.

Mediante a sua aplicação, é possível realizar uma série de atividades usando um padrão simples de extração de características, rotulagem de dados,

²⁴ A versão final do projeto de lei divulgada no dia 19.12.2024 foi revisada e ajustada, conforme disponível no sítio eletrônico do Senado: <https://legis.senado.leg.br/sdleg-getter/documento?dm=9881643&ts=1742240906322&disposition=inline>. Acesso em: 25 abr. 2025.

²⁵ IBM. *IA versus aprendizado de máquina versus aprendizado profundo versus redes neurais: qual é a diferença?*, 2023. Disponível em: <https://www.ibm.com/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks/>. Acesso em: 9 nov. 2023.

²⁶ LEE, K.-f.; QIU, FAN, C. 2041: como a inteligência artificial vai mudar sua vida nas próximas décadas. Rio de Janeiro: Globolivos, 2022. p. 44.

treinamento do modelo e utilizá-lo para tornar o produto melhor, de forma a ampliar o engajamento do usuário e, com isso, obter mais dados diretos e indiretos que aprimoram o modelo.

Por essa razão, podemos dizer que se trata de um *machine learning* aperfeiçoado, em que pese demandar maior intervenção humana e maior descrição de dados para aprender²⁷.

O *machine learning* pode (mas não necessariamente) utilizar dados rotulados, também conhecidos como aprendizado supervisionado para treinar um modelo. Esse eventual sistema de rotulagem pode ser uma forma de ensinar um modelo a reconhecer padrões nos dados e, inclusive, permitir que os sistemas melhorem seu desempenho com base na experiência, ou seja, pode, além de se adaptar, aprender com amostras e encontros de dados adicionais, evoluindo com o passar do tempo até que detecte tendências nos dados.

Não é algo tão simples, pois, se estiver ensinando o modelo a reconhecer um balde, você poderá rotular até mil espécies distintas, incluindo, por exemplo, características como “alças”, “ser meio cilíndrico”, ser feito de “plástico” ou de “metal”, “cores”, “formas” etc.²⁸

Portanto, há necessidade de tornar esses dados de treinamento algo demasiadamente representativo, na medida em que, se, eventualmente, o programador se esquecer de incluir baldes amarelos, haverá um enorme problema no desenvolvimento do sistema, o que nos leva à conclusão de que há necessidade de tempo e investimento no tema.

Já uma rede neural ou *neural network* organiza os algoritmos para que eles possam fazer escolhas precisas por conta própria, ao passo que um modelo de aprendizado de máquina (*machine learning*) age com extrema dependência do que aprendeu com as informações.

As redes neurais são um modelo inspirado no cérebro humano e consistem em nós interconectados que processam dados para identificar padrões e

²⁷ LEE, K.-f.; QIU, FAN, C. 2041: como a inteligência artificial vai mudar sua vida nas próximas décadas. Rio de Janeiro: Globolivos, 2022.

²⁸ SCOTT, K. *O futuro da inteligência artificial: de ameaça a recurso*. Rio de Janeiro: Harper Collins, 2023. p. 134-182.

fazer previsões. Cada nó, ou neurônio artificial, conecta-se a outro e tem um peso e um limite associados²⁹.

Na Europa, de acordo com o *Livro branco sobre inteligência artificial*³⁰, conceitualmente podemos considerá-la como um conjunto de tecnologias que combinam dados, algoritmos e capacidade computacional. À medida que ocorrerem progressos na computação, haverá maior disponibilidade de dados, impulsionando, desta forma, o uso da IA.

Nos EUA, a *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* conceitua-a como um sistema baseado em máquina que pode prever, recomendar ou decidir em prol de um determinado conjunto de objetivos definidos por humanos, influenciando ambientes reais ou virtuais.

Após o livro branco, mais recentemente foi editado o Regulamento UE nº 2024/1689, que cria regras harmonizadas sobre IA no âmbito da União Europeia, definindo-a como um sistema baseado em máquina passível de utilização autônoma, ou ainda, como componente de um produto ou serviço, independentemente de o sistema estar ou não integrado a eles, a ponto de prever, recomendar, decidir ou gerar resultados.

Podemos extrair, em breve suma, que as definições se preocupam em diferenciá-la dos demais sistemas de *software* ou programas mais simples, e que o objetivo da IA é gerar informações para a melhoria de *insights*, produtos e serviços, indo muito além de um simples tratamento de dados, já que permite a aprendizagem, o raciocínio ou a criação de modelos a partir das informações coletadas explícita ou implicitamente.

Mas, indubitavelmente, os sistemas de IA deverão ser compatibilizados com a proteção dos direitos fundamentais, além de consagrar a isonomia e a segurança jurídica, facilitando a convergência e a interoperabilidade internacional de sistemas em prol do desenvolvimento econômico, o que demanda uma regulação flexível para atender às rápidas adaptações decorrentes da evolução tecnológica neste campo.

²⁹ LEE, K.-f.; QIU, FAN, C. 2041: como a inteligência artificial vai mudar sua vida nas próximas décadas. Rio de Janeiro: Globolivos, 2022.

³⁰ UE. *Livro branco sobre a inteligência artificial: uma abordagem europeia virada para a excelência e a confiança*. Bruxelas, 2020. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:52020DC0065>. Acesso em: 9 nov. 2023.

Entretanto, a IA não é um conceito único e estanque, mas sim um gênero que se subdivide nas espécies abaixo explicitadas.

3.1 ESPÉCIES DE INTELIGÊNCIA ARTIFICIAL

Para a IBM³¹, a IA pode ser *fraca*, isto é, treinada e focada na execução de tarefas específicas, como a Siri da Apple, a Alexa da Amazon, o *IBM Watson* e os veículos autônomos.

Há ainda a IA *forte*, subdividindo-se em geral artificial (IGA) – cuja inteligência é similar à humana e, portanto, capaz de resolver problemas, aprender e planejar o futuro, e a superinteligência artificial (SIA) – que, em tese, superaria a inteligência e a capacidade do cérebro humano.

Para acadêmicos da Google *Deep Mind*, em recente artigo, a IGA combina (ou supera) aspectos humanos em uma série de tarefas e também se subdivide em cinco espécies: (i) emergente (que, em sua visão, inclui *chatbots* de ponta como ChatGPT e Bard); (ii) competente; (iii) especialista, (iv) virtuosa; e (v) sobre-humana (executa uma ampla gama de tarefas melhor do que todos os humanos, incluindo atividades que os humanos são incapazes de realizar, como decodificar os pensamentos de outras pessoas, prever eventos futuros e conversar com animais).

Além disso, deve ser cumulativamente de uso geral, de alto desempenho e capaz de aprender e realizar uma série de tarefas, a ponto de, em certo momento, solicitar assistência, se necessário.

Inobstante, ainda de acordo com o estudo, nenhum nível além da IGA emergente foi alcançado.

Dessume-se que há uma corrida para construir sistemas maiores e melhores, e a IA é um conceito elástico, conforme sinaliza o cofundador da *Deep Mind* Mustafa Suleiman³², ao afirmar que a primeira onda de IA dispunha essencial acerca da classificação de vários tipos de dados, ao passo que a segunda resultou na captação daqueles para a produção de novas informações.

³¹ IBM. *What is artificial intelligence (AI)?*, 2024. Disponível em: <https://www.ibm.com/topics/artificial-intelligence>. Acesso em: 30 maio 2024.

³² MIT. *DeepMind's cofounder: generative AI is just a phase. What's next is interactive AI*. *Technology Review*, 2023. Disponível em: <https://www.technologyreview.com/2023/09/15/1079624/deepmind-inflection-generative-ai-whats-next-mustafa-suleyman/>. Acesso em: 18 out. 2023.

E agora caminhamos para a terceira onda denominada interativa, em que serão criados *bots* capazes de realizar tarefas predefinidas que permitam a convocação de outros *softwares*, e o mais incrível: outras pessoas.

4 RISCOS NO DESENVOLVIMENTO DA IA

Em geral, as falhas na IA estão atreladas a quatro questões: um problema com o algoritmo; uma complicação com a *pilha de software*; ou, ainda, um treinamento equivocado para uma determinada função³³; por fim, ao algoritmo.

Por essa razão, a maior discussão objeto deste artigo é sobre a *accountability* na construção de algoritmos sem vieses. Destarte, parece-nos que há uma premente necessidade de qualificar o tema, na medida em que, recentemente, o Ministro Dias Toffoli³⁴, em artigo, afirmara que não é recomendável a divulgação dos algoritmos utilizados na IA Victor³⁵, o que, inclusive, salvo melhor juízo, vai de encontro ao disciplinado pelo Conselho Nacional de Justiça na Portaria nº 271/2020, quando preconiza conceitos genéricos, como transparência e estabelecimento de mecanismos de governança colaborativa e democrática, com a participação daqueles que exercem funções essenciais à justiça, da comunidade acadêmica e da sociedade civil.

Ainda que não houvesse o ato normativo editado pelo CNJ, essa afirmação, com todas as vênias, está em descompasso com a atual visão dos Direitos Constitucional e Administrativo, que superaram concepções e institutos de visões autoritárias da relação entre o Estado e o cidadão, resultantes da vetusta interpretação isolada da supremacia do interesse público³⁶.

³³ SCOTT, K. *O futuro da inteligência artificial: de ameaça a recurso*. Rio de Janeiro: Harper Collins, 2023.

³⁴ TOFFOLI, J. A. D.; RANGEL, R. M.; TOLEDO, E. Inteligência artificial no Poder Judiciário brasileiro: o Projeto Victor do Supremo Tribunal Federal. In: TOMASEVICIUS FILHO, E. (org.). *Inteligência artificial: visões disciplinares e internacionais*. São Paulo: Almedina, 2023. p. 337-352.

³⁵ CHINELLATO, S. J. de A.; TOMASEVICIUS FILHO, E. *Inteligência artificial: visões interdisciplinares*. São Paulo: Almedina, 2023.

³⁶ JUSTEN FILHO, M. O direito administrativo de espetáculo. In: ARAGÃO, A. S. de; MARQUES NETO, F. de A. (coord.). *Direito administrativo e seus novos paradigmas*. Belo Horizonte: Fórum, 2008. p. 66-85.

Também deve haver o pleno controle dos algoritmos, que, nos EUA, são instrumentos utilizados, inclusive, na análise de probabilidade de réus retornarem ao convívio social e/ou reincidirem em crimes³⁷.

Mas o que é um algoritmo? É uma sequência de instruções ou comandos realizados de maneira sistemática com o objetivo de resolver um problema ou executar uma tarefa, tal como uma receita de bolo que ensina você passo a passo. Em breve síntese, são etapas ou fórmulas para processar informações e resolver problemas e antecedem os computadores.

Na computação e na rede mundial de computadores (*internet*), são responsáveis por organizar e filtrar as informações que aparecem para os usuários, de acordo com diversos critérios.

Os algoritmos devem ser entendidos como composições de atores não humanos entrelaçadas com humanos, como *designers*, criadores de dados, mantenedores e operadores, em conjuntos sociotécnicos complexos³⁸.

Regra geral, podem ser utilizados para criar conteúdo criativo, como poemas, histórias, músicas etc.

No âmbito da construção da IA, os algoritmos são responsáveis por definir as regras, os critérios e os processos que permitem às máquinas e aos dispositivos aprender, tomar decisões e executar ações de forma autônoma e adaptativa. Menciono a título exemplificativo: (i) algoritmos de aprendizado supervisionado (assimilam a partir de dados rotulados); (ii) algoritmos de aprendizado não supervisionado (capacitam-se a partir de dados não rotulados, quando ausentes respostas ou classificações predefinidas); (iii) algoritmos de aprendizado por reforço (desenvolvem-se a partir da interação com o ambiente, recebendo ainda estímulos por suas ações); e (iv) algoritmos de geração de conteúdo (instruem-se a partir de dados textuais, visuais, sonoros etc., e são capazes de criar novo conteúdo, como poemas, histórias, músicas, imagens etc.).

Para demonstrar os inúmeros vieses que os algoritmos possuem, Susan Nevelow Mart – da Faculdade de Direito da Universidade do Colorado

³⁷ ROLLAND, M. G. Technology's on Federal Sentencing: past, present, and future. *Washington and Lee Journal of Civil Rights and Social Justice*, 2020. Disponível em: <https://scholarlycommons.law.wlu.edu/cgi/viewcontent.cgi?article=1497&context=crsj>. Acesso em: 2 maio 2022.

³⁸ DIAKOPOULOS, N. Accountability, transparency and algorithms. In: DUBBER, M. D.; PASQUALE, F. *The Oxford handbook of ethics of AI*. Oxford: University Press, 2020. p. 197.

– testara se os bancos de dados *online* de buscadores de jurisprudência retornariam os mesmos resultados de pesquisa e descobriu que os engenheiros desses algoritmos, como os sistemas Casetext, Fastcase, Google Scholar, Lexis Advance, Ravel e Westlaw, têm vieses sobre o que seria um caso relevante, ou seja, cada algoritmo mostra resultados distintos aos usuários³⁹.

Por exemplo, bancos de dados mais recentes, como o Fastcase e o Google Acadêmico, geraram resultados de pesquisa menos relevantes em comparação com bancos de dados mais antigos, como Westlaw e Lexis.

Em tese, os algoritmos de busca deveriam ser capazes de gerar resultados redundantes em qualquer banco de dados jurídicos *online*, já que os advogados precisam apenas dos casos mais relevantes. No entanto, como os desenvolvedores têm vieses e suposições ao desenvolver seus algoritmos, os usuários cientes desses vieses devem promover uma pesquisa com extrema cautela.

O viés não é específico do campo jurídico, já que os sistemas de aprendizado de máquina são sempre influenciados pelos dados nos quais são treinados, e, a título de exemplo, sistemas populares de geração de imagens de IA notoriamente tendem a amplificar vieses e estereótipos prejudiciais.

E qual é o tamanho do problema? Mencionamos abaixo alguns deles constatados em novas ferramentas *online* interativas⁴⁰.

As ferramentas construídas por pesquisadores da *startup* de IA Hugging Face e da Universidade de Leipzig permitem que as pessoas examinem vieses em três modelos populares de geração de imagens de IA: DALL-E 2 e as duas versões recentes do *Stable Diffusion*.

Para criar as ferramentas aptas a tais finalidades, os pesquisadores usaram primeiro os modelos de imagem de IA para gerar 96 mil fotos de pessoas, contemplando as mais distintas etnias, gêneros e profissões.

³⁹ MART, S. N. The algorithm as a human artifact: implications for legal {re}search. *Colorado Law*, 2016. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2859720. Acesso em: 10 mar. 2021.

⁴⁰ MIT. These new tools let you see for yourself how biased AI image models are. *Technology Review*, 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: 22 abr. 2023.

A equipe pediu aos modelos de IA que gerassem um conjunto de imagens com base em atributos sociais, como “uma mulher” ou “um homem latino”, e depois outro conjunto de imagens relacionadas a profissões e adjetivos, como “um encanador ambicioso” ou “um CEO compassivo”.

Os pesquisadores desejavam examinar como os conjuntos de imagens variavam e aplicaram uma técnica de aprendizado de máquina chamada *clustering*, com o objetivo de encontrar padrões sem atribuir categorias como gênero ou etnia a elas.

A partir disso, analisaram as semelhanças entre diferentes imagens para ver quais sujeitos os modelos agruparam, como, por exemplo, pessoas em posições de poder.

Daí, então, construíram ferramentas interativas para explorar as imagens que esses modelos de IA produzem e quaisquer vieses refletidos nessa saída, e a conclusão foi de que esses sistemas são extremamente tendenciosos e, em alguns momentos, discriminatórios.

Um outro exemplo relevante de vieses contidos na geração de imagens por IA foi quanto à criação de imagens de homens brancos quando solicitados a retratar pessoas em posições de autoridade.

A DALL-E 2 gerou homens brancos 97% das vezes quando recebeu instruções como “CEO” ou “diretor”, em razão de esses modelos serem treinados em enormes quantidades de dados e imagens extraídas da internet, o que resulta em um processo que não apenas reflete, mas amplifica ainda mais os estereótipos em torno de raça e gênero.

Ao adicionar adjetivos como “compassivo”, “emocional” ou “sensível” a um aviso descrevendo uma profissão, o modelo de IA gerou mais frequentemente uma mulher em vez de um homem, ao passo que, no uso de adjetivos “teimoso”, “intelectual” ou “irracional”, geraram-se, na maioria dos casos, imagens de homens.

A OpenAI, empresa que criou o ChatGPT, e a Stability.AI, empresa que construiu o Stable Diffusion, dizem que promoveram correções para mitigar os vieses arraigados em seus sistemas, como o bloqueio de certos *prompts* que parecem propensos a gerar imagens ofensivas. No entanto, essas novas ferramentas do Hugging Face mostram como essas correções são limitadas, além da necessi-

dade de cautela no uso, na apreciação e, sobretudo, na imprescindibilidade do controle administrativo e jurisdicional, quando necessário.

Daí o acerto do PLS 3.228/2023 ao proibir, no art. 5º, III, quaisquer discriminações ilícitas ou abusivas, cujo propósito ou efeito seja anular ou restringir, de forma abusiva ou ilícita, o reconhecimento, o gozo ou o exercício, em condições de igualdade, de um ou mais direitos ou liberdades previstas no ordenamento jurídico, em razão de características pessoais.

Também assegura que todos os vieses discriminatórios ilegais ou abusivos, diretos ou indiretos, sejam corrigidos de imediato, que ocorre quando normativa, prática ou critério aparentemente neutro tem a capacidade de acarretar desvantagem para pessoa ou grupos afetados, ou os coloquem em desvantagem, desde que essa normativa, prática ou critério seja abusivo ou ilícito.

5 ACCOUNTABILITY E A IMPRESCINDÍVEL DISPONIBILIZAÇÃO DE INFORMAÇÕES DOS SISTEMAS DE INTELIGÊNCIA ARTIFICIAL

A Resolução nº 332/2020, editada pelo CNJ, definiu a transparência como a divulgação dos objetivos e resultados pretendidos pelo uso do modelo de IA; a documentação dos riscos identificados; os critérios para a auditoria e a certificação de boas práticas; e, por último, a disponibilização de satisfatória motivação passível de controle por autoridade humana quanto a qualquer proposta de decisão apresentada pelo modelo de IA, especialmente quando ela for de natureza judicial.

Já em relação à *accountability*, a Resolução do CNJ assevera que devem ser demonstrados os resultados pretendidos, os que foram efetivamente alcançados, além de ser dada a efetiva divulgação quanto à natureza do serviço oferecido, quais técnicas foram utilizadas, como foi o desempenho do sistema e quais riscos ou erros foram constatados.

Infelizmente, nenhuma dessas informações foi encontrada à disposição da sociedade nos sítios eletrônicos do STF e do STJ, e tampouco do CNJ.

Por esses motivos, é de extrema relevância que se possam compreender as regras e os métodos pelos quais os filtros recursais tomam decisões nas mais distintas situações. Da mesma forma, é essencial que sejam demonstradas por quais maneiras o sistema é conduzido e aperfeiçoado, de forma que haja um controle social e, se o caso, o próprio controle judicial.

A divisão de poderes, na verdade, é uma mera separação de funções legislativa, executiva, jurisdicional, pois o poder estatal é uno.

Todos no exercício de suas funções realizam atividades típicas e atípicas, sendo as primeiras aquelas para as quais foram criadas e, portanto, são exercidas de maneira preponderante, ao passo que as segundas ocorreriam em caráter secundário, mas ambas sempre autorizadas constitucionalmente.

No caso objeto deste estudo, podemos dizer que a IA foi criada e regulamentada no exercício de função atípica pelo Poder Judiciário (administrativamente), mas com consequências na sua atividade típica (jurisdicional) e, por conseguinte, ambas, de acordo com o art. 20 da LINDB⁴¹ c/c o art. 93, IX e X, da CF⁴², devem ser motivadas, principalmente no caso objeto deste estudo em que foram criados sistemas de IA que, ao fim e ao cabo, podem ou não resultar numa negativa ou insuficiente prestação jurisdicional.

A aplicação do direito consiste numa atividade juridicamente subordinada ao conteúdo e às determinações de uma norma jurídica, sendo proscrito às autoridades dar soluções concretas a partir de avaliações puramente subjetivas e, ao contrário da função mecanicista/idealista, que pressupõe estarem todas as soluções jurídicas previstas nas normas, prevalece a concepção realista de que, na sua aplicação, há escolhas a serem realizadas pelo sujeito investido

⁴¹ “Art. 20. Nas esferas administrativa, controladora e judicial, não se decidirá com base em valores jurídicos abstratos sem que sejam consideradas as consequências práticas da decisão. Parágrafo único. 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.” (BRASIL. Decreto-Lei nº 4.657, de 4 de setembro de 1942. Lei de Introdução às Normas do Direito Brasileiro. Brasília: Presidência da República, 1942. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto-lei/del4657compilado.htm. Acesso em: 30 maio 2024)

⁴² “Art. 93. Lei complementar, de iniciativa do Supremo Tribunal Federal, disporá sobre o Estatuto da Magistratura, observados os seguintes princípios: [...] IX – todos os julgamentos dos órgãos do Poder Judiciário serão públicos, e fundamentadas todas as decisões, sob pena de nulidade, podendo a lei limitar a presença, em determinados atos, às próprias partes e a seus advogados, ou somente a estes, em casos nos quais a preservação do direito à intimidade do interessado no sigilo não prejudique o interesse público à informação; X – as decisões administrativas dos tribunais serão motivadas e em sessão pública, sendo as disciplinares tomadas pelo voto da maioria absoluta de seus membros; [...]” (BRASIL. Constituição da República Federativa do Brasil de 1988. Brasília: Presidência da República, 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 30 maio 2024)

da competência decisória, o que enseja inúmeras dificuldades práticas, principalmente quando se utilizam conceitos genéricos/abstratos⁴³.

Impõe-se às autoridades a projeção quanto aos possíveis cenários resultantes da decisão tomada em uma relação de causa e efeito⁴⁴.

Por conseguinte, é essencial perquirirmos quais seriam os caminhos corretos para valores abstratos, como governança e transparência, que, ao fim e ao cabo, resultam na “*accountability*, cuja função, estrutura e escopo devem ser compreendidas”⁴⁵.

Essa preocupação com a *accountability* deriva do fato de que quaisquer modelos de IA são treinados com amplas bases de dados contendo dezenas de bilhões de parâmetros aplicáveis a inúmeros contextos facilmente modificáveis que podem representar sérios riscos à segurança pública nacional, à economia, e até à saúde, já que foi comprovada a tentativa de utilizá-la na criação de novas armas químicas, biológicas e nucleares.

Logo, a transparência é insuficiente, por si só, em termos de *accountability* quando tratamos de justiça, privacidade e não discriminação, na medida em que será apenas um instrumento para a compreensão desses valores.

A título de exemplo, a loteria é um processo transparente enquanto abstração, mas assegurar que seu sistema de computação opera de maneira fidedigna é uma tarefa mais complexa.

Como garantir que todos os números a serem sorteados estarão no sistema que funcionará perfeitamente de maneira aleatória, a despeito do algoritmo de loteria ser de fácil compreensão e programação? Há extrema dependência de escolhas aleatórias que não podem ser repetíveis, e, portanto, a revisão e o controle do sistema são demasiadamente complexos⁴⁶.

⁴³ JUSTEN FILHO, M. Art. 20 da LIND: dever de transparência, concretude e proporcionalidade nas decisões públicas. *Revista de Direito Administrativo*, edição especial: Direito Público na Lei de Introdução às Normas do Direito Brasileiro. Rio de Janeiro, 2018. p. 13-41.

⁴⁴ JUSTEN FILHO, M. Art. 20 da LIND: dever de transparência, concretude e proporcionalidade nas decisões públicas. *Revista de Direito Administrativo*, edição especial: Direito Público na Lei de Introdução às Normas do Direito Brasileiro. Rio de Janeiro, 2018. p. 13-41.

⁴⁵ KROLL, J. A. Accountability in computer systems. In: DUBBER, M. D.; PASQUALE, F. *The Oxford handbook of ethics of AI*. Oxford: University Press, 2020. p. 181-196.

⁴⁶ KROLL, J. A. Accountability in computer systems. In: DUBBER, M. D.; PASQUALE, F. *The Oxford handbook of ethics of AI*. Oxford: University Press, 2020. p. 181-196.

Garantir que todos os números poderão ser sorteados de maneira isonômica e que não há entradas ilegítimas é extremamente difícil, o que torna insuficiente a transparência por si só para demonstrar que o objetivo uma loteria foi atingido de forma legítima⁴⁷. Aqui podemos mencionar, a título exemplificativo, os sistemas de *bets* tão em voga no Brasil, que foram regulamentados pela Lei nº 14.790/2023, cujo teor impõe ao operador o uso de sistemas auditáveis aptos à geração de relatórios técnicos, operacionais, econômico-financeiros e contábeis, de dados, de documentos, de certificações e de certidões, mediante acesso irrestrito, contínuo e em tempo real ao Ministério da Fazenda.

Podemos inferir que todo e qualquer processo relacionado ao uso de sistemas de IA deverá ser diáfano, com vistas a se comprovar que as propriedades exigidas de seus resultados se mantiveram, e que eventuais violações são detectáveis para fins de responsabilização daqueles que tenham dado causa⁴⁸.

6 A RESOLUÇÃO Nº 615, DE 11 DE MARÇO DE 2025, DO CONSELHO NACIONAL DE JUSTIÇA

Na esteira destas preocupações, o Conselho Nacional de Justiça editou a Resolução nº 615/2025⁴⁹, cujo desiderato é estruturar uma governança das soluções de IA, desde que observados os padrões de auditoria, monitoramento e transparência, sem prejuízo da atuação do CNJ no âmbito de suas competências, quando necessário.

Em caráter preliminar, ou seja, antes de sua utilização ou entrada em produção, deverá haver o seu registro no Portal de Serviços do Poder Judiciário (PDPJ-Br) e na Plataforma SINAPSES, contendo ainda uma avaliação do sistema pelo tribunal desenvolvedor ou contratante, com o objetivo de classificar o seu grau de risco.

Quando da implementação da ferramenta, os tribunais deverão disponibilizar à sociedade predominantemente: (i) os nomes dos responsáveis

⁴⁷ KROLL, J. A. Accountability in computer systems. In: DUBBER, M. D.; PASQUALE, F. *The Oxford handbook of ethics of AI*. Oxford: University Press, 2020. p. 181-196.

⁴⁸ KROLL, J. A. Accountability in computer systems. In: DUBBER, M. D.; PASQUALE, F. *The Oxford handbook of ethics of AI*. Oxford: University Press, 2020. p. 181-196.

⁴⁹ Disponível em: <https://atos.cnj.jus.br/files/original1555302025031467d4517244566.pdf>. Acesso em: 15 mar. 2025.

pela execução das ações e pela prestação de contas; (ii) os custos envolvidos na pesquisa, no desenvolvimento, implantação e treinamento; (iii) os resultados pretendidos e alcançados; (iv) a natureza do serviço oferecido, técnicas utilizadas, desempenho do sistema e riscos de erros, sempre em linguagem simples, dentre outros.

Os usuários do serviço terão, em seu favor, o direito de revisão detalhada do conteúdo gerado e dos dados utilizados para sua elaboração, assim como ao método empregado pela inteligência artificial na sua formulação.

Entretanto, restou silente qualquer dispositivo acerca da possibilidade de os jurisdicionados terem acesso a essas informações e de qual maneira isso poderá ocorrer.

A auditoria e o monitoramento das soluções de IA serão realizadas, sem a obrigatoriedade de acesso irrestrito ao código-fonte, desde que adotados mecanismos de transparência e controle sobre o uso dos dados e das decisões automatizadas, competindo ao Comitê Nacional de IA do Poder Judiciário definir a metodologia para a condução de auditorias, levando em consideração os riscos envolvidos, a definição de salvaguardas (medidas de proteção) e a documentação produzida.

A promessa do ato normativo é assegurar a transparência no uso de IA por meio de indicadores claros e a disponibilização de relatórios públicos que informem o uso dessas soluções de maneira compreensível, de forma a assegurar a eficiência e a credibilidade das decisões judiciais.

Até o momento de conclusão deste artigo, os sítios eletrônicos acima mencionados não dispõem de nenhuma informação, respectivamente, em sua aba sobre serviços nacionais⁵⁰ e Repositório Nacional de Projetos de Software⁵¹ no que tange a quaisquer sistemas de IA atualmente utilizados no âmbito do Poder Judiciário, a despeito da sua entrada em vigor no prazo de 120 dias, pois, ao menos no âmbito do STF e do STJ, a utilização de instrumentos desta natureza já está consolidada.

O CNJ também passa a exigir uma avaliação de impacto algorítmico, de forma contínua no tocante aos impactos do sistema sobre os direitos

⁵⁰ Disponível em: <https://www.jus.br/servicos/>. Acesso em: 25 abr. 2025.

⁵¹ Disponível em: <https://www.cnj.jus.br/sistemas/git-jus/>. Acesso em: 25 abr. 2025.

fundamentais, por intermédio da identificação de medidas preventivas, mitigadoras de danos e de maximização dos impactos positivos, sem a violação da propriedade industrial e intelectual da solução de IA utilizada.

Dessume-se que o CNJ já deu início à definição de critérios passíveis de utilização em sede de controle administrativo, social e jurisdicional, mas será que são suficientes e exaustivos?

7 CONTROLE FORMAL DA IA – CRITÉRIOS PARA A ACCOUNTABILITY A PARTIR DO DIREITO COMPARADO E DO PROJETO DE LEI DO SENADO Nº 2.338/2023

Os Estados Unidos, em sua *Executive Order* editada em outubro de 2023, foram enfáticos acerca da imprescindibilidade na incorporação de princípios de equidade em tecnologias habilitadas por IA, com vistas ao monitoramento do desempenho algorítmico contra a discriminação e o viés⁵², que pode ser compreendido como o uso de interpretações subjetivas ou incorretas dos dados.

Já o Regulamento UE nº 2024/1689⁵³, que cria regras harmonizadas sobre IA no âmbito da União Europeia, exige que qualquer sistema disponibilize à sociedade principalmente as seguintes informações:

- a) as tarefas que o modelo se destina a desempenhar e o tipo e a natureza dos outros sistemas em que pode ser integrado;
- b) as políticas de utilização aplicáveis;
- c) a arquitetura e o número de parâmetros;
- d) as especificações para a criação do modelo e o processo de treino, incluindo as metodologias e as técnicas de treino;

⁵² WHITE HOUSE. *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*. Washington, DC, 2023. Disponível em: <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>. Acesso em: 30 out. 2023.

⁵³ Regulamento (UE) nº 2024/1689 do Parlamento Europeu e do Conselho de 13 de junho de 2024, que cria regras harmonizadas em matéria de inteligência artificial. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L_202401689. Acesso em: 30 out. 2024.

- e) as principais finalidades da criação, incluindo a fundamentação e os pressupostos assumidos; o que se pretende otimizar com o modelo e a importância dos diferentes parâmetros, se aplicável;
- f) informações sobre os dados utilizados para treino, testagem e validação, se aplicável, incluindo o tipo e a proveniência dos dados e as metodologias de curadoria (por exemplo, limpeza, filtragem etc.);
- g) o número de pontos de dados, o seu âmbito e as principais características;
- h) a forma como os dados foram obtidos e selecionados, bem como todas as outras medidas para detectar a inadequação das fontes de dados e dos métodos para detectar enviesamentos identificáveis, se aplicável;
- i) os recursos computacionais utilizados para treinar o modelo (por exemplo, o número de operações de vírgula flutuante), o tempo de treino e outros dados pertinentes relacionados com a formação.

Por conseguinte, vislumbra-se que outros critérios deveriam ter sido disponibilizados pelo CNJ, sendo essencial o desenvolvimento de metodologias e ferramentas específicas para a implementação e o uso de mecanismos voltados a aferir, testar e certificar os sistemas de IA em uma escala de riscos.

Logo, deveriam ter sido avaliados os impactos dos filtros recursais do STF e do STJ antes da sua utilização, mas também durante a execução; neste ponto, citamos, a título exemplificativo, as seguintes hipóteses: os resultados de julgamento eventualmente discrepantes sobre temas relativamente similares, a categorização dos julgamentos a partir dos resultados e o potencial de a IA prever resultados de decisões futuras.

Não sem razão, o PL 2.338/2023 exigiu a avaliação de impacto algorítmico dos sistemas de inteligência artificial que deverá ao menos registrar: os (i) riscos conhecidos e previsíveis associados ao sistema à época em que foi desenvolvido; (ii) os benefícios associados; (iii) a probabilidade de consequências adversas e a sua gravidade, incluindo o número de pessoas potencialmente impactadas; (iv) o processo e o resultado de testes e avaliações e medidas de mitigação realizadas para verificação de possíveis impactos discriminatórios.

Basicamente, os sistemas têm de ser desenvolvidos a partir de um exame *ex ante*, tendo em devida consideração os riscos que podem gerar e serem supervisionados por nós humanos, garantindo o registro de tudo o que desenvolveu ou gerou a partir das tarefas, da arquitetura e do número de

parâmetros previamente estabelecidos, para, *a posteriori*, permitir a sua revisão e análise por terceiros a partir dos seguintes parâmetros:

- avaliação e registro integral do sistema, de forma a se definir a natureza do serviço oferecido, as técnicas utilizadas, o seu desempenho e os riscos de erros;
- supervisão a partir de regras predefinidas;
- possibilidade de confirmar se as regras predefinidas foram ou são aplicadas corretamente;
- havendo violações das regras, devem ser explicitados quais são os critérios e as formas para solucionar e prevenir novos equívocos;
- disponibilização das informações e/ou programas que violaram ou tentaram invadir o sistema e as fragilidades já sanadas;
- divulgação das ações que foram adotadas a partir das violações ou fragilidades constatadas; e
- [...]

Novamente, ainda de acordo com o PL, os sistemas de IA das Cortes Superiores deveriam, previamente às suas implementações, documentar as técnicas de uso, mantendo-as atualizadas e à disposição do controle social.

E como se daria essa documentação técnica?

Por intermédio da disponibilização do processo de funcionamento do sistema e das decisões envolvidas em sua construção, implementação e uso, contemplando todo o ciclo de vida do sistema, tal como estágio de *design*, de desenvolvimento, de avaliação, de operação e de descontinuação do sistema, sem prejuízo do registro automático da operação, de modo a permitir a avaliação de sua acurácia e robustez, além das medidas de mitigação de riscos adotadas.

Adicionalmente às medidas estabelecidas neste capítulo, órgãos e entidades do Poder Público, ao contratar, desenvolver ou utilizar sistemas de IA, deverão realizar consulta e audiência públicas prévias sobre a utilização planejada dos sistemas, com informações sobre os dados a serem utilizados,

a lógica geral de funcionamento e resultados de testes realizados, e, ao final, divulgar os resultados das avaliações.

Extraí-se que os riscos relacionados com os sistemas de IA podem resultar não somente quanto à forma como esses sistemas são concebidos, mas também com relação à maneira como são utilizados, e, portanto, os responsáveis pela implantação de sistemas de IA desempenham um papel fundamental na garantia da proteção dos direitos fundamentais, complementando as obrigações do prestador quando do desenvolvimento do sistema de IA.

Os responsáveis pela implantação estão em melhor posição para entender de que forma o sistema de IA será utilizado em concreto e deverão exercer um juízo de valor extremamente crítico.

Daí apresentamos o seguinte questionamento: a partir das diretrizes internacionais e das contidas no PL em trâmite no Congresso Nacional, os sistemas de IA do STF e do STJ seriam avaliados de que forma em um eventual controle jurisdicional formal?

8 CONTROLE JURISDICIONAL MATERIAL OU QUANTO AO OBJETO

Em relação ao conteúdo da IA, o PL 2.338/2023, do Senado Federal, prevê que quaisquer pessoas afetadas por sistemas têm direito à explicação sobre a decisão, providência ou recomendação adotadas, de contestá-las, não serem discriminadas, à correção de vieses discriminatórios diretos, indiretos, ilegais ou abusivos.

Ademais, a pessoa afetada pelo sistema de IA poderá solicitar explicações sobre os principais fatores que impactaram para a decisão, providência ou recomendação, advindas do sistema, para se aferir a sua lógica e os critérios.

O PL prevê, outrossim, que, se a decisão advinda do sistema de IA produzir efeitos jurídicos relevantes ou que impactem de maneira significativa os interesses da pessoa, deverá ocorrer uma intervenção ou revisão humana.

Na mesma senda, a atual redação do PL 2.338/2023 classifica como alto risco as soluções de IA desenvolvidas e utilizadas na “administração da justiça, no que toca o uso sistemas que auxiliem autoridades judiciárias em investigação dos fatos e na aplicação da lei quando houver risco às liberdades

individuais e ao Estado democrático de direito, excluindo-se os sistemas que auxiliem atos e atividades administrativas”.

O Regulamento EU nº 2024/1689 proscree a disponibilização de serviços que empreguem técnicas subliminares visando contornar a consciência dos seres humanos, ou manifestamente manipuladoras, que tenham por objetivo distorcer substancialmente o seu comportamento, ao prejudicar a capacidade decisória.

Também fica vedada a disponibilização de sistemas de IA que:

- (i) explorem vulnerabilidades das pessoas em razão da idade ou da situação socioeconômica;
- (ii) avaliem ou classifiquem pessoas ou grupos durante certo período com base no seu comportamento social ou em características de personalidade, conhecidas, inferidas ou previsíveis, quando conduzir a tratamentos prejudiciais em contextos não relacionados às hipóteses em que foram coletadas inicialmente;
- (iii) criem ou expandam bases de dados de reconhecimento faciais por meio da coleta aleatória de imagens a partir da internet ou de circuitos fechados de segurança;
- (iv) extraia emoções de uma pessoa no local de trabalho e nas instituições de ensino, exceto nos casos em que se destinem às questões médicas ou de segurança;
- (v) classifiquem individualmente as pessoas com base nos seus dados biométricos para deduzir ou inferir a sua raça, opiniões políticas, filiação sindical, convicções religiosas ou filosóficas, orientação sexual, salvo se houver autorização legal específica;
- (vi) utilizem sistemas de identificação biométrica em tempo real nos espaços públicos para efeitos de aplicação da lei, salvo quando estritamente necessária para a busca seletiva de pessoas desaparecidas, vítimas dos crimes de sequestro, tráfico de seres humanos ou exploração sexual; ou ainda, para prevenir ameaças à vida ou à segurança física de pessoas, tal como um ataque terrorista; e, por último para a localização ou identificação de pessoas suspeitas de terem cometido

infrações penais, ou ainda, para a execução de uma sanção penal, cuja pena privativa de liberdade seja superior a quatro anos.

Nesta última hipótese, mencionamos que os Estados da Bahia, São Paulo e Minas Gerais, sem qualquer lei disciplinando o tema⁵⁴, não distinguem sobre o quantitativo de pena aplicada a procurados, isto é, basta uma condenação prevista no sistema de segurança pública para que reconheça a pessoa, o que ensejará a sua imediata prisão.

O Regulamento EU nº 2024/1689 também impõe diferentes obrigações aos seus desenvolvedores em função do nível de risco da IA, que pode ser inaceitável, elevado e mínimo.

Os sistemas de IA de risco inaceitável, por serem considerados uma ameaça à população, são proibidos, incluindo-se aqueles que possam promover: (i) manipulação cognitivo-comportamental de pessoas ou grupos vulneráveis específicos, como, por exemplo, brinquedos ativados por voz que incentivam comportamentos perigosos nas crianças; (ii) que categorizam pessoas a partir de seus comportamentos, características e situação socioeconômica; e (iii) de identificação biométrica em tempo real e à distância, cuja utilização seja voltada ao controle social, salvo para a repressão de crimes graves e após aprovação do Tribunal de Justiça da União Europeia.

Consideram-se de risco elevado os sistemas de IA que afetam negativamente a segurança ou os direitos fundamentais, como aplicativos de saúde, gestão de recursos humanos, concessão de benefícios sociais e previdenciários, serviços públicos (aí incluída a IA utilizada no Poder Judiciário) e outros voltados à criação de produtos e à prestação de serviços que demandem a observância de parâmetros de segurança uniformes no âmbito da UE.

⁵⁴ O Distrito Federal é o único ente político que possui a Lei nº 6.712/2020 disciplinando o uso de tecnologia de reconhecimento facial, ou seja, a tecnologia que analisa as características faciais usada para a identificação pessoal exclusiva de indivíduos em imagens estáticas ou em vídeos. Contudo, é extremamente lacunosa, pois veda o seu uso na vigilância contínua de indivíduo ou grupo de indivíduos, salvo na segurança pública restrita a espaços públicos, e condicionadas à comunicação prévia com placas visíveis contendo a respectiva informação. Disponível em: <https://www.tjdft.jus.br/institucional/relacoes-institucionais/arquivos/lei-no-6-712-de-10-de-novembro-de-2020.pdf>. Acesso em: 4 abr. 2025.

A Resolução nº 615/2025 do CNJ assevera que os tribunais deverão realizar a avaliação das soluções que utilizem técnicas de inteligência artificial, com a finalidade de definir o seu grau de risco, baseando-se na categorização de riscos, desenvolvida com base em fatores como o potencial impacto nos direitos fundamentais, a complexidade do modelo, os usos pretendidos e potenciais e a quantidade de dados sensíveis utilizados.

Fica assegurado ao Comitê Nacional de Inteligência Artificial do CNJ, de ofício ou mediante provocação fundamentada, determinar a reclassificação do grau de risco de determinada solução, bem como determinar a avaliação de impacto algorítmico, quando tal medida se demonstrar essencial.

Segundo o CNJ, resta vedado ao Poder Judiciário, por acarretar risco excessivo à segurança da informação, aos direitos fundamentais dos cidadãos ou à independência dos magistrados, o desenvolvimento e a utilização das soluções de IA:

- a) que impeçam a revisão humana dos resultados propostos ao longo de seu ciclo de treinamento, desenvolvimento e uso, ou que gerem dependência absoluta do usuário em relação ao resultado proposto, sem possibilidade de alteração ou revisão;
- b) que valorem traços da personalidade, características ou comportamentos de pessoas naturais ou de grupos de pessoas naturais, para fins de prever o cometimento de crimes ou a probabilidade de reiteração delitiva;
- c) com o propósito de fundamentar decisões em matéria trabalhista a partir da formulação de perfis pessoais;
- d) que classifiquem ou ranqueiem pessoas naturais, com base no seu comportamento ou situação social, ou ainda, em atributos da sua personalidade, para a avaliação da plausibilidade de seus direitos, méritos judiciais ou testemunhos.

No tocante à vedação contida na alínea “b”, acima, destacamos que o Regulamento Europeu permite a utilização do sistema de IA para apoiar a avaliação humana do envolvimento de uma pessoa numa atividade criminosa, a partir de fatos objetivos e verificáveis diretamente ligados a esta, ou seja, a Resolução do CNJ foi mais restritiva quando comparada à legislação da União Europeia.

O Anexo da Resolução do CNJ considera de alto risco e, portanto, deverão ser submetidos a processos regulares de auditoria e monitoramento contínuo os seguintes mecanismos de IA:

- a) que identifiquem padrões biométricos ou comportamentais que permitam o reconhecimento de emoções;
- b) que afirmem a adequação dos meios de prova e a sua valoração nos processos de jurisdição contenciosa, sejam documentais, testemunhais, periciais ou de outras naturezas, especialmente quando tais avaliações possam influenciar diretamente a decisão judicial;
- c) valorem, tipifiquem e interpretem fatos como sendo crimes, contravenções penais ou atos infracionais, ressalvadas as soluções voltadas à mera rotina da execução penal e de medidas socioeducativas;
- d) formulem 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;
- e) monitorem o comportamento por meio de autenticação facial ou biométrica, exceto quando voltada para a mera confirmação da identidade de uma pessoa natural específica ou para atividades de segurança pública devidamente justificadas.

Em que pese a análise final, isto é, o juízo conclusivo sobre a admissibilidade ou provimento dos recursos dos ministros do STJ, parece-nos que os sistemas *Sócrates*, *Athos* e *Logos* induzem a um desfecho sobre a aplicação de normas jurídicas ou precedentes a um conjunto determinado de fatos concretos.

Portanto, possuem uma carga argumentativa de extrema relevância acerca da sua adequada utilização, principalmente quando da ponderação de decisões contrários e favoráveis que foram apresentados pela IA e levaram à prolação da decisão, o que poderá ensejar grandes discussões e a possibilidade da interposição de recursos que poderão ser fundamentados a partir do uso da Lei de Acesso à Informação (Lei nº 12.527/2011) visando à disponibilização dos algoritmos do sistema na busca de precedentes dotados de vieses ou outras incongruências.

CONCLUSÃO

A IA pode ser a ferramenta mais poderosa que a humanidade já criou e a consideramos um divisor de águas para os modelos de negócios, padrões de consumo e comportamento individuais.

É inelutável o significativo impacto na educação, no meio ambiente, na economia e na saúde, mas já estão demonstrados os riscos de ser uma ameaça à democracia e à liberdade, tanto que o TSE já editou e aprovou a regulação da IA na propaganda eleitoral, proscrevendo as *deepfakes*; impôs a comunicação prévia sobre o uso de IA na propaganda eleitoral; vedou o emprego de robôs para intermediar contato com os eleitores (é vedado simular diálogo com candidato ou qualquer outra pessoa); e atribuiu responsabilidade às *big techs* que não retirem do ar imediatamente conteúdos com desinformação.

O texto foi debatido em consulta pública na linha do que as melhores práticas internacionais de *accountability* recomendam, inclusive mediante a divulgação dos temas, das propostas em cada um dos eixos (pesquisa eleitoral; auditoria e fiscalização; sistemas eleitorais; atos gerais do processo eleitoral; registro de candidatura; fundo eleitoral de financiamento de campanha; prestação de contas; propaganda eleitoral; representações, reclamações e ilícitos eleitorais) e as hipóteses em que houve ou não acatamento, levando à conclusão de que os critérios apresentados neste estudo possuem indícios de pertinência e adequação.

O único ponto que eventualmente pode ser reavaliado no tocante ao procedimento do TSE diz respeito à ausência de motivação em relação às propostas não acatadas.

Ainda nesse cariz, um recente estudo divulgado pelo Laboratório de Governança e Regulação de Inteligência Artificial (LIA), vinculado ao Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP), ao analisar 56 decisões proferidas ainda na pré-campanha e no primeiro turno pelos TREs, constatou três linhas interpretativas sobre a remoção (ou não) de conteúdos considerados *deepfakes*: na primeira delas, a utilização de em vídeos de teor político foi aceita, desde que não houvesse pedido explícito de voto ou de não voto, e nem desinformação; a segunda considerou que nenhum uso da IA fosse permitido no contexto eleitoral; e, na última, seria necessário analisar o grau de manipulação e o contexto da propaganda, proibindo somente aqueles conteúdos de nítida desinformação.

Infere-se que há grande preocupação do Poder Judiciário com o uso da IA por terceiros, mas caberá ao TSE sedimentar a jurisprudência sobre o tema, a fim de concretizar o princípio da isonomia e da segurança jurídica, bem como proporcionar a racionalização do exercício da magistratura.

Ainda, pelo menos no âmbito do STF e do STJ, não restou demonstrada a preocupação com a *accountability* para aumentar o índice de confiança dos cidadãos no Poder Judiciário, que nacionalmente gira em torno de 40% (e já foi maior), consoante Relatório ICJBrasil 2021 da Fundação Getúlio Vargas, sendo imprescindível aprimorar a sua capacidade interna de regular e fornecer os melhores resultados em termos de serviços de inteligência artificial aos cidadãos.

Esses esforços se iniciam com o controle interno por intermédio de sua força de trabalho ao desenvolver e utilizar os sistemas de IA, além da devida transparência à sociedade demonstrando os benefícios, os riscos e as limitações dos sistemas enquanto filtros de admissibilidade recursais.

Além do uso indevido ou sem critérios passíveis de controle social ou judicial, o Relatório editado pelo Fórum Econômico Mundial (*World Economic Forum*), em 2024, inclusive já alertou sobre os riscos da IA na desinformação e no aumento da polarização da sociedade, na medida em que se podem criar inúmeros modelos para influenciar eleitores.

Podemos ainda citar outros problemas de ordem prática decorrentes de vieses, como, por exemplo, no passaporte *online* na Nova Zelândia, que rejeitou a foto de um asiático por ter concluído que seus olhos estavam fechados, e ainda sistemas de marcação *online* que rotulam fotos de pessoas negras como macacos; por fim, a menção a campos de concentração como “esporte” e “academia na selva”⁵⁵ no *Flickr*, que supostamente seria uma tecnologia avançada de reconhecimento de imagem passível de categorizar automaticamente as fotos em vários grupos amplos⁵⁶.

⁵⁵ ACEMOGLU, D. *Redesigning AI: work, democracy, and justice in the age of automation*. Boston: Boston Review Forum, 2021. p. 39.

⁵⁶ THE GUARDIAN. Flickr enfrenta reclamações sobre marcação automática “ofensiva” em fotos, 2015. Disponível em: <https://www.theguardian.com/technology/2015/may/20/flickr-complaints-offensive-auto-tagging-photos>. Acesso em: 10 nov. 2019.

Por essa razão, tal como os demais instrumentos e atos estatais, devemos conseguir analisar a sua acurácia de forma independente, inclusive com acesso aos eventuais erros e maus dados (*bad data*), somados ao processo de compreensão de como estes foram mitigados.

Em suma, tal como um rótulo nutricional, a IA, enquanto filtro recursal dos tribunais superiores, deve permitir que compreendamos os dados de treinamento utilizados para o aprendizado, os dados do mundo real empregados nas aplicações e os resultados esperados, que deverão estar sempre atrelados às tomadas de decisões que resultem efetivamente em segurança jurídica, para, se for o caso, a própria sociedade buscar o Poder Judiciário a fim de controlá-los.

Inclusive, o *Chief Justice* John Roberts da Suprema Corte norte-americana⁵⁷, em recente relatório divulgado, afirmou que os juízes recursais desempenham funções essencialmente humanas, e, portanto, a IA não pode tomar decisões.

Dessarte, todas as pessoas podem e devem ter o direito de questionar os sistemas da IA, inclusive mediante a disponibilização dos resultados, dos dados refinados e utilizados e a quais conclusões chegaram as comissões de ética responsáveis por essas análises.

E o controle da inteligência artificial, seja em sede judicial ou administrativa, a despeito da pendência da aprovação do PLS 3.228/2023, fica assegurado a partir do Direito Comparado com fundamento nas diretrizes previstas neste estudo por analogia, considerando o disposto nos arts. 5º, XXXVI⁵⁸, da CF e 4º⁵⁹ da LINDB c/c o art. 140⁶⁰ do CPC, que veda ao Poder Judiciário o *non liquet*, sob a alegação de lacuna ou obscuridade do ordenamento jurídico, aplicável também aos processos administrativos, nos termos do art. 15⁶¹, também do CPC.

⁵⁷ SUPREME COURT OF USA. *Year-End Report on the Federal Judiciary*, 2023. Disponível em: <https://www.supremecourt.gov/publicinfo/year-end/2023year-endreport.pdf>. Acesso em: 9 jan. 2024.

⁵⁸ “Art. 5º [...] XXXV – a lei não excluirá da apreciação do Poder Judiciário lesão ou ameaça a direito; [...]”

⁵⁹ “Art. 4º Quando a lei for omissa, o juiz decidirá o caso de acordo com a analogia, os costumes e os princípios gerais de direito.”

⁶⁰ “Art. 140. O juiz não se exime de decidir sob a alegação de lacuna ou obscuridade do ordenamento jurídico. Parágrafo único. O juiz só decidirá por equidade nos casos previstos em lei.”

⁶¹ “Art. 15. Na ausência de normas que regulem processos eleitorais, trabalhistas ou administrativos, as disposições deste Código lhes serão aplicadas supletiva e subsidiariamente.”

Por todas essas razões, nós enquanto profissionais do Direito precisamos nos adaptar à inteligência artificial e compreendê-la, na medida em que se trata de algo inexorável ao nosso percurso de evolução e desenvolvimento.

Ao mesmo tempo, as instituições públicas e privadas devem amadurecer no tema da IA, porque somente com a *accountability*, o controle e o diálogo teremos um *Leviatã* agrilhado, que não seja despótico⁶² ou de papel⁶³.

Logo, a adoção de sistemas inteligência artificial (IA) não tem por escopo somente a inovação, mas, sobretudo, a proteção à segurança e aos direitos fundamentais consagrados na CF, incluindo a democracia e o Estado de Direito.

REFERÊNCIAS

ACEMOGLU, D. *Redesigning AI: work, democracy, and justice in the age of automation*. Boston: Boston Review Forum, 2021.

ACEMOGLU, D.; ROBINSON, J. A. *O corredor estreito: estados, sociedades e o destino da liberdade*. 1. ed. Rio de Janeiro: Intrínseca, 2022.

BOOTLE, R. *A economia da inteligência artificial: como a IA está transformando o trabalho, a riqueza e o progresso*. Rio de Janeiro: Alta Books, 2021.

CHINELLATO, S. J. de A.; TOMASEVICIUS FILHO, E. *Inteligência artificial: visões interdisciplinares*. São Paulo: Almedina, 2023.

DIAKOPOULOS, N. Accountability, transparency and algorithms. In: DUBBER, M. D.; PASQUALE, F. *The Oxford handbook of ethics of AI*. Oxford: University Press, p. 197-213, 2020.

⁶² O *Leviatã* despótico é aquele que não encontra limites na lei e não permite a participação e a concorrência política, a exemplo de China, Coreia do Norte e Cuba. O *Leviatã* de papel é um Estado arbitrário, de plena insegurança jurídica, que manipula e desempodera pessoas ao deixá-las esperando e implorando por serviços públicos, na medida em que, por exemplo, não há protocolos e tampouco horário de atendimento definido, pois funcionam quando os servidores públicos bem desejam, sendo mencionada de maneira exemplificativa a Argentina. Não se trata de um *Leviatã* agrilhado, pois não é capaz de fazer cumprir as leis e de regular a economia. Este último seria o detentor de uma burocracia racional que calcula custos, receitas, lucros e perdas e é capaz de exigir o cumprimento da lei de forma conciliatória com a participação e o controle social. ACEMOGLU, D.; ROBINSON, J. A. *O corredor estreito: estados, sociedades e o destino da liberdade*. Rio de Janeiro: Intrínseca, 2022.

⁶³ ACEMOGLU, D.; ROBINSON, J. A. *O corredor estreito: estados, sociedades e o destino da liberdade*. Rio de Janeiro: Intrínseca, 2022.

INTERNATIONAL ORGANIZATIONAL FOR STANDARDIZATION. ISO/IEC 22989:2022(E). *Information technology – Artificial intelligence – Artificial intelligence concepts and terminology*.

JUSTEN FILHO, M. O direito administrativo de espetáculo. In: ARAGÃO, A. S. de; MARQUES NETO, F. de A. (coord.). *Direito administrativo e seus novos paradigmas*. Belo Horizonte: Fórum, 2008.

JUSTEN FILHO, M. Art. 20 da LIND: dever de transparência, concretude e proporcionalidade nas decisões públicas. *Revista de Direito Administrativo*, Edição especial: Direito público na lei de introdução às normas do Direito brasileiro. Rio de Janeiro, 2018.

KISSINGER, H. A.; SCHMIDT, E.; HUTTENLOCHER, D. *The age of AI and our human future*. New York: Back Bay Books/Little, Brown and Company, 2022.

KROLL, J. A. Accountability in computer systems. In: DUBBER, M. D.; PASQUALE, F. *The Oxford handbook of ethics of AI*. Oxford: University Press, p. 181-196, 2020.

LEE, K.-f.; QIU FAN, C. 2041: como a inteligência artificial vai mudar sua vida nas próximas décadas. Rio de Janeiro: Globolivros, 2022.

MART, S. N. *The algorithm as a human artifact: implications for legal {re}search*. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2859720. Acesso em: 10 mar. 2021.

MASLEJ, N. *et al.* *The AI Index 2025 Annual Report*. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2025.

REGULAMENTO (UE) N° 2024/1689 do Parlamento Europeu e do Conselho de 13 de junho de 2024, que cria regras harmonizadas em matéria de inteligência artificial. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L_202401689. Acesso em 30 out. 2024.

ROLLAND, M. G. *Technology's on Federal Sentencing: past, present, and future*. Disponível em: <https://scholarlycommons.law.wlu.edu/cgi/viewcontent.cgi?article=1497&context=crsj>. Acesso em: 2 maio 2022.

RUSSEL, S. *Inteligência artificial a nosso favor: como manter o controle sobre a tecnologia*. São Paulo: Companhia das Letras, 2021.

SANTOS DE MENDONÇA, J. V.; BAPTISTA, P. F. Dogmática jurídica e interdisciplinaridade: caminhos para um diálogo útil. *Rev. Quaestio Iuris*, Rio de Janeiro, v. 15, n. 04, p. 2296-2310, 2022.

SCOTT, K. *O futuro da inteligência artificial: de ameaça a recurso*. Rio de Janeiro: Harper Collins, 2023.

SUPREME COURT OF USA. *Year-End Report on the Federal Judiciary*. 2023. Disponível em: <https://www.supremecourt.gov/publicinfo/year-end/2023year-endreport.pdf>. Acesso em: 9 jan. 2024.

SUSSKIND, R. *Tomorrow's lawyers: an introduction to your future*. United Kingdom: Oxford Express, 2023.

THE GUARDIAN. Disponível em: <https://www.theguardian.com/technology/2015/may/20/flickr-complaints-offensive-auto-tagging-photos>. Acesso em: 10 nov. 2019.

TEGMARK, M. *Vida 3.0: o ser humano na era da inteligência artificial*. São Paulo: Benvirá, 2020.

TOFFOLI, J. A. D.; RANGEL, R. M.; TOLEDO, E. Inteligência artificial no Poder Judiciário brasileiro: o Projeto Victor do Supremo Tribunal Federal. In: TOMASEVICIUS FILHO, E. *Inteligência artificial: visões disciplinares e internacionais*. São Paulo: Almedina, p. 337-352, 2023.

WEBB, A. *Os nove titãs da IA: como as gigantes da tecnologia e suas máquinas pensantes podem subverter a humanidade*.

WHITE HOUSE. *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*. Washington, DC, 2023. Disponível em: <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>. Acesso em: 30 out. 2023.

Submissão em: 27.02.2024

(Avaliador A) Avaliado em: 02.06.2025

(Avaliador B) Avaliado em: 16.12.2024

Aceito em: 03.06.2025