

INTELIGÊNCIA ARTIFICIAL NO JUDICIÁRIO BRASILEIRO: ESTUDO EMPÍRICO SOBRE ALGORITMOS E DISCRIMINAÇÃO

Caroline Somesom Tauk¹

Luis Felipe Salomão²

Resumo

O texto analisa as iniciativas que envolvem inteligência artificial (IA) no Judiciário brasileiro, em suas diferentes fases, desde a ideação até a produção, com base em dados extraídos de pesquisa empírica que mapeou as iniciativas em 2022. O texto apresenta o estado da arte relativo aos sistemas inteligentes, dividindo-se em dois eixos. O primeiro trata dos aspectos gerais e técnicos, explicando como funcionam os sistemas de IA no Judiciário brasileiro, a forma de treinamento dos algoritmos, as tarefas que desempenham, as equipes e as regiões do país em que estão distribuídos. O segundo conclui que os atuais sistemas observam os princípios éticos extraídos de documentos regulatórios, nacionais e internacionais, com ênfase na discussão sobre opacidade, transparência, supervisão humana e discriminação algorítmica.

Palavras-chave: Inteligência artificial. Sistema de Justiça. Princípios éticos. Opacidade. Discriminação.

Abstract

The text analyzes the initiatives involving artificial intelligence (AI) in the Brazilian Judiciary, in its different phases, from ideation to production, based on data gathered from empirical research in 2022. The text presents the state of the art regarding intelligent systems, divided into two parts. The first deals with general and technical aspects, explaining how AI systems work in the Brazilian Judiciary, how the algorithms are trained, the tasks they perform, the teams and the regions of the country where they are distributed. The second concludes that current systems observe ethical principles extracted from national and international regulatory documents, with emphasis on the discussion of opacity, transparency, human supervision and algorithmic discrimination.

Keywords: Artificial intelligence; Justice System; Ethical principles; Opacity; Discrimination.

1. INTRODUÇÃO

¹ Juíza Federal no Rio de Janeiro; Mestre em Direito Público – UERJ; Doutoranda em Direito Civil – USP; Coordenadora Acadêmica do Centro de Inovação, Administração e Pesquisa do Judiciário (CIAPJ/FGV Conhecimento); ORCID: <https://orcid.org/0009-0005-1850-1326> Email: carolinetauk@gmail.com.

² Ministro do Superior Tribunal de Justiça; Doutor honoris causa em Ciências Sociais e Humanas pela Universidade Cândido Mendes-RJ; Professor Emérito da Escola da Magistratura do Estado do Rio de Janeiro e da Escola Paulista da Magistratura; Coordenador Geral do Centro de Inovação, Administração e Pesquisa do Judiciário (CIAPJ/FGV Conhecimento); ORCID: <https://orcid.org/0009-0006-2769-2408> E-mail: gab.luis.f.salomao@stj.jus.br.

Victor, Elis, Julia, Larry, Judi. Pessoas? Não, robôs. Essas são algumas das dezenas de ferramentas de inteligência artificial (IA) em funcionamento no Judiciário brasileiro. Conferir nomes a máquinas não as personifica, mas demonstra um desejo, ainda que inconsciente, de aproximá-las da figura do ser humano.

A utilização de máquinas para tarefas tradicionalmente desempenhadas por magistrados e outros servidores públicos nasceu de uma necessidade, muito mais do que de uma escolha de gestão pública. O Conselho Nacional de Justiça (CNJ) contabilizou, em 2021, 77,3 milhões de processos em tramitação no Brasil, dos quais 80,8% estão em meio eletrônico (CNJ, 2022, p. 103 e 192). Isso significa milhões de páginas de documentos de texto, em formato não estruturado, que exigem considerável esforço humano para tarefas de triagem, análise e tomada de decisão. O volume de dados que se origina dos processos requer o desenvolvimento de soluções escaláveis destinadas ao domínio jurídico. Os Tribunais perceberam essa necessidade.

Foi assim que o Judiciário brasileiro passou, em especial desde 2019, a ter interesse em áreas específicas da inteligência artificial, na busca de soluções capazes de extrair e organizar as informações não estruturadas presentes em documentos, vindo, a partir daí, a usar os dados em sistemas de suporte à decisão, destacando-se a categorização e triagem de processos e a automação de fluxos de trabalho, bem como em modelos preditivos, capazes de antever possíveis resultados de um processo judicial, como o percentual conciliação.

O crescimento de iniciativas que envolvem inteligência artificial no Judiciário motivou a pesquisa realizada pelo Centro de Inovação, Administração e Pesquisa do Judiciário (CIAPJ), da Fundação Getúlio Vargas (FGV Conhecimento), intitulada “*Tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário*” (CIAPJ/FGV, 2022), publicada em abril de 2022, a qual fez abrangente levantamento das tecnologias de IA no CNJ, no Supremo Tribunal Federal (STF), no Superior Tribunal de Justiça (STJ), no Tribunal Superior do Trabalho (TST), nos Tribunais Regionais Federais (TRFs), nos Tribunais de Justiça (TJs) e nos Tribunais Regionais do Trabalho (TRTs).

Na pesquisa, que está na segunda edição, foram identificadas 64 ferramentas de IA espalhadas por 44 Tribunais (STJ, STJ, TST, os cinco TRFs, 23 Tribunais de Justiça e 13 TRTs), além da Plataforma Sinapses do CNJ. As informações mapeadas incluíram a equipe, os aspectos técnicos, a base de dados, a avaliação e o monitoramento de cada sistema.

Nesse cenário, partindo das informações contidas no relatório de pesquisa do CIAPJ/FGV e da literatura a respeito da inteligência artificial, o presente texto busca apresentar

o estado da arte relativo aos sistemas inteligentes no Judiciário brasileiro, dividindo-se em dois eixos.

O primeiro eixo objetiva cobrir os aspectos gerais e técnicos relacionados ao uso da IA no sistema de justiça, enfatizando a indispensável participação humana ao longo do desenvolvimento e implantação dos sistemas. Além da contextualização do crescimento das inovações tecnológicas no sistema de justiça e da apresentação do *machine learning*, subcampo da IA bastante comum nos sistemas estudados, é feita uma análise, consideradas as limitações do texto, de como funcionam os sistemas de IA no Judiciário, a forma de treinamento dos algoritmos, as tarefas que desempenham, as equipes e as regiões do país em que estão distribuídos.

No segundo eixo, são debatidos os princípios éticos extraídos de documentos regulatórios, nacionais e internacionais. Também é analisada sua observância pela ampla paisagem de sistemas de IA no Judiciário do país, a fim de averiguar em que medida causam risco a direitos fundamentais dos cidadãos. Pretende-se dar enfoque às indagações que surgem a respeito dos princípios da transparência, do controle do usuário e da não discriminação, de modo a respondê-las a partir de diferentes perspectivas e seus contrapontos, com atenção aos tipos de tarefas atualmente desempenhadas pelos sistemas.

2. A INTELIGÊNCIA ARTIFICIAL E O *MACHINE LEARNING* - BREVE INTRODUÇÃO

Em 1950, Alan Turing iniciou, em artigo seminal, discussão acadêmica sobre a possibilidade de existência de máquinas inteligentes que pudessem imitar o comportamento humano (TURING, 1950, p. 433). Apesar de o debate não ser recente, inexistente conceito amplamente aceito de inteligência artificial, seja porque os sistemas de IA interagem com a inteligência humana de inúmeras formas, seja porque são projetados para operar com diversos níveis de autonomia. Por isso, uma forma simples é descrever os sistemas de IA como “todo sistema computacional que simula a capacidade humana de raciocinar e resolver problemas, por meio de tomadas de decisão baseadas em análises probabilísticas” (MULHOLLAND, 2020).

Atualmente, o subcampo da IA mais promissor é o *machine learning* ou aprendizado de máquina, assim chamado em razão da sua capacidade de aprendizagem com a

experiência: tal sistema permite que computadores possam aprender a partir de experiências passadas, representadas por um conjunto de dados. A novidade é ensinar máquinas em vez de programá-las. Em geral, sistemas que utilizam *machine learning* ensinam o programa de computador a identificar padrões em dados passados e a aplicar o conhecimento obtido em novos dados, a fim de que possam lidar com *tarefas descritivas*, como agrupar dados em subconjuntos de dados semelhantes (como agrupar consumidores conforme o perfil), ou com *tarefas preditivas*, fazendo previsão de resultados (como sugestão de compras a consumidores e de conexões em redes sociais) (GOODFELLOW, BENGIO e COURVILLE, 2016, p. 96-101). Da próxima vez que receber a tarefa, o desempenho do sistema de IA melhora com a experiência.

O processo de *machine learning* depende da análise de dados por algoritmos. O algoritmo “é um conjunto de instruções matemáticas, uma sequência de tarefas para alcançar um resultado esperado em um tempo limitado” (KAUFMAN, 2018). Em outras palavras, trata-se de comando, codificado em linguagem matemática, para permitir a leitura por computadores. Nem todo algoritmo utiliza inteligência artificial. Calculadoras simples, por exemplo, utilizam algoritmos para operações matemáticas, mas não são inteligentes. Os algoritmos inteligentes, por outro lado, usam técnicas mais sofisticadas que buscam aproximar o sistema da ação racional.

No Judiciário brasileiro, a pesquisa do Centro de Inovação, Administração e Pesquisa do Judiciário, da FGV Conhecimento, revelou que a aprendizagem de máquina é atualmente a técnica mais comum entre os sistemas de IA, presente em 77% dos sistemas (CIAPJ/FGV, 2022, p. 256), tanto para o desempenho de tarefas descritivas quanto de preditivas.

Diante da diversidade, as espécies de algoritmos de *machine learning* são classificadas de muitas formas. Uma delas é a que divide os algoritmos conforme o tipo de treinamento do modelo computacional e a forma como os dados de treinamento são usados, classificando-os em algoritmos *supervisionados*, *não supervisionados* e *de aprendizado por reforço*, somando 77% dos sistemas de IA mencionados acima. As ferramentas de *machine learning* podem combinar mais de uma forma de treinamento. Nos outros 23% dos sistemas, não é utilizado aprendizado de máquina, embora possa haver uso da IA. Essa classificação auxilia a verificação, também, do tipo de contribuição humana em cada situação.

No aprendizado supervisionado (*supervised learning*), são apresentados ao algoritmo padrões de treinamento acompanhados de um conjunto de rótulos (ou classes)

previamente conhecidos. O programador seleciona os dados de treinamento, rotula-os para ensinar o algoritmo a chegar ao resultado desejado – exigindo-se, portanto, *maior participação humana* no treinamento – e o sistema vai “aprendendo” por tentativa e erro, a fim de que possa realizar previsões para novos padrões, em situações em que não haja a rotulagem (DREXL, HILTY et al, 2019, p. 8-9).

Imagine-se, por exemplo, que se pretenda utilizar sistema de IA para identificar animais, como gatos. Nesse caso, o modelo computacional seria treinado com conjunto de dados rotulados, de modo que cada imagem viria acompanhada pela informação de que há ou não um gato. Depois de concluído o treinamento, o modelo deveria ser capaz de reconhecer, a partir de imagem sem rótulo, se um gato aparece nela. O uso prático poderia ser a inserção desse sistema em carro autônomo, para permitir que freie quando deparar com um gato.

No Judiciário, o aprendizado supervisionado é utilizado em 49% dos sistemas de IA, conforme pesquisa do CIAPJ/FGV. Por exemplo, no TST, o sistema Bem-te-vi utiliza essa forma de treinamento para auxiliar na triagem nos gabinetes. Os dados de processos decididos nos últimos dois anos por cada Ministro foram utilizados como insumos para os algoritmos, a fim de permitir que sejam extraídas previsões, como o formato da decisão (acórdão ou decisão monocrática), o assessor mais experiente na matéria e a análise de transcendência, requisito para admissibilidade do recurso.

A outra abordagem do aprendizado de máquina consiste em utilizar algoritmos não supervisionados, e o treinamento não depende de dados rotulados. O modelo computacional é treinado para identificar semelhanças, paralelos e/ou diferenças nos dados de entrada, com o objetivo de organizar amostras (clusterização ou *clustering*), sendo bastante empregados em tarefas de agrupamento e associação. Como a rotulagem dos dados de treinamento não é necessária, exige-se menos participação humana no treinamento; no entanto, é necessária maior interpretação humana em relação ao resultado (DREXL, HILTY et al, 2019, p. 8-9).

Imagine-se, por exemplo, que certa empresa pretenda dividir em grupos os clientes com base no poder de compra, para personalizar as ofertas. A utilização do aprendizado não supervisionado permitiria identificar semelhanças e diferenças nos dados dos clientes e agrupá-los conforme o poder de compra, idade, gênero etc.

O aprendizado não supervisionado é utilizado em 26% dos sistemas de IA no Judiciário, conforme pesquisa do CIAPJ/FGV. Por exemplo, no STJ, o sistema Athos utiliza essa forma de treinamento para a categorização, classificação e agrupamento de documentos, com o fim de realizar as tarefas de identificação de temas repetitivos e monitoramento de temas

de processos com peças semelhantes. De forma similar, o sistema Larry, do Tribunal de Justiça do Paraná (TJPR), utiliza esse tipo de algoritmo para o agrupamento e a identificação de petições iniciais que descrevem fatos similares.

Por fim, a terceira abordagem do aprendizado de máquina consiste no aprendizado por reforço (*reinforcement learning*). Diferentemente do aprendizado supervisionado e do não supervisionado, esse enfoque não depende de conjuntos de dados pré-existentes, mas da coleta de dados de simulações ou jogos. O algoritmo “descobre” as regras e toma as decisões com base no *feedback* contínuo que recebe em relação às ações realizadas durante o treinamento, gerando sistema de pontuações. Nesse caso, o envolvimento humano restringe-se à mudança do ambiente ou ao ajuste do sistema de pontuações no jogo ou na simulação (DREXL, HILTY et al, 2019, p. 8-9).

Essa forma de aprendizado foi usada, por exemplo, no treinamento de um agente, conhecido como AlphaGo, para o jogo *Go*, sendo obtidos resultados melhores do que o de humanos. Em tal jogo, considerado mais difícil que o xadrez, o algoritmo nunca foi ensinado sobre as estratégias para vencer, tendo apenas jogado contra si mesmo várias vezes, alcançando, ao fim de cada partida, o *feedback* com base na pontuação final (SILVER, HUANG, MADDISON et al, 2016, p. 484-489).

No Judiciário brasileiro, apenas 3% dos sistemas de IA utilizam a aprendizagem por reforço segundo a citada pesquisa. No Tribunal de Justiça de Santa Catarina (TJSC), o Classificador de Petições em Execuções Fiscais utiliza essa forma de treinamento para classificar textos de petições, de modo a auxiliar na localização dos processos e a permitir o impulso em bloco de grandes volumes de processos. No Tribunal Regional do Trabalho da 17ª Região (TRT17), o Gemini usa o aprendizado de reforço para agrupar documentos de processos no segundo grau por similaridade de temas, contribuindo, assim, na elaboração de votos por matéria nos gabinetes.

Como se percebe, a participação humana, em pequena, média ou grande intensidade, está presente em todas as formas de aprendizagem mencionadas. Apesar da ideia que o nome gera, dizer que algoritmos de aprendizado de máquina se capacitam de modo autônomo depende da noção que se dê ao termo autonomia, uma vez que a participação humana é indispensável para o resultado final, podendo ser maior no treinamento ou na interpretação do resultado. Esta conclusão deixa clara a necessidade de atenção à formação das equipes de pesquisa, desenvolvimento e implantação dos sistemas.

3. INTELIGÊNCIA ARTIFICIAL FORTE E FRACA: FICÇÃO CIENTÍFICA VS. REALIDADE

Além da importância da noção trazida no tópico anterior, no sentido de que a participação humana é indispensável para o desenvolvimento, implantação e produção dos sistemas de IA, outra noção fundamental é a compreensão do atual estado da arte da IA.

Os grandes avanços da IA até o momento, no Brasil e no mundo, foram todos feitos em domínios limitados. A constatação evidencia a diferença entre inteligência artificial forte e fraca.

Considera-se que há inteligência artificial fraca ou restrita (*weak* ou *narrow*) quando máquinas desempenham tarefas específicas e altamente restritas, tomando decisões limitadas ao objetivo do treinamento (RUSSEL e NORVIG, 2010, p. 27-28). A *performance*, portanto, é setORIZADA, de modo que a inteligência não se mostra completa, diferentemente da inteligência humana. Por exemplo, algoritmos treinados para agrupamento de processos judiciais similares não responderiam bem a uma tarefa de agrupamento de exames de raio-X de pacientes.

A inteligência artificial forte ou geral (*strong* ou *general*), por sua vez, refere-se a sistemas capazes de tomar decisões inteligentes em diversas áreas, permitindo *performance* abrangente, bem como de pensar e executar tarefas por conta própria, similar a um ser humano. Filmes de ficção científica têm gosto por essa forma de IA: alguns personagens são robôs que desempenham tarefas em diversas áreas do conhecimento e interagem com humanos de igual para igual. Não passam de histórias de ficção. Em razão da grande dificuldade de criação, não existe atualmente a inteligência artificial forte.

Mesmo o Watson da IBM, um dos mais bem-sucedidos sistemas de IA existentes, é altamente limitado. Utilizando processamento de linguagem natural (PLN), tal sistema foi capaz de vencer no programa de televisão *Jeopardy!*, em que é feito quiz com perguntas e respostas. No entanto, não funcionaria àquela época se houvesse perguntas que nunca foram feitas ou perguntas para as quais as respostas não tinham sido anteriormente registradas (GOVINDARAJULU, BRINGSJORD e LICATO, 2013, p. 1-2). Watson é inteligente, mas tem limitações.

Embora atualmente a IA fraca seja a única existente, isso não significa, de modo algum, que seus efeitos não sejam importantes para a sociedade sob o aspecto social, econômico

e político. No Judiciário, as ferramentas de IA têm sido usadas, na maior parte, para tarefas voltadas à melhoria na gestão do sistema de justiça, em busca de qualidade e efetividade na prestação jurisdicional. No próximo tópico, é visto o percurso até a adoção da IA no sistema público de justiça.

4. DESENVOLVIMENTO TECNOLÓGICO DO JUDICIÁRIO BRASILEIRO

A inserção da tecnologia para a administração dos conflitos no Judiciário brasileiro desenvolveu-se a partir da *virtualização* ou *digitalização*. A Lei 11.419/2006 inaugura o processo judicial eletrônico (PECK, 2021, p. 365), disciplinando o uso de meio eletrônico na tramitação de processos, comunicação de atos e transmissão de peças nos processos civil, penal e trabalhista, bem como nos juizados especiais, em qualquer grau de jurisdição.

A virtualização merece algumas observações.

Em primeiro lugar, os processos eletrônicos materializam-se em milhões de páginas de documentos, em formato não estruturado, que demandam grande esforço humano para tarefas de triagem, análise e tomada de decisão. Portanto, o processo judicial eletrônico é composto por diversos dados *desestruturados*, ou seja, dados de documentos textuais que não possuem organização (AMARAL, 2016), a exemplo de todas as informações que constam na petição inicial, na defesa e na sentença. Algumas informações, por outro lado, já vêm *estruturadas* no sistema de tramitação processual do Tribunal, isto é, organizadas em banco de dados ou planilha do sistema, previamente desenhada para armazená-las, como o nome das partes, a classe e o assunto processual. Muito pode ser feito a partir de ambos os tipos de dados. O Tribunal Regional Federal da 3ª Região, por exemplo, informou na pesquisa do CIAPJ/FGV que está desenvolvendo repositório de dados estruturados e não estruturados dos processos judiciais (*data lake*), os quais poderão ser usados para construir painéis de *Business Intelligence* e iniciativas de IA.

Em segundo lugar, ferramentas tecnológicas “entendem” o significado das peças dos processos eletrônicos, em geral, por meio da utilização da técnica de processamento de linguagem natural (PLN) ou *Natural Language Processing* (NLP).

O PLN é uma subárea da inteligência artificial que permite aos computadores processar a linguagem humana, escrita ou falada, chamada de linguagem natural, e “entender” seu significado. O processamento converte as palavras escritas ou faladas em números, ou seja,

em dados manipuláveis pelo computador (CHOWDHURY, 2003, p. 51). Entre as diversas aplicações atuais, o PLN vem sendo utilizado para reconhecimento de fala ou conversão de voz para texto e categorização ou classificação de documentos.

Concluídas as duas observações sobre a virtualização ou digitalização, retomase o percurso da inserção da tecnologia no Judiciário brasileiro. A virtualização permitiu a *automatização* de tarefas humanas. Utiliza-se a tecnologia com o objetivo de aumentar a produtividade, reduzir o tempo e os custos da tramitação do processo, por meio da automação, com a possibilidade de ser utilizada inteligência artificial. Fala-se em automação simples quando as máquinas respondem a um comando e o uso destina-se a tarefas repetitivas ou de baixa complexidade cognitiva. É o caso, por exemplo, de sistemas de tramitação processual no Judiciário que permitem programar a publicação e intimação da decisão, uma vez assinada pelo juiz, dispensando que esses atos sejam feitos pelo assessor.

Quando há um salto de sofisticação na automação, fala-se em inteligência artificial, ocasião em que as máquinas têm a capacidade de escolher a melhor ação a ser tomada para atingir determinado objetivo, considerando dados disponíveis. É nesse sentido a definição proposta pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE) em 2019, ao anunciar princípios para o desenvolvimento da IA, tendo como signatários 42 países signatários, entre eles o Brasil (OCDE, 2019)³.

Por fim, sobretudo mais recentemente, ganhou fôlego a *desmaterialização* da justiça (BRAGANÇA, 2021, p. 33-35), marcada pelo conjunto de medidas destinadas à realização de atos processuais por videoconferência, como as audiências virtuais e o contato virtual entre partes e magistrados. Desenvolveu-se a ideia de conceber a justiça como um serviço (*justice as a service*) e não mais associada apenas a um lugar, como os fóruns. É nesse sentido que Richard Susskind afirma que funcionar *online* não é alternativa ao sistema judicial, sendo o próprio sistema judicial (SUSSKIND, 2019, p. 19).

São exemplos a autorização do CNJ para a adoção do “Juízo 100% Digital”, no âmbito do qual todos os atos processuais serão exclusivamente praticados por meio eletrônico, inclusive a produção de meios de prova (Resolução CNJ n. 345/2020 - CNJ, 2020), bem como o “Balcão Virtual”, ferramenta de videoconferência que permite imediato contato de partes e advogados com as secretarias das varas e gabinetes de Desembargadores durante o horário de atendimento ao público (Resolução CNJ n. 372/2021 - CNJ, 2021).

³ Segundo a OCDE, “um sistema de IA é um sistema baseado em máquina que pode, para um determinado conjunto de objetivos, fazer previsões, recomendações ou decisões que influenciam ambientes reais ou virtuais”.

5. INTELIGÊNCIA ARTIFICIAL NO JUDICIÁRIO BRASILEIRO: ESTUDO EMPÍRICO

A pesquisa do CIAPJ/FGV identificou 64 modelos computacionais de IA no Judiciário, em diferentes fases - em ideação, em desenvolvimento ou já implantados –, os quais podem ser divididos, para fins didáticos, em quatro grupos principais em relação às tarefas que desempenham (SALOMÃO e TAUKE, 2022).

No *primeiro grupo*, inclui-se pequena parcela de sistemas de IA destinados a assistir nas *atividades-meio* do Judiciário relacionadas à administração, objetivando gerir de modo mais eficiente os recursos financeiros e de pessoal, e não auxiliar o magistrado na prestação jurisdicional. Citem-se, como exemplo, no Tribunal de Justiça do Rio Grande do Sul (TJRS), o *Chatbot Digep*, que responde dúvidas dos servidores quanto aos assuntos relacionados à gestão de pessoas; no Tribunal de Justiça de São Paulo (TJSP), o *Judi Chatbot*, que fornece orientação ao cidadão para ajuizamento de ações relacionadas ao Juizado Especial Cível; no Tribunal de Justiça do Distrito Federal (TJDFT), o *Amon*, que faz o reconhecimento facial de quem ingressa no Tribunal, a partir de fotografias, visando aumentar a segurança do fórum.

O segundo e o terceiro grupo englobam modelos computacionais que contemplam a *atividade-fim* do Judiciário, qual seja, a prestação jurisdicional.

O *segundo grupo* abrange a maioria dos modelos, que se destinam à automação dos fluxos de movimentação do processo e das atividades executivas de apoio aos juizes por meio da execução de tarefas pré-determinadas. Esses modelos computacionais dão apoio à gestão de secretarias e gabinetes, fazendo triagem e agrupamento de processos similares, classificação da petição inicial, transcrição de audiências etc.

São alguns exemplos desses modelos: no STJ, o *Athos* faz a identificação e o monitoramento de temas repetitivos, assim como há ferramenta que otimiza a identificação e indexação das peças processuais vindas com os autos originários; no Tribunal Regional Federal da 3ª Região (TRF3), o *Julia* (Jurisprudência Laborada com Inteligência Artificial) auxilia na localização de processos sobrestados, cujas decisões devam ser reformadas em função do julgamento do tema de repercussão geral ou recurso repetitivo; no Tribunal Regional Federal da 4ª Região (TRF4), há ferramenta que faz o agrupamento de apelações por similaridade de

sentença; no Tribunal de Justiça do Amapá (TJAP), o Tia auxilia na identificação de demandas repetitivas, havendo a mesma ferramenta do Tribunal de Justiça da Bahia (TJBA); no Tribunal de Justiça de Alagoas (TJAL), o Hércules faz a triagem de petições em processos de execução fiscal; no Tribunal de Justiça do Acre (TJAC) e do Rio de Janeiro (TJRJ), bem como no TJDFT, neste último, nomeada Toth, há mecanismos que prestam assistência aos advogados na classificação correta das petições.

Pode-se listar nesse grupo, ainda, os seguintes modelos: no TJDFT, o Horus digitaliza o acervo físico das Varas de Execução Fiscal e faz o reconhecimento ótico dos caracteres da peça processual para identificar os tipos de documento; no Tribunal de Justiça de Goiás (TJGO), o Berna (Busca Eletrônica em Registros usando linguagem Natural) analisa petições iniciais e verifica se há outros processos com pedidos semelhantes; no Tribunal de Justiça do Pará (TJPA), há ferramenta que faz migração de processos físicos para o PJe; no TJPR, o Larry reconhece processos similares que ingressam no Estado; no Tribunal de Justiça do Rio Grande do Norte (TJRN), o GPSMED busca identificar, a partir da análise de conteúdo de petições iniciais, a espécie de demanda de processos de saúde pública (tratamento e medicamento solicitado, tipo de cirurgia e de doença); no Tribunal de Justiça de Rondônia (TJRO), o Peticionamento Inteligente ampara as delegacias de polícia a enviar documentos (termo circunstanciado, inquérito policial etc) ao PJe; no Tribunal de Justiça de Roraima (TJRR), o Mandamus utiliza inteligência artificial para o cumprimento e certificação de mandados judiciais; no TJRS, o Grafo transcreve os áudios das audiências, e o IA Execução Fiscal auxilia na análise e classificação da petição inicial, tarefa similar, embora não restrita à execução fiscal, à do Sistema de Classificação de Petições Judiciais do Tribunal de Justiça de Tocantins (TJTO); no TJSP, há ferramenta que analisa as guias de recolhimento das custas processuais; no Tribunal Regional do Trabalho da 18ª Região (TRF18), há o Sistema Automatizado de Busca Patrimonial, que faz buscas automáticas nos principais convênios, como Infojud, Renajud, CNIB e Censec.

No *terceiro grupo*, em menor quantidade, há modelos computacionais de inteligência artificial que colaboram na elaboração de minutas com conteúdo decisório de sentença, votos ou decisões interlocutórias.

São exemplos desses modelos: no STF, o Victor facilita a identificação da presença de temas de repercussão geral; no STJ e nos Tribunais Regionais do Trabalho da 8ª e da 9ª Regiões (TRT8 e TRT9), há ferramentas para auxiliar na admissibilidade dos recursos; no Tribunal Regional Federal da 1ª Região (TRF1), o Alei (Análise Legal Inteligente), dentre

outras tarefas, pretende associar, ao processo judicial em análise, julgados anteriores e buscar jurisprudência; no Tribunal de Justiça do Espírito Santo (TJES), o Argos, e no Tribunal de Justiça do Paraíba (TJPB), o Midas, dão suporte às decisões sobre justiça gratuita; no Tribunal de Justiça do Mato Grosso do Sul (TJMS), a ferramenta Jurimetria com Inteligência Artificial aponta tendências de julgamento do órgão baseado nos acórdãos e na jurisprudência; no Tribunal de Justiça de Pernambuco (TJPE), o Elis realiza a triagem das petições iniciais dos processos de execução fiscal e, se não houver inconsistências, elabora automaticamente a minuta e a encaminha para a conferência do magistrado; no TJSC e no Tribunal Regional do Trabalho da 9ª Região (TRF9), há instrumentos para incrementar a busca de jurisprudência, como o Magus trabalhista.

A divisão entre os grupos foi feita apenas para facilitar a compreensão, mas não é estanque. Modelos computacionais para agrupamento de processos, incluídos no segundo grupo, podem auxiliar na elaboração de minutas de decisão se tal agrupamento identificar, por exemplo, causas de prescrição.

Finalmente, incluem-se no *quarto grupo* iniciativas relacionadas a formas adequadas de resolução de conflitos, em que se usam informações de processos similares para amparar as partes na busca da melhor solução. Como exemplo, cite-se, no Tribunal Regional do Trabalho da 4ª Região (TRT4), o Índice de Conciliabilidade por Inteligência Artificial (Icia), que estima a probabilidade de o processo ser conciliado no estágio em que se encontra, tarefa similar à desempenhada pelo Concilia JT, do Tribunal Regional do Trabalho da 12ª Região (TRF12).

Os dados da pesquisa deixam claro que, em nenhuma das iniciativas, há a interpretação de textos legais, a elaboração de argumentação jurídica e, muito menos, a tomada de decisão pela máquina. Mesmo nos modelos computacionais que fornecem subsídios para a elaboração de minutas com conteúdo decisório, a tarefa da máquina limita-se a identificar temas repetitivos ou fundamentos presentes nas peças, a pesquisar jurisprudência e a sugerir decisões simples como de gratuidade de justiça, sendo todos os resultados sujeitos à supervisão do juiz. Em suma, as máquinas atuais são capazes de utilizar a inteligência de dados, mas não a consciência hermenêutica própria do raciocínio humano (MARRAFON, 2020).

Em síntese, os sistemas utilizam métodos de predição por probabilidade, ou seja, “extraem os resultados de uma série de decisões judiciais e as correlacionam com fatores como o tipo de demanda, valor envolvido e o Tribunal em que a demanda é julgada”. Justamente por isso, os sistemas de aprendizado de máquina “não levam em consideração qualquer justificção

normativa sobre como deve ser a decisão a partir das características e argumentos do caso” (MARANHÃO, FLORÊNCIO e ALMADA, 2021, p. 165).

Assim, pode-se afirmar que, no estado da arte atual, inexistente tecnologia capaz de substituir juízes por robôs na tomada de decisões jurisdicionais no Brasil. Ademais, tais modelos gerariam a preocupação relativa ao alto risco gerado, sobretudo riscos discriminatórios, a exigir maior cautela e controle.

Cabem duas observações finais neste tópico relativas: i) à *distribuição das iniciativas por região*, percebendo-se o impacto da pandemia no aumento do ritmo de crescimento observado em relação aos anos anteriores e ii) às *equipes técnicas*, de desenvolvimento e curadoria dos sistemas.

Conforme informações da pesquisa do CIAPJ/FGV, a maior parte das iniciativas está localizada no Centro-Oeste, totalizando vinte sistemas, principalmente nos Tribunais localizados em Brasília. Em seguida, vem a região Sul com treze iniciativas, a Sudeste com doze, a Nordeste com dez e a Norte com nove. Portanto, o desenvolvimento tecnológico no Judiciário ocorre não apenas nos grandes centros urbanos, estando presente de norte a sul do país, a exemplo da plataforma aberta Sinapses, que pode hospedar qualquer modelo e tecnologia de IA, desenvolvida pelo TJRO e atualmente ampliada para o âmbito nacional pelo CNJ, nos termos da Portaria CNJ n. 271/2020.

A criação de plataforma comum, com modelos de IA que utilizam preferencialmente *software* de código aberto, nasceu da necessidade de compartilhamento de conhecimento entre os Tribunais e da *interoperabilidade tecnológica*, prevista no art. 194 do Código de Processo Civil, definida como característica intrínseca dos sistemas operacionais, possibilitando a troca de informações com outros sistemas de organizações conexas (IWAKURA, 2016, p. 109). Na prática, no entanto, a maioria das iniciativas origina-se dos próprios Tribunais, de modo que é frequente o uso de tecnologias e representações de dados distintas entre os sistemas, tornando pouco usual que sistemas bem-sucedidos sejam (ou possam ser) replicados em outros contextos.

As equipes de desenvolvimento das soluções inteligentes são, majoritariamente, internas, ou seja, 91% das iniciativas são formadas por magistrados, servidores ou terceirizados do Tribunal. Somente 6% vêm de equipe externa, oriundas da iniciativa privada ou universidades, e 3% resultam da atuação interna e externa, a exemplo do sistema Victor do STF, desenvolvido por equipe interna em conjunto com a Universidade de Brasília (UnB).

Em qualquer caso, a formação das equipes deverá ter caráter *interdisciplinar*, incluindo profissionais de Tecnologia da Informação e de outras áreas do conhecimento científico, bem como orientada pela busca da *diversidade* dos profissionais em mais amplo espectro, como exige o art. 20 da Resolução CNJ n. 332/2020.

6. A REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL

Em todo o mundo, órgãos governamentais nacionais e internacionais e organismos privados vêm publicando documentos com diretrizes para o desenvolvimento e aplicação ética e responsável de sistemas de IA em diversos setores.

A União Europeia tem forte atuação na regulação de novas tecnologias e na proteção de dados pessoais. Além do Regulamento Europeu de Proteção de Dados (PARLAMENTO EUROPEU, 2016), a União Europeia apresentou, em 2021, a Proposta de Regulamento do Parlamento Europeu e do Conselho, a estabelecer normas harmonizadas em matéria de inteligência artificial (*AI Act*) com o objetivo de trazer regras sobre a confiabilidade de tais sistemas (COMISSÃO EUROPEIA, 2021).

Esse não foi o primeiro movimento da União Europeia no sentido de buscar regulação da inteligência artificial. Em outubro de 2020, o bloco publicou várias resoluções atinentes ao uso dessa tecnologia (UNIÃO EUROPEIA, 2020), como a resolução sobre o regime relativo aos aspectos éticos da inteligência artificial, da robótica e das tecnologias conexas, a resolução acerca do regime de responsabilidade civil aplicável à inteligência artificial e a resolução sobre os direitos de propriedade intelectual para o desenvolvimento de tecnologias ligadas à inteligência artificial.

Os Estados Unidos, embora com atuação bem mais incipiente, propôs o “*Algorithmic Accountability Act of 2019*” (SENADO DOS ESTADOS UNIDOS, 2019), que orienta a Comissão Federal de Comércio a exigir a condução de avaliações de impacto dos sistemas de decisão automatizada pelas entidades que usam, armazenam ou compartilham informações pessoais. O Estado de Nova York aprovou pouco antes, em 2018, a Lei n. 49, que busca abordar os efeitos nocivos dos algoritmos no setor público por meio de investigação e da criação de uma força-tarefa (CONSELHO DA CIDADE DE NOVA YORK, 2018).

No Brasil, desde 2020, está em tramitação o Projeto de Lei n. 21/2020, de autoria do Deputado Federal Eduardo Bismarck, a estabelecer fundamentos, princípios e diretrizes para

o desenvolvimento e a aplicação da IA⁴, sendo publicada, em 2021, a Estratégia Brasileira de Inteligência Artificial (Ebia) (MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES, 2021).

Especificamente em relação ao ambiente do Direito, destacam-se dois documentos pioneiros no tratamento do tema: a Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais, de 2018 (COMISSÃO EUROPEIA PARA A EFICÁCIA DA JUSTIÇA, 2018), e, no Brasil, a Resolução CNJ n. 332/2020, que dispõe sobre a ética, a transparência e a governança na produção e no uso de IA no Poder Judiciário, inspirado no ato europeu. Ambos os documentos adotam princípios éticos similares.

Entre os princípios adotados pelos dois documentos acima, três serão selecionados para análise quanto à respectiva observância no estado da arte do Judiciário brasileiro, considerando as informações fornecidas pela pesquisa do CIAPJ/FGV, a saber: i) princípio da transparência: o usuário precisa ter ciência, de forma clara e compreensível, dos métodos de tratamento de dados, além do que devem ser autorizadas auditorias externas; ii) princípio do controle do usuário: há possibilidade de revisão da proposta de decisão pelo magistrado e demais servidores públicos da unidade judicial, bem como dos dados utilizados para a elaboração da minuta; iii) princípio da não discriminação: é necessário prevenir que sistemas de IA incorporem vieses que possam intensificar qualquer discriminação entre indivíduos.

Nos tópicos a seguir, serão abordadas as preocupações sobre o atendimento a esses princípios pelos sistemas de IA e em que medida causam risco a direitos fundamentais dos cidadãos.

7. OPACIDADE: É POSSÍVEL TER RESULTADOS EXPLICÁVEIS?

A inteligência artificial e, mais especificamente, a técnica de *machine learning* costumam ser descritas como opacas ou como uma “caixa preta” (*black box*). A chamada opacidade significa que, embora os especialistas possam explicar como o modelo algorítmico funciona, eles não podem elucidar com precisão porque foi gerado resultado concreto com base nos dados de entrada. Em outras palavras, a opacidade refere-se à ideia de que “os seres

⁴ Uma análise da versão do PL 21/2020 aprovada pela Câmara dos Deputados é feita em: SALOMÃO, Luis Felipe (Coord). *Nota Técnica sobre o Projeto de Lei 21/2020*. Rio de Janeiro: FGV, 2021. Disponível: [https://ciapj.fgv.br/sites/ciapj.fgv.br/files/notatecnica_substitutivo_pl21_2020.pdf]. Acesso: 07.03.2023.

humanos são cada vez mais incapazes de entender, explicar ou prever o funcionamento interno dos algoritmos, seus vieses e eventuais problemas” (DONEDA e ALMEIDA, 2016, p. 60).

Existem modelos mais e menos opacos. Ainda conforme Doneda e Almeida, a dificuldade aumenta quando se está diante de sistemas de inteligência artificial que utilizam algoritmos de aprendizado de máquina, os quais melhoram o desempenho com a experiência e podem alterar o funcionamento interno, realizando ações que não estão de acordo com uma programação pré-definida e que resultam da interação com os dados e com o meio. Além disso, os conjuntos de dados, nesses casos, podem ser compostos por milhões ou bilhões de dados, especialmente diante do uso de redes neurais profundas, tipo de aprendizado de máquina baseado no uso de arquitetura de modelo mais complexa. No entanto, diferentemente de computadores, seres humanos não são capazes de processar grandes quantidades de dados. Esses fatores causam a opacidade de diversos sistemas de IA.

Diante da opacidade, como explicar o funcionamento interno que levou a determinado resultado produzido pelo algoritmo de *machine learning* com base em determinada entrada? Como fornecer os motivos que levaram ao resultado produzido pela máquina? As perguntas relacionam-se diretamente com a transparência perante os jurisdicionados.

Em se tratando de sistemas usados no Judiciário, como visto, a grande maioria dos modelos destina-se ao apoio à gestão dos gabinetes, fazendo triagem de processos similares e classificação de petições ou recursos. Em menor quantidade, há os sistemas de IA que auxiliam na elaboração de minutas com conteúdo decisório, a exemplo do Victor do STF e do Elis do TJPE. De todo modo, mesmos nesses dois últimos exemplos, trata-se de decisões muito simples, relativas a temas processuais e sujeitas à revisão humana. Portanto, ainda que se diga haver opacidade nos sistemas de IA atualmente em funcionamento no Judiciário, o risco de causar prejuízo às partes não parece alto, nem sequer existirá na maior parte das hipóteses de utilização dos referidos sistemas.

De qualquer forma, considerando que o uso de *machine learning* é tendência e que todos os sistemas de IA desse tipo, e não apenas os do Judiciário, terão de lidar com a opacidade, há algumas propostas de solução para combatê-la.

A primeira proposta de solução discutida vem da propriedade intelectual (SALOMÃO e TAUKE, 2020, p. 739-762) e refere-se à abertura do *código-fonte*. Sistemas de IA podem ser apoiados em programa de computador, e o código-fonte é onde fica a linguagem de programação fundamental para a compreensão do programa de computador, ou seja, o código contém todas as instruções necessárias para que o programa funcione. Tratados

internacionais, a exemplo do Acordo Trips (do inglês *Agreement on Trade-Related Aspects of Intellectual Property Rights*, art. 10.1⁵), e a legislação brasileira (art. 2º da Lei n. 9.609/98⁶) conferem proteção de direito autoral ao código-fonte, tornando-o fechado como regra. Alternativamente à proteção conferida pelo direito autoral, é possível optar pela proteção do código-fonte como segredo comercial ou industrial, hipótese em que será considerada informação confidencial, nos termos do artigo 39.2 do Trips.⁷ Portanto, pela legislação nacional e internacional, a abertura do código-fonte para demonstrar o funcionamento interno do algoritmo poderia esbarrar em questões relacionadas à propriedade intelectual.

De todo modo, independentemente de o código-fonte ser protegido pelo direito autoral ou por segredo comercial ou industrial, não há garantia de que o acesso e abertura trarão a explicação que se deseja (BURRELL, 2016, p. 5). Na prática, ainda que se possa visualizar com clareza os grandes conjuntos de dados utilizados e mesmo que o código seja escrito em linguagem compreensível, a complexidade surge da interação entre o código e os dados durante o funcionamento do algoritmo.

Nesse cenário, mesmo com acesso às linhas ou páginas do código-fonte, não é tarefa simples a efetiva compreensão de todos os fatores envolvidos no funcionamento do algoritmo, nem mesmo pelos programadores. Como explica Jenna Burrell, a lógica do algoritmo pode ser quase totalmente compreensível para os seres humanos. No entanto, quando o algoritmo está em funcionamento, justamente por ser de aprendizagem de máquina, ocorre mudança na lógica de decisão interna, já que o algoritmo desenvolve-se com os dados de treinamento, aumentando o grau de complexidade. A opacidade representa, assim, a “complexidade de o ser humano entender o algoritmo em ação, operando nos dados” (BURRELL, 2016, p. 5).

Em razão das dificuldades acima, atualmente uma das principais áreas de pesquisa é a *explicabilidade* ou *interpretabilidade* do aprendizado de máquina, referido como “*Explainable Artificial Intelligence*” ou “*XAI*”. É daí que vem a segunda proposta de solução para o problema da opacidade.

⁵ Acordo TRIPS: “Art. 10.1. Programas de computador, em código fonte ou objeto, serão protegidos como obras literárias pela Convenção de Berna (1971)”.

⁶ Lei n. 9.609/98: “Art. 2º O regime de proteção à propriedade intelectual de programa de computador é o conferido às obras literárias pela legislação de direitos autorais e conexos vigentes no País, observado o disposto nesta Lei.”

⁷ Acordo TRIPS: “SEÇÃO 7: PROTEÇÃO DE INFORMAÇÃO CONFIDENCIAL ARTIGO 39.2. Pessoas físicas e jurídicas terão a possibilidade de evitar que informação legalmente sob seu controle seja divulgada, adquirida ou usada por terceiros, sem seu consentimento, de maneira contrária a práticas comerciais honestas, desde que tal informação: a) seja secreta, no sentido de que não seja conhecida em geral nem facilmente acessível a pessoas de círculos que normalmente lidam com o tipo de informação em questão, seja como um todo, seja na configuração e montagem específicas de seus componentes”.

O termo “*Explainable AI*”, cunhado em 2004, refere-se a uma discussão que nasceu nos anos 70. O objetivo é criar conjunto de técnicas que produzam mais modelos explicáveis, sem interferir nos altos níveis de desempenho do sistema, de modo que a decisão algorítmica possa ser explicada aos usuários finais e outras partes interessadas em termos não técnicos (ADADI e BERRADA, 2018, p. 3). Especialistas vêm testando técnicas de interface homem-máquina capazes de traduzir modelos algorítmicos em diálogos de explicação compreensível e útil para o usuário final externo, fornecendo variedade de opções de *design* dos sistemas aos futuros desenvolvedores.

O aprofundamento das propostas de solução foge ao objeto deste texto. O que importa, por ora, é saber que a opacidade é tema de crescente atenção pelos especialistas, o que demonstra sinal positivo de preocupação com a transparência e a confiança.

8. TRANSPARÊNCIA DA MÁQUINA VS. TRANSPARÊNCIA HUMANA

Importante discussão relativa à opacidade refere-se à necessidade de dar transparência ao funcionamento da máquina e de manter a confiança do usuário quanto ao uso do sistema (DOMINGUES, SILVA e SOUZA, 2021, p. 6-7). No âmbito do sistema de justiça, uma parte essencial da transparência concretiza-se por meio da fundamentação das decisões judiciais pelo magistrado. No Brasil, a exigência de fundamentação é constitucional (art. 93, IX, da Constituição da República Federativa do Brasil - CRFB) e permite que as partes possam discordar da conclusão judicial, exercendo o contraditório e a ampla defesa (art. 5º, LV, da CRFB), o que pressupõe a inteligibilidade humana da decisão.

No Judiciário brasileiro, como visto, a grande maioria dos modelos computacionais implementados e em desenvolvimento destina-se a tarefas processuais de apoio à gestão dos gabinetes. A preocupação com a transparência aumenta, contudo, se o sistema de IA for utilizado para a elaboração de minuta de decisão judicial.

Atento à necessidade de transparência, o CNJ editou a Resolução n. 332/2020, a qual dispõe, no art. 8º, que a transparência consiste em (i) “divulgação responsável, considerando a sensibilidade própria dos dados judiciais”; (ii) “indicação dos objetivos e resultados pretendidos pelo uso do modelo de Inteligência Artificial”; (iii) “documentação dos riscos identificados e indicação dos instrumentos de segurança da informação e controle para seu enfrentamento”; (iv) “possibilidade de identificação do motivo em caso de dano causado

pela ferramenta de Inteligência Artificial”; (v) “apresentação dos mecanismos de auditoria e certificação de boas práticas”; (vi) “fornecimento de explicação satisfatória e passível de auditoria por autoridade humana quanto a qualquer proposta de decisão apresentada pelo modelo de Inteligência Artificial, especialmente quando essa for de natureza judicial”.

A intenção do CNJ parece ter sido compatibilizar o problema da caixa-preta de algoritmos de *machine learning* com a necessidade de transparência das decisões judiciais, por meio da busca por modelos menos opacos e mais explicáveis e da divulgação aos interessados. Embora as pesquisas nessa área ainda estejam se desenvolvendo em todo o mundo, trata-se de um passo importante do CNJ, que encontra fundamento na doutrina, a exemplo de Ana Frazão, que destaca: “Casos assim, também chamados de black boxes, requerem particular atenção e a adoção de outras medidas de explicabilidade, tais como a rastreabilidade, a auditabilidade e a comunicação transparente sobre as capacidades dos sistemas” (FRAZÃO, 2019, p. 513).

A preocupação com a opacidade da IA tem contraponto que merece ser considerado. A tomada de decisões judiciais por magistrados – assim como por todos os seres humanos – também possui alguma “opacidade”, estando marcada por vieses que lhes são inerentes e que, naturalmente, não serão demonstrados na fundamentação judicial, que expõe somente parcela do processo decisório. Assim, a fundamentação das decisões humanas traria, em alguns casos, apenas uma ilusão de transparência quando comparada com a decisão algorítmica.

A psicologia comportamental moderna trabalha com os chamados vieses de cognição humana, que decorrem do funcionamento do cérebro. Vieses de cognição (*cognitive bias*) são distorções ou ilusões cognitivas, erros sistêmicos de avaliação, que afetam o entendimento e as crenças da pessoa em relação ao mundo a sua volta (KAHNEMAN, 2012). Em outras palavras, todas as pessoas são enviesadas e tendenciosas em boa parte do tempo. Magistrados, assim como todos os seres humanos, por mais que confiem na sua racionalidade, igualmente estão sujeitos, ainda que inconscientemente, a simplificações, inclinações e preconceitos que podem afetar as decisões no trabalho. É por isso que são aplicadas algumas hipóteses para a mitigação da influência dos vieses de cognição, ou seja, o desenviesamento (*debiasing*) (SUNSTEIN e JOLL, 2005, p. 3). A primeira delas é a educação e informação aos julgadores da possibilidade de que estejam decidindo com base em critérios enviesados (PI, PARISI e LUPPI, 2013, p. 8).

Diversos vieses foram identificados pela psicologia e ciência cognitiva. Um exemplo é o viés de ancoragem (*anchoring bias*), que consiste na tendência de que as pessoas

confiem em informações que primeiro tiveram sobre determinado assunto, passando a usá-las como referencial (ou seja, como âncoras), ainda que sejam aleatórias ou infundadas, influenciando diretamente o sentido das suas decisões. Em razão desse viés, “não é difícil imaginar que no recurso de apelação o tribunal se ancore na sentença apelada e que no julgamento colegiado o resto da turma julgadora se ancore no voto do relator” (COSTA, 2016, p. 110).

Outro exemplo é o viés de *status quo* (*status quo bias*), tendência do tomador de decisão de deixar as coisas como estão, pois a mudança exigirá emprego de novo raciocínio (NUNES, LUD e PEDRON, 2018, p. 115-118). O viés pode gerar no magistrado a predisposição de manter o entendimento jurídico que vinha adotando, por simplificação cognitiva, ainda que a jurisprudência que sobreveio não mais embase a conclusão.

Toda essa discussão, contudo, é complexa e de difícil solução atualmente. Em que pese a relevância do contraponto acima, essencial enfatizar que o Poder Judiciário tem a legitimidade embasada na fundamentação das decisões judiciais. Por mais que estas sejam proferidas por magistrados – que possuem, pela própria humanidade, vieses e pré-conceitos, os quais nem sempre ficam claramente expostos na fundamentação –, obrigatoriamente haverá na decisão os motivos de fato e de direito da conclusão judicial, permitindo-se discordância e recurso pelas partes. Diversamente, a decisão que resulte de um sistema de IA não transparente não viabilizaria a compreensão dos motivos em todos os casos.

9. A DESCONHECIDA INTERAÇÃO HOMEM-MÁQUINA

O art. 19, parágrafo único, da Resolução CNJ n. 332/2020 cuidou da supervisão humana, exigindo que “sistemas computacionais que utilizem modelos de Inteligência Artificial como ferramenta auxiliar para a elaboração de decisão judicial” permitam “a supervisão do magistrado competente”. A exigência justifica-se, sobretudo, quando os sistemas desempenham tarefas de minutar decisões.

Em 83% das iniciativas de IA no Judiciário, os resultados obtidos pela máquina passam por algum tipo de validação humana, conforme a pesquisa do CIAPJ/FGV. Em 14% das iniciativas, não há essa validação, tratando-se, em geral, de *chatbots* e sistemas que não desempenham atividades-fim no Judiciário, ou seja, não auxiliam na elaboração de decisões.

A supervisão humana sobre o resultado da atuação da máquina é uma das formas de interação homem-máquina. A configuração dos sistemas de modo a permitir que ambos os lados – homem e máquina – interajam continuamente dá origem a mecanismo chamado "*human in the loop*" (GREEN e CHEN, 2019, p. 8). Um dos principais objetivos da revisão humana sobre o resultado produzido pela máquina é manter o controle da tomada de decisão nas mãos humanas. No entanto, é problemático confiar excessivamente nessa supervisão.

Uma das razões do problema é o viés de automação (*machine bias*), outro viés cognitivo humano, que consiste na tendência humana de privilegiar os resultados gerados pelo sistema automatizado, pela crença de que estes estão embasados em operações matemáticas e, por isso, são científicos e estão corretos. A consequência é que o tomador de decisão acaba por diminuir sua discordância em relação ao resultado de sistemas de IA, aderindo total ou parcialmente, seja porque não reconhece quando os sistemas automatizados erram, seja porque nem sequer dá importância a eventuais informações contraditórias (GREEN e CHEN, 2019, p. 2).

O viés de automação é ponto fundamental da relação homem-máquina. Como alerta Dierle Nunes, justamente pela existência dos vieses humanos, não basta que haja supervisão humana sobre a decisão automatizada. É preciso monitoramento rigoroso na implementação dos alguns sistemas de IA no sistema de justiça, “para que modelos algorítmicos enviesados não possam gerar erros em escala exponencial ou que modelos precisos não sejam utilizados com o intuito de agravar a situação de alguns grupos que já sofrem preconceito” (NUNES, 2021). Novamente, tal monitoramento rigoroso justifica-se, sobretudo, quando os sistemas desempenham tarefas de minutar decisões.

No Judiciário brasileiro, a tendência de seguir o resultado sugerido pela máquina tem mais um motivo. A forma de medir o desempenho de unidades judiciais, seja pelos Tribunais locais, seja pelo CNJ, é eminentemente quantitativa. A produtividade dos magistrados é medida, por exemplo, por indicadores que computam a média de processos baixados. Varas e gabinetes que possuem os menores acervos são mais bem avaliados do que os demais, não havendo análise, em geral, quanto à qualidade das respectivas decisões. O objetivo dessa afirmação não é fazer qualquer crítica à forma de controle de produtividade feita pelo CNJ ou pelos Tribunais locais. A intenção é apenas ressaltar que, em cenário de avaliação numérica da produtividade e em sistema de justiça como o brasileiro, com recorde de processos em tramitação, o viés de automação pode ser ainda mais intenso.

10. DISCRIMINAÇÃO ALGORÍTMICA NAS DECISÕES JUDICIAIS: RISCOS E PRECAUÇÕES

O art. 7º da Resolução CNJ n. 332/2020 dispõe que “as decisões judiciais apoiadas em ferramentas de inteligência artificial devem preservar a igualdade, a não discriminação, a pluralidade e a solidariedade, auxiliando no julgamento justo, com criação de condições que visem eliminar ou minimizar a opressão, a marginalização do ser humano e os erros de julgamento decorrentes de preconceitos”.

A Carta Europeia de Ética sobre o Uso de Inteligência Artificial em Sistemas Judiciais prevê, a seu turno, que o princípio da não discriminação visa prevenir, especificamente, o desenvolvimento ou a intensificação de qualquer discriminação entre indivíduos ou grupos de indivíduos. O documento europeu reforça que deve ser dada especial atenção tanto na fase de desenvolvimento como na de implantação, sobretudo quando o tratamento se baseia, direta ou indiretamente, em dados sensíveis.

Importante discussão relacionada ao princípio da não discriminação refere-se ao risco de ocorrência dos vieses algorítmicos. Esse viés constitui erro de decisão algorítmica e ocorre quando o sistema de IA comporta-se de modo a refletir os valores humanos implícitos nos dados ou nas escolhas associadas à programação, afetando os resultados gerados pela máquina.

As preocupações mais comuns relacionam-se com o risco de incorporação de valores humanos presentes nas decisões anteriores usadas para o treinamento do sistema. Um algoritmo é tão bom quanto os dados usados para treiná-lo, de modo que, se “o algoritmo se baseia em dados históricos repletos de preconceitos, ele reproduzirá, de forma automatizada, os mesmos padrões preconceituosos utilizados como base de seu processamento” (DONEDA, MENDES, SOUZA e ANDRADE, 2018, p. 5).

Exemplo muito citado de ocorrência do viés algorítmico diz respeito ao sistema Compas (*Correctional Offender Management Profiling for Alternative Sanctions* - Perfil de Gerenciamento de Criminoso Correcional para Sanções Alternativas), utilizado na justiça criminal dos Estados Unidos para auxiliar o magistrado na previsão de risco de reincidência do réu.

O sistema ficou conhecido mundialmente em razão do caso *Loomis vs. Wisconsin*. Eric Loomis, acusado de dirigir veículo roubado e fugir da polícia, foi condenado a seis anos

de prisão em agosto de 2013. Ao decidir, o juiz afirmou que o réu foi identificado, por meio da avaliação do Compas, como indivíduo de alto risco para a comunidade. Loomis recorreu, arguindo que o uso do sistema de IA na sentença seria violação do direito ao devido processo legal, já que o funcionamento do modelo computacional é segredo comercial e os magistrados não poderiam avaliar como os fatores são ponderados para a análise da reincidência. O recurso foi negado. Em última instância, a Suprema Corte americana decidiu pela ausência de violação dos direitos do réu, ainda que a metodologia usada para a avaliação de risco de reincidência não tenha sido divulgada⁸.

O algoritmo do Compas não usa dados de raça como entrada, o que poderia sugerir que o sistema é cego para raça. No entanto, uma análise feita por certa empresa particular coletou e analisou o resultado de pontuações de reincidência fornecidas pelo sistema em mais de 10 mil casos de réus criminais no Condado de Broward, na Flórida (ANGWIN, LARSON, MATTU e KIRCHNER, 2016). De acordo com os resultados obtidos, o sistema previa que réus negros tinham quase duas vezes mais chances de serem erroneamente classificados como de maior risco em comparação com réus brancos (45% contra 23%), o que demonstrou a existência de viés discriminatório. Outro estudo sobre o Compas chega a conclusão similar: o sistema indicava que negros são de maior risco, e brancos são de menor risco (TAN, CARUANA, HOOKER e LOU, 2018, p. 305).

Os resultados enviesados gerados pelo Compas, entretanto, não representam a regra em relação aos sistemas de IA usados no Judiciário brasileiro. Para comprovar essa afirmação, serão analisados a seguir os riscos de ocorrência de discriminação em razão de vieses conforme as tarefas atribuídas a cada tipo de sistema.

O primeiro grupo de sistemas, anteriormente identificado, é destinado a auxiliar nas atividades-meio do Judiciário, relacionadas à administração do fórum, como *chatbots* e o reconhecimento facial de quem ingressa no Tribunal, para fins de segurança. Os *chatbots* citados não são preocupantes em relação ao risco de discriminação, já que nem sequer utilizam dados sensíveis. A ferramenta que exige mais acompanhamento é a que realiza reconhecimento facial, tendo em vista os relatos de efeitos discriminatórios em alguns casos (THE GUARDIAN, 2019), embora em aplicações diversas da presente. Note-se, ademais, que o Amon não está diretamente associado ao julgamento da causa.

⁸ State v. Loomis: 881 N.W.2d 749 (Wis. 2016). *Harvard Law Review*, vol. 130, p. 1530-1537, 10 mar 2017, aqui, p. 1532.

O segundo grupo inclui sistemas que ajudam na atividade-fim, na prestação jurisdicional, mas é restrito ao apoio à gestão de secretarias e gabinetes, fazendo triagem e agrupamento de processos similares, classificação da petição inicial, transcrição de audiências etc. Aqui igualmente não há preocupação relevante em relação à incorporação de vieses implícitos. Eventuais erros no resultado gerado pelo sistema podem ser identificados e corrigidos pelo servidor ou magistrado, pois os dados referem-se a situações de natureza processual, e o sistema é usado para análises objetivas, voltadas à automação de volumosas tarefas repetitivas (FERRARI e MENDONÇA, 2022, p. 455), o que confere maior confiabilidade ao resultado e maior produtividade em relação ao ser humano.

O terceiro grupo é composto por menor quantidade de sistemas, os quais auxiliam, direta ou indiretamente, na elaboração de minutas com conteúdo decisório. Difícil prever situações de risco discriminatório em relação aos sistemas que buscam jurisprudência ou fazem a admissibilidade recursal, seja porque os dados do treinamento desses sistemas são de natureza processual e, portanto, dizem respeito a situações objetivas, seja porque é exigida a revisão humana. Essa mesma conclusão aplica-se a diversos sistemas que elaboram minutas instantaneamente. Por exemplo, o sistema Elis, do TJPE, é empregado para agilizar a etapa de conferência e deferimento da petição inicial da execução fiscal. A ferramenta elabora, automaticamente, minuta padrão de decisão e encaminha para a análise e aprovação do magistrado. Segundo a estimativa do TJPE, antes do uso do Elis, 70 mil petições iniciais levavam cerca de um ano e meio para serem conferidas manualmente. Atualmente, esse mesmo volume é analisado pela inteligência artificial em 15 minutos.

Finalmente, o quarto grupo é composto por ferramentas que auxiliam na solução adequada dos conflitos. Os sistemas ICIA, do TRT4, e Concilia JT, do TRF12, não elaboram proposta de acordo, mas apenas analisam o histórico de conciliações do passado e estimam a probabilidade de o processo ser conciliado no estágio em que se encontra. A base de dados é reflexo de conciliação do passado, portanto, pode haver viés originário da prática dos servidores. De todo modo, não parece haver aumento do risco que normalmente ocorreria na seleção feita pelos servidores.

Os sistemas desse quarto grupo não constituem sistemas informatizados de resolução de conflitos por meio da conciliação e mediação (Sirec). Portanto, não representam sistema de *Online Dispute Resolution* (ODR), que estruturam e facilitam a coleta, a troca e a gestão de informações entre as partes, havendo, em alguns casos, proposta de acordo feita de forma automática. Tais plataformas são regulamentadas, no âmbito do Judiciário, pela

Resolução CNJ n. 358/2020 e vêm sendo incorporadas ao sistema público de justiça com o objetivo de torná-lo mais acessível, mais célere e menos oneroso para o cidadão (TIMM e GABRIEL, 2021). As plataformas geram, no entanto, preocupações éticas e maiores cautelas nos Tribunais (TAUK, 2022), diversas das que ocorrem em relação aos sistemas de IA que integram o quarto grupo.

Falta analisar o uso de sistemas de IA no Judiciário brasileiro relacionados à questão criminal, tendo em vista os impactos que podem gerar para os direitos fundamentais do acusado ou réu. No Judiciário, não há sistemas que auxiliam na avaliação da reincidência.

No TJRO, o Peticionamento Inteligente ajuda as delegacias de polícia a enviar documentos (termo circunstanciado, inquérito policial etc) ao PJe. De acordo com as informações prestadas pelo Tribunal para a pesquisa do CIAPJ/FGV, de fato, não foram identificados riscos, incluindo discriminatórios, uma vez que o sistema consiste em acelerar o preenchimento dos campos.

No TJDF, o Saref controla a presença dos apenados da Vara de Execuções Penais em regime aberto por meio de reconhecimento facial, criando alternativa à apresentação presencial de uma população de cerca de 20 mil apenados. Em vez de comparecer pessoalmente, o apenado comprova sua localização por meio do aplicativo. O risco, como informado pelo Tribunal na pesquisa do CIAPF/FGV, é a possibilidade de reconhecer um apenado com metadados de outro, bem como a tentativa de fraude com simulação de imagem digital prévia. A princípio, inexistente identificação de risco de discriminação, no entanto, é relevante que haja acompanhamento constante no seu uso e que a base com os dados pessoais dos apenados não seja utilizada para outros fins, em observância ao art. 6º, I, da LGPD.

Ante a importância do tema, o CNJ, ao editar a Resolução n. 332/2020, dispôs no art. 23 que “a utilização de modelos de Inteligência Artificial em matéria penal não deve ser estimulada, sobretudo com relação à sugestão de modelos de decisões preditivas”, ressaltando seu uso relativo “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”. Com nítida influência das consequências indesejadas decorrentes do caso Compas, em que o sistema de análise de risco de reincidência incorporou vieses discriminatórios, o ato normativo veda que modelos de IA destinados à verificação de reincidência penal indiquem “conclusão mais prejudicial ao réu do que aquela a que o magistrado chegaria sem sua utilização”.

Há observação importante a ser feita neste ponto. Embora a vedação acima seja uma tentativa de evitar efeitos negativos do uso desses sistemas, não se pode esquecer dos vieses de automação e do viés de ancoragem a que estão sujeitos os juízes. Sistemas de análise de risco de reincidência, ainda que apenas auxiliem o magistrado em um dos aspectos da tomada de decisão criminal, podem gerar no juiz a (falsa) impressão de que as sugestões da máquina estão corretas por serem científicas. Por isso, eventual adoção de modelo para prever risco de reincidência precisa vir acompanhada de mecanismos reforçados de precaução, tanto para buscar que o modelo se comporte de maneira não discriminatória, quanto para ensinar os juízes a interagir com o sistema.

A análise feita acima permite concluir que, no estado atual do uso da inteligência artificial no Judiciário brasileiro, as iniciativas em produção e em desenvolvimento não têm aplicação que permita identificar relevante risco discriminatório.

Essa conclusão não dispensa a necessidade de deveres de cuidado no desenvolvimento e implementação da IA no Judiciário, assim como constante monitoramento e auditorias das iniciativas atuais e futuras. Ainda há pouco conhecimento acerca do potencial lesivo dessa tecnologia, bem como sobre as formas de responsabilização por danos dela decorrentes (DE TEFFÉ e MEDON, 2020, p. 301-333), o que pode gerar – indevidamente – menor atenção com as cautelas necessárias para sua utilização no dia a dia.

A urgência por cautelas uniu Elon Musk e Yuval Harari. Em março de 2023, em carta aberta assinada por eles e por mais de 1.300 cientistas, empresários de tecnologia e representantes do meio acadêmico, pediu-se que os experimentos com IA mais sofisticados que o ChatGPT-4 sejam pausados por pelo menos seis meses (FUTURE OF LIFE INSTITUTE, 2023). O ChatGPT-4 (*Generative Pre-trained Transformer*) é a quarta edição do programa que impressionou os usuários, ao envolvê-los em conversas semelhantes às humanas, respondendo a perguntas de acompanhamento, admitindo seus erros, contestando premissas incorretas e rejeitando solicitações inadequadas. O objetivo do pedido de pausa? Os signatários entendem que sistemas poderosos de IA devem ser desenvolvidos apenas quando houver protocolos de segurança compartilhados, a fim de que os sistemas sejam mais seguros, interpretáveis, transparentes e confiáveis, gerando efeitos positivos e riscos administráveis.

11. CONSIDERAÇÕES FINAIS

No estado atual do desenvolvimento tecnológico, a participação humana é indispensável ao longo da ideação, implantação e produção dos sistemas de inteligência artificial, mesmo que haja utilização de *machine learning*. Neste texto, destacou-se a atuação humana nos tópicos relacionados ao funcionamento os sistemas de IA no Judiciário, à forma de treinamento dos algoritmos e à limitação de tarefas que desempenham. Por isso, dizer que tais sistemas possuem algum grau de autonomia depende da noção que se dê a este termo. Afirmar no sentido de que os sistemas não possuem qualquer interferência humana mais confundem do que auxiliam os profissionais do Direito a se familiarizar com os principais aspectos referentes à adoção de IA no Judiciário.

Os resultados apresentados pela pesquisa do Centro de Inovação Administração e Pesquisa do Judiciário, da Fundação Getúlio Vargas, indicaram que, na grande maioria, as iniciativas já em produção e as que estão em desenvolvimento desempenham tarefas de estruturação de dados, com ênfase na categorização e triagem de processos, automação de fluxos de trabalho e recuperação de informações. Justamente por isso, não se identificou relevante risco de discriminação algorítmica. Do mesmo modo, a opacidade inerente aos sistemas, que não permite à máquina fornecer a fundamentação exigida por lei, não parece ser obstáculo para sua adoção nesses casos, em virtude da limitação das tarefas desempenhadas pela IA nos Tribunais.

De forma realista, em nenhuma das iniciativas, a decisão judicial decorre de ato automatizado produzido por suposto juiz robô, entendido este como uma máquina que profere decisões interlocutórias ou sentenças sem supervisão humana. Tais modelos gerariam a preocupação relativa ao alto risco gerado, sobretudo riscos discriminatórios e relativos à opacidade e à desumanização do processo de decisão, a exigir maior cautela e controle. Entretanto, diante da rapidez com que a tecnologia se desenvolve, não é inimaginável haver tarefas cada vez mais sofisticadas desempenhadas pela IA no Judiciário – por isso é preciso agir com precaução desde já.

Em vista da crescente implementação, considerando o conteúdo da Carta Europeia de Ética sobre o Uso de Inteligência Artificial em Sistemas Judiciais e da Resolução CNJ n. 332/2020, sugere-se que os Tribunais: i) mapeiem e avaliem os resultados, quantitativos e qualitativos, do uso dos sistemas em comparação com a situação anterior à respectiva utilização; ii) deem ênfase à transparência, fornecendo explicação sobre os sistemas de IA em suas páginas eletrônicas, em especial noticiando suas tarefas e sua finalidade, de modo acessível e de acordo com padrões de *design* adequados à compreensão pelo público externo; e iii) dediquem-se ao

controle prévio sobre o treinamento do modelo computacional, a exemplo dos cuidados com os dados utilizados e com a formação de equipe, interdisciplinar e que observe a diversidade, para pesquisa, desenvolvimento e implantação, já que o controle posterior é mais difícil e nem sempre é efetivo.

REFERÊNCIAS

ADADI, Amina; BERRADA, Mohammed. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, v. 6, 17 set. 2018, p. 1-23. Disponível: [<https://ieeexplore.ieee.org/document/8466590>]. Acesso: 07.03.2023.

AMARAL, Fernando. *Introdução à Ciência de Dados: Mineração de Dados e Big Data*. Rio de Janeiro: Alta Books, 2016, versão Kindle.

ANGWIN, Julia; LARSON, Jeff; MATTU, Surya; KIRCHNER, Lauren. *How we analyzed the Compas recidivism algorithm*. Pro Publica, 23, mai 2016. Disponível: [<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>]. Acesso: 07.03.2023.

BRAGANÇA, Fernanda. *Justiça digital: implicações sobre a proteção de dados pessoais, soluções on-line de conflitos e desjudicialização*. Londrina: Thoth, 2021.

BURRELL, Jenna. How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, jan-jun, 2016, p. 1-12. Disponível: [<https://journals.sagepub.com/doi/pdf/10.1177/2053951715622512>]. Acesso: 7.03.2023.

CHOWDHURY, Gobinda. Natural language processing. *Annual review of information science and technology*, v. 37, n. 1, pp. 51-89, 2003. Disponível: [<https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/aris.1440370103>]. Acesso: 07.03.2023.

CNJ. *Justiça em números 2022*. Brasília: CNJ, 2022, p.103 e 192. Disponível: [<https://www.cnj.jus.br/wp-content/uploads/2022/09/justica-em-numeros-2022-1.pdf>]. Acesso: 07.03.2023.

COSTA, Eduardo José da Fonseca. *Levando a imparcialidade a sério: proposta de um modelo interseccional entre direito processual, economia e psicologia*. 2016, 187 f. Tese (Doutorado) – Programa de Pós-Graduação em Direito Processual, Pontifícia Universidade Católica de São Paulo, São Paulo, 2016. Disponível: [<https://tede2.pucsp.br/handle/handle/6986>]. Acesso: 07.03.2023.

DE TEFFÉ, Chiara Spadaccini; MEDON, Filipe. Responsabilidade civil e regulação de novas tecnologias: questões acerca da utilização de inteligência artificial na tomada de decisões empresariais. *REI - revista estudos institucionais*, v. 6, p. 301-333, 2020.

DOMINGUES, Juliana; SILVA, Alaís da; SOUZA, Henrique de. Inteligência artificial nas relações de consumo: reflexões à luz do histórico recente. In: *Inteligência artificial: sociedade, economia e Estado*, Capítulo 13. São Paulo, Thomson Reuters Brasil. Disponível: [https://www.direitorp.usp.br/wp-content/uploads/2021/11/AI-nas-Relacoes-de-consumo_FINAL.pdf]. Acesso: 07.03.2023.

DONEDA, Danilo; ALMEIDA, Virgilio A.F. What is Algorithm Governance? *IEEE Internet Computing*, vol. 20, no. 4, pp. 60-63, julho-ago. 2016, Disponível: [https://ieeexplore.ieee.org/document/7529042]. Acesso: 07.03.2023.

DONEDA, Danilo; MENDES, Laura; SOUZA, Carlos Affonso Pereira de; ANDRADE, Norberto Nuno Gomes de. Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal. *Pensar*, Fortaleza, v. 23, n. 4, p. 1-17, out./dez. 2018.

DREXL, Josef; HILTY, Reto M. *et al.* Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective. *Max Planck Institute for Innovation and Competition Research Paper Series*, v.1, outubro 2019. Disponível: [https://ssrn.com/abstract=3465577]. Acesso: 07.03.2023.

FERRARI, Isabela; MENDONÇA, José Vicente Santos de. Sua Excelência o Robô – conceito, riscos e oportunidades das Cortes on-line. In: FUX Luiz; MARTINS Humberto; SHUENQUENER, Valter (Coords). *O Judiciário do futuro*. São Paulo: Revista dos Tribunais, 2022.

FRAZÃO, Ana. Responsabilidade civil de administradores de sociedades empresárias por decisões tomadas com base em sistemas de inteligência artificial. In: *Inteligência Artificial e Direito: ética, regulação e responsabilidade*. FRAZÃO, Ana; MULHOLLAND, Caitlin (Coords). São Paulo: Thomson Reuters Brasil, 2019.

FUTURE OF LIFE INSTITUTE. *Pause Giant AI Experiments: An Open Letter*. Março, 2023. Disponível: [https://futureoflife.org/open-letter/pause-giant-ai-experiments/]. Acesso: 07.03.2023.

GOODFELLOW, Ian; BENGIO, Yoshua e COURVILLE, Aaron. *Deep Learning*. MIT: Cambridge, 2016. Disponível: [http://www.deeplearningbook.org]. Acesso: 07.03.2023.

GOVINDARAJULU, Naveen Sundar; BRINGSJORD, Selmer e LICATO, John. On Deep Computational Formalization of Natural Language. *Computer Science & Cognitive Science Rensselaer Polytechnic Institute (RPI)*, 2013. Disponível: [http://kryten.mm.rpi.edu/SELPAP/2013.FormalMagic/main.pdf]. Acesso: 7.03.2023.

GREEN, Ben; CHEN, Yiling. Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments. *FAT**, janeiro, 2019, Atlanta, Estados Unidos. Disponível: [https://www.benzvegreen.com/wp-content/uploads/2019/02/19-fat.pdf]. Acesso: 07.03.2023.

IWAKURA, Cristiane Rodrigues. *Acesso à justiça e processo civil eletrônico*. Tese de doutorado – Universidade do Estado do Rio de Janeiro. Faculdade de Direito, 2016, 296f. Disponível: [https://www.bdtd.uerj.br:8443/bitstream/1/9331/2/Cristiane%20Rodrigues%20Iwakura_Total.pdf]. Acesso: 07.03.2023.

KAHNEMAN, Daniel. *Rápido e devagar: duas formas de pensar*. São Paulo: Objetiva, 2012 (versão Kindle).

KAUFMAN, Dora. Os Meandros da Inteligência Artificial: Conceitos-chave para Leigos. *Estadão*. 01. fev. 2018. Disponível: [<https://estadodaarte.estadao.com.br/os-meandros-da-inteligencia-artificial-conceitos-chave-para-leigos/>]. Acesso: 07.03.2023.

MALONE, Hugo; NUNES, Dierle. *Manual da Justiça Digital*. Salvador: Editora Juspodivm, 2022.

MARANHÃO, Juliano Souza de Albuquerque; FLORÊNCIO, Juliana Abrusio e ALMADA, Marco. Inteligência artificial aplicada ao direito e o direito da inteligência artificial. *Suprema - Revista de Estudos Constitucionais*, v. 1, n. 1, p. 154-180, jan./jun. 2021. Disponível: [<https://suprema.stf.jus.br/index.php/suprema/article/view/20>]. Acesso: 07.03.2023.

MARRAFON, Marco Aurélio. Filosofia da linguagem e limites da IA na interpretação jurídica (parte II).. *Conjur*, 8 jun 2020. Disponível: [<https://www.conjur.com.br/2020-jun-08/constituicao-poder-filosofia-limites-ia-interpretacao-juridica-parte-ii>]. Acesso: 07.03.2023.

MULHOLLAND, Caitlin. Apresentação. In: *Inteligência Artificial e direito: ética, regulação e responsabilidade*. FRAZÃO, Ana; MULHOLLAND, Caitlin (Coords). São Paulo: Thomson Reuters Brasil, 2020, versão Kindle.

NUENS, Dierle. A supervisão humana das decisões de inteligência artificial reduz os riscos? *Conjur*, 25 jun 2021. Disponível: [https://www.conjur.com.br/2021-jun-25/nunes-supervisao-humana-decisoes-ia-reduz-riscos#_ftnref13]. Acesso: 07.03.2023.

NUNES, Dierle.; LUD, Natanael.; PEDRON, Flávio. *Desconfiando da Imparcialidade dos Sujeitos Processuais: um estudo sobre os vieses cognitivos, a mitigação de seus efeitos e o debiasing*. Salvador: Juspodivm, 2018.

O'NEIL, Cathy. *Algoritmos de Destruição em Massa*. São Paulo: Editora Rua do Sabão, 2021.

PASQUALE, Frank. *The black box society: the secret algorithms that control money and information*. Cambridge: Harvard University Press, 2015.

PI, Daniel; PARISI, Francisco; LUPPI, Barbara. Biasing, debiasing and the law. *University of Minnesota Law School*. Research paper n. 13-08. Disponível: [https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2215117]. Acesso: 07.03.2023.

PECK, Patrícia Pinheiro. *Direito Digital*. 7ª ed. São Paulo: Saraiva, 2021.

RUSSEL, Stuart J. e NORVIG, Peter. *Artificial intelligence: a modern approach*. 3 ed. New Jersey: Pearson Education, 2010.

SALOMÃO, Luis Felipe (Coord). *Tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário*. 2 ed. Rio de Janeiro: CIAPJ/FGV, 2022. Disponível: [https://ciapj.fgv.br/sites/ciapj.fgv.br/files/relatorio_ia_2fase.pdf]. Acesso: 07.03.2023.

SALOMÃO, Luis Felipe; TAUK, Caroline Somesom. Inteligência artificial e direito da propriedade intelectual: fundamentos teóricos e legais de proteção. In: *O direito civil na era da inteligência artificial*. TEPEDINO, Gustavo; SILVA, Rodrigo da Guia (Coords). São Paulo: Revista dos Tribunais, 2020.

SALOMÃO, Luis Felipe; TAUK, Caroline Somesom. Objetivos dos sistemas de inteligência artificial: estamos perto de um juiz robô? *Conjur*, 11 mai 2022. Disponível: [<https://www.conjur.com.br/2022-mai-11/salomao-tauk-estamos-perto-juiz-robo>]. Acesso: 07.03.2023.

SILVER, D., HUANG, A., MADDISON, C. *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature*, v. 550, n. 7676, p. 484–489, 27 Jan 2016. Disponível: [<https://www.nature.com/articles/nature24270>]. Acesso: 7.01.2023. Acesso: 07.03.2023.

SUNSTEIN, Cass, JOLL, Christine. *Debiasing through law*. Chicago: The University of Chicago, 2005. Disponível: [https://papers.ssrn.com/sol3/papers.cfm?abstract_id=590929]. Acesso: 07.03.2023.

SUSSKIND, Richard. *Online Courts and the Future of Justice*. Nova York: Oxford University Press, 2019.

TAN, Sarah; CARUANA, Rich; HOOKER, Giles; LOU, Yin. Detecting Bias in Black-Box Models Using Transparent Model Distillation. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Societ*, dezembro, 2018, p. 303–310. Disponível: [https://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_96.pdf]. Acesso: 07.03.2023.

TAUK, Caroline Somesom. Grandes litigantes e online dispute resolution: alguns princípios éticos. In: FUX Luiz; MARTINS Humberto; SHUENQUENER, Valter (Coords). CHINI, Alexandre; GABRIEL, Anderson de Paiva; PORTO, Fábio Ribeiro (Orgs). *O Judiciário do futuro*. São Paulo: Revista dos Tribunais, 2022.

TIMM, Luciano Benetti; GABRIEL, Anderson de Paiva. Plataformas digitais para resolução consensual de conflitos e acesso à Justiça. *Jota*, 26.04.2021. Disponível: [<https://www.jota.info/opiniao-e-analise/colunas/coluna-da-abde/plataformas-digitais-para-resolucao-consensual-de-conflitos-e-o-acesso-a-justica-26042021>]. Acesso: 07.03.2023.

TURING, Allan. The Computing Machinery and Intelligence. *Mind - A Quarterly Review of Psychology and Philosophy*, Volume LIX, Issue 236, outubro, 1950. Disponível: [<https://academic.oup.com/mind/article/LIX/236/433/986238>].