RPPGD/UFBA
ISSN 2358-4777

INTELIGÊNCIA ARTIFICIAL, VIESES COGNITIVOS E DECISÕES JUDICIAIS

ARTIFICIAL INTELLIGENCE, COGNITIVE BIASES AND JUDICIAL DECISIONS

DOI:

Maurício Requião [1]
Doutor em Direito das Relações Sociaispela Universidade Federal da Bahia.
EMAIL: maurequiao@gmail.com
ORCID: https://orcid.org/0000-0002-1638-6548

RESUMO: O texto analisa as possíveis relações da inteligência artificial (IA) com os vieses cognitivos nas decisões judiciais. Para tanto, inicia apresentando panorama sobre o atual estado de uso da inteligência artificial no Poder Judiciário. Em seguida, discute algumas situações de vieses cognitivos, relacionando-as com as decisões judiciais. Por fim, analisa se o uso da inteligência artificial se coloca como fator favorável para combater os vieses cognitivos nas decisões judiciais, apresentando ainda algumas propostas de solução dos problemas detectados.

PALAVRAS-CHAVE: Inteligência artificial; Vieses cognitivos; Decisões judiciais; Ato de julgar.

ABSTRACT: The text analyzes the possible relationships between artificial intelligence (AI) and cognitive biases in judicial decisions. To this end, it begins by presenting an overview of the current state of use of artificial intelligence in the Judiciary. It then discusses some situations of cognitive biases, relating them to judicial decisions. Finally, it analyzes whether the use of artificial intelligence is a favorable factor in combating cognitive biases in judicial decisions, also presenting some proposals for solving the problems detected.

KEY-WORDS: Artificial intelligence; Cognitive biases; Judicial decisions; Decision making.

1 Introdução

O avanço da inteligência artificial (IA) tem sido recebido com entusiasmo em diversos setores da sociedade, e não é diferente no Poder Judiciário. Suas diversas aplicações, com a possibilidade de facilitar atividades humanas, aumentar a produtividade e, potencialmente, a qualidade de vida, são inegáveis.

As inúmeras vantagens proporcionadas pela IA, entretanto, vêm também acompanhadas de diversas preocupações. Algumas destas são relativas aos impactos da IA no mundo, tal qual, por exemplo, a discussão sobre eventual crise nos postos de

[1] Doutor em Direito das Relações Sociais pela Universidade Federal da Bahia, mestre em Direito Privado, professor adjunto da Faculdade de Direito da UFBA e da Faculdade Baiana de Direito.

trabalho, por conta da substituição do trabalho humano por sistemas de IA. Outras, a seu turno, se relacionam com o próprio funcionamento da IA, e envolvem questões como opacidade, transparência e explicabilidade. A pesquisa que dá origem a este artigo se aproxima mais das discussões deste segundo grupo de problemas.

A proposta deste texto passa pela ideia da verificação de se os atuais modelos de IA estariam aptos a resolver um antigo problema não relacionado com a tecnologia, mas sim com o próprio pensamento humano: o viés cognitivo. Não se entrará nesta introdução em detalhes técnicos sobre o que é o viés cognitivo, o que será feita ao longo do desenvolvimento do texto. Entretanto, apenas a título de esclarecimento inicial, apresenta-se aqui que ele se constitui como um uso inadequado de atalhos mentais no processo do pensamento humano.

Deste modo, a pesquisa que se realizou partiu do objetivo de responder a seguinte questão-problema: como o uso da IA pode influenciar na ocorrência de vieses cognitivos nos julgamentos?

A realização de estudos envolvendo vieses cognitivos nas decisões judiciais, bem como nestas a partir da interferência da inteligência artificial se justifica já que os resultados enviesados podem trazer prejuízos à própria realização da justiça. Isso porque a ocorrência de viés cognitivo acaba por comprometer a correção da decisão judicial (HORTA; COSTA, 2020, p.85). Os estudos em tal campo tanto podem gerar conhecimento sobre a função cognitiva indicada pela existência do viés, como são aptos a oferecer propostas de correção para evitá-lo (CAVERNI; FABRE; GONZALES, 1990, p.10), contribuindo, em qualquer dos casos, para o caminho na busca por decisões judiciais mais adequadas.

Em busca da solução à questão-problema proposta, na primeira seção deste artigo se realiza abordagem que traça o panorama atual do uso da inteligência artificial nos Tribunais brasileiros. Isso é feito a partir do cruzamento de informações de pesquisas anteriores, apresentando não apenas uma revisão da literatura sobre o tema, mas também a elaboração deste conhecimento com reflexões que só se tornaram possíveis em período temporal já posterior ao das referidas pesquisas.

Na segunda seção é abordada a figura do viés cognitivo, tanto para sua conceituação, como para apresentação de algumas de suas variantes. Também nesta já

se realiza a análise do problema o vinculando de modo específico à questão das decisões judiciais.

Numa terceira e última seção do desenvolvimento, se realiza análise do impacto do uso da IA nos vieses cognitivos. Para tanto, há não apenas análise de situações já ocorridas, como também são formuladas hipóteses de possíveis problemas ainda não diagnosticados, sendo ao final oferecidas algumas soluções.

Por fim, nas conclusões, se responde de modo mais objetivo à questão-problema apresentada, com base nos resultados obtidos ao longo da pesquisa.

2 Breve panorama da inteligência artificial nos Tribunais brasileiros

Como já salientado na introdução, o uso da IA pelos Tribunais brasileiros vem crescendo em forte ritmo. Pesquisas realizadas pelo Conselho Nacional de Justiça (CNJ) permitem identificar o uso crescente da IA no Judiciário brasileiro. De acordo com o comparativo entre pesquisas realizadas pelo órgão nos anos de 2021 e 2022, houve aumento de 171% no número de projetos envolvendo uso de IA no Poder Judiciário (AGÊNCIA CNJ DE NOTÍCIAS, 2022).

Em artigo de Tauk e Salomão (2023) também podem ser encontradas várias notícias sobre o uso da IA nos Tribunais nacionais, com base em dados de pesquisa empírica realizada no ano de 2022. Isto significa que, por mais que os números apontados a seguir já sejam surpreendentes, a tendência é que no atual momento as iniciativas envolvendo IA já sejam ainda mais numerosas.

De acordo com os autores

> 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 (TAUK; SALOMÃO, 2023, p.3).

Destes sistemas de IA, 77% utilizam *machine learning*[2], seja nos modelos de algoritmos supervisionados, não supervisionados ou de aprendizado por reforço (TAUK;

[2] Machine learning é um campo da IA em que o sistema computacional se torna progressivamente capaz de resolver tarefas sem que precise receber instruções específicas para tanto. Assim, progressivamente, a máquina vai "aprendendo". Existem diversos modelos de machine learning, com variados graus de interferência humana no processo acima descrito.

SALOMÃO, 2023, p.5), dado que é condizente com o cenário mundial atual, já que este tem sido o modelo de IA dominante na atualidade.

A mesma pesquisa encontrou 64 modelos de IA no Judiciário, entre os em ideação, os em desenvolvimento e os já implantados, que os autores dividiram, para fins didáticos, em quatro grupos (TAUK; SALOMÃO, 2023, p.11), cuja reprodução é útil para as reflexões que serão feitas no presente artigo.

Assim, no primeiro grupo estariam os sistemas de IA que se prestam à atividade-meio do Judiciário, como administração, gestão de recursos e de pessoal, não auxiliando o magistrado na prestação jurisdicional. Nos segundo e terceiro grupos, ao revés, estariam as IA que se prestam à atividade-fim da prestação jurisdicional. A distinção entre este segundo e terceiro grupos se daria porque no segundo estariam os modelos que, compondo a maioria, "se destinam à automação dos fluxos de movimentação do processo e das atividades executivas de apoio aos juízes por meio da execução de tarefas pré-determinadas". O terceiro, a seu turno, abrangeria os que "colaboram na elaboração de minutas com conteúdo decisório de sentença, votos ou decisões interlocutórias" (TAUK; SALOMÃO, 2023, p.11-12), atuando, portanto, de modo mais direto na formação das decisões. Por fim, no quarto grupo, estariam "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" (TAUK; SALOMÃO, 2023, p.13).

Para além da classificação oferecida, os autores salientam que, em nenhum dos modelos avaliados, "há a interpretação de textos legais, a elaboração de argumentação jurídica e, muito menos, a tomada de decisão pela máquina" (TAUK; SALOMÃO, 2023, p.13). Até o momento das referidas pesquisas, portanto, afirmam os autores que estariam mantidos o raciocínio jurídico e o dever de decidir exclusivamente ao magistrado. Entretanto, discorda-se parcialmente desta conclusão dos autores, já que, especialmente as atividades previstas no terceiro grupo, interferem de modo direto no ato de julgar.

Outra iniciativa que deve ser destacada neste panorama é a criação pelo CNJ da Plataforma Sinapses/Inteligência Artificial como consequência da sua Resolução 332/2020 do órgão. Foi a Sinapses instituída "como plataforma nacional de armazenamento, treinamento supervisionado, controle de versionamento, distribuição

e auditoria dos modelos de Inteligência Artificial, além de estabelecer os parâmetros de sua implementação e funcionamento" (CNJ, 2020), o que reforça o interesse crescente sobre o tema no âmbito do Judiciário.

Também a fala do Ministro Barroso (2023), no Seminário Direito e Tecnologia, foi bem representativa deste movimento progressivo do uso da IA no Judiciário, ao reiterar sua iniciativa em abrir chamada pública para três encomendas no campo da tecnologia, duas delas diretamente ligadas à IA.

A primeira, seria a criação de uma IA capaz de analisar um processo judicial para fazer um resumo dos fatos relevantes, da decisão de primeiro grau, da decisão de segundo grau e razões do recurso. Destacou o Ministro que isto seria uma revolução para os que atuam nos Tribunais, que não mais precisariam percorrer milhares de páginas para o julgamento, embora destaque que tal procedimento deveria se dar sempre com supervisão humana.

A segunda demanda ligada à IA seria a criação de uma espécie de ChatGPT estritamente jurídico, "alimentado com a Jurisprudência do Supremo, STJ e de todos os Tribunais Estaduais, e que desenvolva a capacidade de, alimentado com o fato relevante, propor uma minuta de decisão". Sobre este ponto destacou ainda que com isso não pretende acabar com a atividade do magistrado, já que o que a IA faria seria o mesmo trabalho que hoje é feito pelos assessores, cabendo ainda o estudo da proposta oferecida para a decisão final por parte do juiz[3].

Dialogando com a fala do Ministro Barroso, e ainda na análise sobre o panorama do uso da IA no ato de julgar, são necessárias algumas considerações para além dos dados informados pelas pesquisas anteriormente citadas. Uma primeira, é que estes são dados colhidos a partir de iniciativas oficiais de implantação da IA no Judiciário, não dando conta do uso individual de IA pelos magistrados ou demais servidores. Uma segunda, diretamente vinculada à primeira, é que as pesquisas se realizaram antes da chegada do ChatGPT, que foi disponibilizado ao público a partir de novembro de 2022.

Neste intervalo de tempo de um ano, muita coisa mudou em termos de popularização no uso da IA, por conta do ChatGPT e demais sistemas de IA generativa

[3] A terceira demanda, ligada ao mundo digital, mas não à IA, seria a criação de uma interface comum para todos os Tribunais, com o objetivo de facilitar o acesso dos advogados e demais usuários.

que vêm sucessivamente surgindo. Já houve, por exemplo, episódio amplamente divulgado, em que juiz baseou decisão em "jurisprudência" inexistente do Superior Tribunal de Justiça (STJ), inteiramente inventada pelo ChatGPT, e que levou a investigação pelo CNJ (CONJUR, 2023). Paralelamente, cresce o número de juízes que se manifestam em suas redes sociais sobre o uso de ferramentas como o ChatGPT na sua prática cotidiana, alguns daqueles, inclusive, ministrando cursos cujo objetivo é otimizar o uso do ChatGPT na atividade jurídica.

3 Vieses cognitivos e julgamentos por humanos

O ser humano, de modo consciente ou não, toma decisões a todo momento. Elas vão desde escolhas simples e normalmente pouco refletidas, como sobre por qual lado vai levantar da cama, até decisões mais complexas e racionalizadas, como, por exemplo, o planejamento de sua aposentadoria. De todo modo, cada uma dessas decisões decorre de um processo de raciocínio, em que o sujeito considera uma série de informações numa estrutura lógica, mas também é constantemente influenciado pelo que se costuma considerar como instinto.

Ao tomar decisões, para além das evidências factuais de cada caso, a pessoa leva em consideração, às vezes de modo inconsciente, coisas como seus aprendizados e experiências pretéritas. De toda sorte, a pessoa humana possui limitação na sua capacidade de pensamento, não conseguindo acessar e processar racionalmente todas as variáveis para cada situação analisada. Por conta disso, desenvolve o processo de heurística, que se caracteriza como um atalho mental para chegar a determinadas soluções, atuando, de certa forma, como uma intuição. O viés cognitivo, por sua vez, se caracteriza como a disposição humana de, dada a limitação acima apontada, utilizar a heurística de modo inapropriado para realizar seu raciocínio e chegar a suas conclusões. No viés cognitivo, portanto, há uma distorção no processo de análise da informação, que acaba por levar a um resultado de alguma maneira inadequado (KORTELING; TOET, 2020, p.7-8).

O modo de pensar humano, de acordo com Kahneman (2012, p.29), é dividido na psicologia entre dois sistemas. O Sistema 1 é responsável pelo pensar de modo rápido e automático, executando diversas tarefas que muitas vezes são realizadas quase sem reflexão. Imaginar a resposta para uma conta matemática simples, como

1+1=?, é algo para o que o cérebro humano processa a resposta ainda que o sujeito não queira, como destaca o autor.

Já o Sistema 2 trata das atividades mentais que requerem mais trabalho, comumente associadas "com a experiência subjetiva de atividade, escolha e concentração". Se a conta indicada como exemplo fosse 134x41=?, entraria em ação o Sistema 2, requerendo que o sujeito pare, foque sua atenção e realize algum esforço para chegar no resultado.

Assim, o pensar humano normalmente se dá a partir do Sistema 1, alimentando o Sistema 2 com "impressões, intuições, intenções e sentimentos", que podem ser endossadas pelo segundo se tornando então ações voluntárias. E prossegue Kahneman (2012, p.33-34): "Quando tudo funciona suavemente, o que acontece na maior parte do tempo, o Sistema 2 adota as sugestões do Sistema 1 com pouca ou nenhuma modificação. Você geralmente acredita em suas impressões e age segundo seus desejos, e tudo bem – normalmente."

A partir desta explicação sobre os dois modos de pensar, se percebe que o viés cognitivo acontece, como já dito em outras palavras, quando esta heurística, que é o que é realizado pelo Sistema 1, leva o Sistema 2 a decidir, sob a aparência da mais plena lógica e racionalidade, de um modo mais informado por outros fatores, como os desejos e sentimentos do sujeito.

O viés cognitivo aparece de modo persistente nas mais variadas situações. Ela é parte do modo como a sociedade funciona, ainda que a maioria dos sujeitos o realize sem ter qualquer consciência disso. Inclusive, estudos realizados também apontam que a maior parte das pessoas acredita ser menos afetada pelo viés cognitivo do que a média geral (KORTELING; TOET, 2020, p.8). Estas informações, por si, apontam que o viés cognitivo, mais do que unicamente um defeito do modo de pensar humano, é, ao revés, uma característica deste.

A identificação de algo como um viés cognitivo costuma partir da comparação com o que se considera a ausência de viés. Esta ausência seria a norma, o modo adequado de funcionamento, e o viés, por sua vez, seria o desvio de tal norma, que acontece não como um erro qualquer, mas sim de modo sistemático e recorrente (CAVERNI; FABRE; GONZALES, 1990, p.7). Assim, por exemplo, se uma pessoa erra sobre quanto material vai precisar para certa obra por ter feito uma conta errada, não

estaria caracterizado o viés cognitivo. Ao revés, se ela realiza uma estimativa otimista demais, comprando menos material do que seria necessário, há chances de que isso tenha acontecido por um viés cognitivo.

As razões para a ocorrência do viés cognitivo ainda não são conhecidas, havendo teorias que, por exemplo, o vinculam com a recente, em termos evolutivos, capacidade cognitiva do ser humano (KORTELING; TOET, 2020, p.12). Estas discussões, entretanto, não serão tratadas no presente artigo, por fugirem do objeto mais específico aqui pretendido. O que importa é que ocorrem, ainda que não se possa conhecer com certeza a sua causa.

Para além de pesquisas realizadas por pessoas com formação no Direito, importante destacar também que as pesquisas nos campos de Psicologia Experimental e Economia Comportamental, inclusive sobre vieses cognitivos, já vêm sendo feitas há mais de vinte anos analisando as práticas do Judiciário brasileiro, notadamente do Supremo Tribunal Federal, tanto por conta da maior visibilidade, como pela maior facilidade no acesso à informação (HORTA, COSTA, 2020, p.78). Entretanto, há mais avanço nas pesquisas que tratam do comportamento e influência de fatores políticos ou ideológicos, do que propriamente de psicologia da decisão judicial (HORTA, COSTA, 2020, p.81).

Dentro do campo jurídico, um dos fatores apontados como gerador de vieses cognitivos seria o “limitado tempo ofertado para o deslinde de cada caso” (NUNES, p.74). Esta afirmação possui lógica se pensarmos as razões pelas quais os vieses se ocasionam, já que, dispondo de menos tempo, aumenta a possibilidade de uso da heurística com falhas no processo decisório.

Estes vieses cognitivos são explorados na literatura com diversas classificações, que tentam dissecar de modo mais pontual as possíveis ocorrências deste fenômeno. Ao analisar a produção científica, se nota que muitas destas classificações se sobrepõem em alguma medida, ou dialogam de modo claro. Muitas delas, por exemplo, compartilham a característica de dar preferência às conclusões que são compatíveis a crença do sujeito que está tomando a decisão (KORTELING; TOET, 2020, p.8). Não há neste texto, portanto, o objetivo de listar de modo exaustivo tais classificações, pois resultaria unicamente em uma enumeração incompleta e, ainda assim, rasa.

Feito este esclarecimento preliminar, nesta seção serão apresentadas algumas destas classificações, bem como será analisado como os vieses cognitivos poderiam interferir no ato de julgar por parte de um magistrado, explorando ainda possíveis correções.

O chamado viés de confirmação é a tendência de a pessoa interpretar fatos e informações do modo que melhor confirme sua visão preestabelecida (KORTELING; TOET, 2020, p.8). Ou seja, o sujeito dará maior valor às suas ideologias e crenças de como as coisas funcionam, do que às evidências que lhe são apresentadas. (TABAK; AMARAL, 2018. p. 477). Assim, novas informações a que o sujeito tenha acesso, serão valoradas de maneira diferente, conforme confirmem ou venham de encontro ao que ele pensa (ANDRADE, 2019, p.519).

Imagine-se, por exemplo, uma pessoa que admira certo político. Ela tenderá a acreditar mais em notícias que confirmem a idoneidade deste sujeito, do que em notícias que lhe sejam desabonadoras.

Os algoritmos das redes sociais se valem deste viés de confirmação, criando as já famosas “bolhas”, ao selecionar para a visualização pelo usuário assuntos que sejam ao agrado do seu modo de pensar. Assim, por exemplo, uma pessoa que tenha inclinação política mais de esquerda, receberá conteúdos que reforcem esse posicionamento, o mesmo ocorrendo com as pessoas que se situam no espectro político oposto.

No ato de decidir, o viés de confirmação de um juiz pode o inclinar a dar decisões que *confirmem* o seu modo de pensar. Isso ocorre porque, ao se deparar com os litígios, embora o ideal fosse que sua mente processasse as alegações dos advogados e as provas produzidas no processo dentro da racionalidade, sua limitação humana acabará por o levar a incorporar na decisão aspectos decorrentes do seu viés de confirmação. O peso que dê ao depoimento de uma testemunha, por exemplo, pode ser influenciado pela questão de esta confirmar ou não sua impressão pessoal sobre o caso.

O viés de adesão, por sua vez, se caracteriza como “a tendência de pensar, acreditar ou decidir de uma determinada forma porque outras pessoas assim o fazem” (ANDRADE, 2019, p.520). Ele serve para explicar, de certa maneira, o efeito manada por vezes visto no comportamento humano.

Já o viés de ancoragem “ocorre quando o indivíduo é exposto a uma informação ou experiência previamente à decisão que servirá de base (ou âncora) a seu raciocínio ao considerar estimativas e realizar julgamentos” (TABAK; AMARAL, 2018. p. 478). Assim, se uma pessoa é exposta a um número qualquer, antes de ter que fazer certa estimativa para uma quantidade desconhecida, ela tenderá a realizá-la de modo a aproximá-la deste número (KAHNEMAN, 2012, p.152-153). Esta informação inicial, portanto, pode influenciar no modo como a pessoa irá realizar sua estimativa.

No ato de julgar isso pode ser encontrado, por exemplo, a partir do uso irrefletido de súmulas e ementas, sem que se realize leitura mais aprofundada dos fundamentos, como causa para decidir em determinado sentido (NUNES, 2015, p.73). Imagine-se, por exemplo, um juiz que deva fixar uma indenização numa sentença cível ou o tempo de detenção num caso criminal. A exposição dele a certos números, pode influenciar na sua decisão final em cada um dos casos.

E o viés da ação ocorre quando há uma reação exagerada, ou a ausência de ação, diante de uma situação de risco ou incerteza (TABAK; AMARAL, 2018. p. 481). Na atuação do juiz, se poderia exemplificar um viés de ação quando aquele, ao julgar um caso penal que tenha alcançado grande clamor popular, acaba por fixar sentença mais alta do que usualmente fixaria.

Apesar da inevitabilidade dos vieses cognitivos, isto não significa que não existam medidas que possam ser adotadas para tentar minimizar a sua ocorrência. É o que se chama na psicologia de desenviesamento (*debiasing*).

No campo estritamente jurídico, NUNES (2015, p.73) afirma que a própria realização do que chama de processo constitucionalizado, com garantia do contraditório e devido processo constitucional, seria um modo de desenviesamento. Aponta ainda que a colegialidade, ou seja, o julgamento realizado por colegiados, em contraponto à decisão monocrática, poderia vir a ser também um caminho para afastar os vieses, embora tal afirmação ainda dependeria de pesquisa empírica para ser comprovada, e haja, em contraponto, pesquisas que indicam que a colegialidade não se realiza do modo devido (NUNES, 2015, p.74-75).

Wojciechowski e Rosa (2021) também oferecem soluções para combater o enviesamento nas decisões judiciais, tratando de modo mais específico dos campos do Direito Penal e Direito Processual Penal. Iniciam suscitando a necessidade de tomada

de consciência pelos magistrados do uso das heurísticas e vieses nos julgamentos, como ponto de partida para o seu combate. Na sequência, uma das soluções apontadas é a sugestão de que os juízes utilizem constantemente uma técnica de falseamento em relação aos fatos que presumam como verdadeiros, para, por exemplo, evitar formar juízo prévio, logo, viés cognitivo, em relação a um acusado por conta de ele morar numa região onde o índice de criminalidade é alto. Sugerem também a superação do princípio da verdade real, por vezes defendido no Direito Penal, pois ao atuar para além do limite das provas trazidas pelas partes, com o intuito de buscar uma inalcançável verdade real, os juízes estariam favorecendo a ocorrência dos seus vieses de confirmação. Com base neste mesmo raciocínio, defendem o seguimento do sistema processual-penal acusatório, já que o distanciamento de tal modelo, através, por exemplo, da busca de provas de ofício pelo magistrado, também favoreceria a ocorrência de vieses.

4 Vieses cognitivos e IA: um fator de correção?

Teoricamente, a IA pode se apresentar como uma ferramenta com potencial para diminuir a ocorrência de vieses cognitivos. Afinal, como tem capacidade de analisar um número muito maior de dados do que os humanos, estaria também apta a chegar a soluções mais racionais e, portanto, adequadas.

Além disso, os vieses cognitivos, conforme visto, são uma característica presente no pensamento humano. Isso poderia levar à conclusão de que, em tese, não estaria presente numa racionalidade não-humana, como aquela encontrada nas inteligências artificiais. Até o presente momento, entretanto, esta hipótese não parece se confirmar. Os vieses têm também sido encontrados nas IAs que funcionam com base em machine learning, apresentando vieses, bem como estereótipos e preconceitos tipicamente humanos (KORTELING; TOET, 2020, p.11). E estes são apenas problemas por conta dos vieses cognitivos que, destaque-se, estão longe de serem os únicos relacionados ao uso da IA para tomar decisões judiciais[4].

De modo preliminar à análise dos vieses, se faz necessária alguma discussão sobre a importância da regulação da IA, e o modo como isso deve ocorrer. Isto porque, tais pontos centrais, são decisivos em relação às soluções que podem ser adotadas.

[4] Pense-se, por exemplo, nas questões que envolvem opacidade, transparência e explicabilidade.

MEIRA (2023), durante sua fala no Seminário Direito e Tecnologia, afirmou textualmente que “a regulação tem uma regra básica, que é o seguinte, não faça como o Brasil está fazendo”. Justificou tal crítica diante de uma ausência de conhecimento no Brasil sobre o tema da IA para realizar uma regulação global acerca do tema. Defendeu, em seguida, o que afirmou ser o modelo norte-americano, que demanda que a empresa desenvolvedora de IA informe se sua atividade é de risco, para que, posteriormente, se houver falsidade em tal declaração venha ela a ser responsabilizada. Haveria, então, em tal modelo, uma construção inovadora da regulação em conjunto com a indústria, que é quem está criando o problema e também as oportunidades, plataformas e dimensões da IA.

É certo que há críticas sobre o PL 2338/2023, e que cabem discussões sobre modelos de regulação, como, por exemplo, se deve ela ser geral ou setorial. Entretanto, esperar que a IA se desenvolva, para só então se trazer qualquer regulação, parece medida que pode trazer prejuízos à própria democracia, e que só beneficia às Big Techs. Seja pela aprovação do PL 2338/2023, ou pela criação de outros instrumentos normativos, a regulação é necessária.

Ademais, o próprio foco sugerido por Meira em responsabilização, seja civil ou penal, já que não houve especificação em sua fala, é insuficiente, pois se desvia do principal ponto do problema. Afinal, é certo que há muitos danos que são irreparáveis, e o melhor caminho é sempre evitar sua ocorrência, o que só é possível a partir da prevenção e regulação. Evitar a lesão é sempre melhor do que indenizar o dano.

Como exemplo de algum início de regulação setorial da IA se tem, para o próprio Judiciário, a Resolução 332/2020, do Conselho Nacional de Justiça (CNJ), que “dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências”. Esta, inclusive, dialogará com uma das soluções apontadas neste artigo para afastar o viés cognitivo da IA, conforme caso a seguir apresentado.

Um comum problema que pode ser apontado como decorrente de vieses cognitivos e que, em concreto, já foi detectada como reproduzido pela IA, é o preconceito. A ocorrência deste, dentre outros possíveis vieses, dialoga francamente com o previamente explicado viés de confirmação.

O'Neil (2020) apresenta análise sobre o sistema *Level of Service Inventory* (LSI-R), aplicado a milhares de detentos desde sua invenção em 1995, para os categorizar como sendo de risco alto, médio ou baixo. Tal categorização seria utilizada para incluir os com pontuação de alto risco em programas antirreincidência após já estarem presos, mas foi também utilizado para outros fins, como uma espécie de pontuação para guiar os juízes na determinação das sentenças criminais.

Embora tal programa não perguntasse nada sobre a raça do detido, posto que seria ilegal, as perguntas realizadas sobre sua vida pessoal, tais quais "qual foi a primeira vez em que você se envolveu com a polícia", e antecedentes criminais de parentes e amigos, faziam com que os homens negros, em sua maioria mais pobres, alcançassem maiores pontuações de risco em relação àqueles que tiveram uma vida mais abastada. Como aponta a autora, é necessário avaliar "se de fato eliminamos o viés humano ou simplesmente o camuflamos com tecnologia".

O caso trazido por O'Neil mostra o que é convencionalmente referido como discriminação algorítmica, que se dá não apenas no contexto de julgamentos judiciais, mas em diversos outros, que apenas não serão aqui explorados por fugirem da proposta deste artigo.

Na mesma linha, o Ministro Barroso (2023), em sua fala no Seminário Direito e Tecnologia, destacou os prejuízos da discriminação algorítmica, relatando a existência de pesquisas em outros países que apontaram a ocorrência de discriminação em razão da raça ou condição social para o cálculo de progressão de regime prisional.

A discriminação algorítmica acontece, em regra, por duas razões: "i) quando os algoritmos refletirem os preconceitos humanos (conscientes ou não) embutidos desde a programação; ii) quando entrarem em contato com bases de dados contendo vieses preconceituosos, o que faz com que o algoritmo 'aprenda' a discriminar." (REQUIÃO; COSTA, 2022, p.4).

A preocupação em combater a discriminação algorítmica pode ser deduzida, inclusive, a partir da própria Resolução 332/2020, do CNJ, que, em seu art.7°, trata da não discriminação[5]. Um dos fatores essenciais para o alcance da não discriminação é a

[5] Art. 7º 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.

existência de equipes plurais, tanto no desenvolvimento da IA, como ao longo do seu uso. Administração pública e empresas privadas poderiam, inclusive, quando necessário, se valer do instrumento das ações afirmativas para compor quadros plurais (REQUIÃO; COSTA, 2022, p.19-20).

Passando a outro tópico, no que toca aos vieses de adesão e ancoragem, acredita-se que o uso da IA pode levar ao aumento destes. Sistemas de IA que lidem com a atividade-fim dos magistrados, ou seja, que de alguma forma auxiliem no ato de julgar, podem fazer com que estes se tornem meros repetidores das sugestões das máquinas.

Pense-se, por exemplo, um sistema de IA que ajude a modular penas criminais, ou agrupe jurisprudência estabelecendo valores médios de indenização fixados em certos tipos de casos cíveis. Haverá a natural tendência, sobretudo por parte dos magistrados menos experientes, em seguir a sugestão automatizada, até por ser maior o ônus de discordar.

Se, por um lado, isso pode favorecer maior uniformização da jurisprudência e diminuição do ruído[6] nas decisões judiciais, por outro, pode aumentar a mecanização da atividade de julgar e gerar magistrados menos confiantes a decidirem casos mais complexos por si mesmos. Além disso, caso haja alguma outra espécie de viés na própria base de dados da IA, estes seriam ainda mais amplamente propagados, através dos efeitos decorrentes dos vieses de ancoragem e adesão.

Por outro lado, pensando num argumento espelho em relação ao acima apresentado, parece promissor o uso da IA para evitar o viés de ação. Afinal, as chances de uma reação exagerada ou uma inação indevida, por parte do magistrado,

---

§ 1º Antes de ser colocado em produção, o modelo de Inteligência Artificial deverá ser homologado de forma a identificar se preconceitos ou generalizações influenciaram seu desenvolvimento, acarretando tendências discriminatórias no seu funcionamento.
§ 2º Verificado viés discriminatório de qualquer natureza ou incompatibilidade do modelo de Inteligência Artificial com os princípios previstos nesta Resolução, deverão ser adotadas medidas corretivas.
§ 3º A impossibilidade de eliminação do viés discriminatório do modelo de Inteligência Artificial implicará na descontinuidade de sua utilização, com o consequente registro de seu projeto e as razões que levaram a tal decisão.

[6] Há ruído quando pessoas que deveriam estar de acordo sobre certo tema alcançam resultados muito díspares entre si. Assim, por exemplo, quando juízes, examinando casos muito parecidos emitem decisões enormemente diferentes, muitas vezes por fatores alheios ao próprio caso (KHANEMAN; SIBONY; SUNSTEIN, 2021).

quando confrontado com uma situação de risco ou incerteza, tendem a diminuir caso tenha maior acesso a soluções facilitadas pela IA.

Ademais, a IA pode também ser útil de uma forma geral para o combate aos vieses cognitivos. Os magistrados, diante da alta demanda do Judiciário, tendem a recorrer mais a heurísticas para julgar com mais rapidez, o que favorece a ocorrência de viés cognitivo. Afinal, como adverte ANDRADE (2019, p.523), “a pressa pode significar eficiência sem justiça”.

Assim, com a IA realizando algumas tarefas rotineiras, que se constituem como atividade-meio, como as indicadas na primeira seção deste artigo, mas que também consomem tempo do magistrado e demais servidores levaria a mais tempo dedicado ao ato de julgar, reduzindo, em teoria, a chance de que o julgamento seja feito com alguns dos vieses cognitivos.

Acredita-se que o sucesso do uso da IA no Judiciário depende de constante análise para enfrentamento dos vieses. Isto pode ser feito a partir da criação e manutenção dos chamados *red teams*. O termo *red team*

> tem sido utilizado para abranger uma ampla variedade de métodos de avaliação de riscos para sistemas de inteligência artificial, incluindo descoberta qualitativa de capacidades, testes de estresse de mitigação, "red teaming" automatizado usando modelos de linguagem, fornecendo feedback sobre a escala de risco para uma vulnerabilidade específica, etc. (OPENAI, 2023)

Assim, a manutenção de uma equipe multidisciplinar, que possa tanto verificar problemas na programação, como analisar fatores éticos e jurídicos, é de extrema importância para que se possa verificar e corrigir a existência de vieses na programação de qualquer IA. Em se tratando de uma que atue em ponto tão importante para o funcionamento da sociedade, como é o Poder Judiciário, a necessidade do *red team* se apresenta como ainda mais relevante.

Por fim, a realização de cursos para os magistrados e demais sujeitos do Poder Judiciário envolvidos no ato de julgar, que versem sobre os temas dos vieses cognitivos e do próprio funcionamento da IA, também se colocam como excelente medida para combater os vieses cognitivos. A tomada de consciência sobre um problema é, sem dúvidas, o primeiro passo para a sua solução.

5 Conclusões

Embora a IA se coloque como uma tecnologia revolucionária, o seu uso, por si só, não possui o condão mágico de resolver os problemas humanos. Conforme visto, as consequências são plurais e diversas nas relações entre o seu uso e os vieses cognitivos.

Por um lado, a IA se coloca como instrumento útil, capaz de poupar o tempo dos sujeitos envolvidos no ato de julgar, propiciando que possam investir maior esforço para a realização desta atividade-fim e, por conseguinte, teoricamente, sejam capazes de decidir de modo menos enviesado. Além disso, sua capacidade de tornar parâmetros racionais de decisão mais disponíveis, também é fator que pode ser positivo no combate ao enviesamento.

Por outro lado, entretanto, é já provado que o uso da IA também pode reproduzir ou favorecer a ocorrência de vieses cognitivos. Talvez isso signifique que os vieses não são restritos à inteligência humana, mas simplesmente à inteligência, e continue a se reproduzir também na IA, ainda que com futuros avanços.

Tais constatações, portanto, não levam ao total impedimento do uso da IA no ato de julgar, sobretudo quando destacada a importância da supervisão humana. Trazem à luz, porém, a necessidade de contínuos esforços, a exemplo das soluções oferecidas neste texto, para diminuir as chances de julgamentos enviesados a partir do uso da IA como ferramenta.

6 Referências

AGÊNCIA CNJ DE NOTÍCIAS. **Justiça 4.0: Inteligência Artificial está presente na maioria dos tribunais brasileiros**. Disponível em <https://www.cnj.jus.br/justica-4-0-inteligencia-artificial-esta-presente-na-maioria-dos-tribunais-brasileiros/>. Acesso em 10 dez. 2023.

ANDRADE, Flávio da Silva. A tomada da decisão judicial criminal à luz da psicologia: heurísticas e vieses cognitivos. **Revista Brasileira de Direito Processual Penal**, Vol.5, n.1, 2019, p.507-540. Disponível em: https://www.redalyc.org/articulo.oa?id=673971413015 Acesso em: 21 fev. 2024.

BARROSO, Luís Roberto. **Seminário Direito e Tecnologia**, 2023. Disponível em <https://app.openfy.co/home/>. Acesso em 21 nov. 2023.

CAVERNI, Jean-Paul; FABRE, Jean-Marc; GONZALES, Michel. **Advances in psychology, 68: Cognitive Biases**. Amsterdam: Elsevier, 1990.

CONJUR. **CNJ vai investigar juiz que usou tese inventada pelo ChatGPT para escrever decisão**. Disponível em < https://www.conjur.com.br/2023-nov-12/cnj-vai-investigar-juiz-que-usou-tese-inventada-pelo-chatgpt-para-escrever-decisao/>. Acesso em 20 nov. 2023.

CNJ. **Plataforma Sinapses**. Disponível em <https://www.cnj.jus.br/sistemas/plataforma-sinapses/>. Acesso em 10 fev. 2024.

KHANEMAN, Daniel. **Rápido e devagar: duas formas de pensar.** Rio de Janeiro: Objetiva, 2012.

KHANEMAN, Daniel; SIBONY, Olivier; SUNSTEIN, Cass R. **Ruído: uma falha no julgamento humano**. Rio de Janeiro: Objetiva, 2021.

KORTELING, J.E.; TOET, A. Cognitive biases. S. *Della Sala:* ***Reference Module in Neuroscience and Biobehavioral Psychology***. Amsterdam-Edinburgh: Elsevier ScienceDirect. 2020. Disponível em <https://doi.org/10.1016/B978-0-12-809324-5.24105-9>. Acesso em 12 fev. 2024.

MEIRA, Sílvio. **Seminário Direito e Tecnologia**, 2023. Disponível em <https://app.openfy.co/home/>. Acesso em 22 nov. 2023.

NUNES, Dierle. ***Colegialidade corretiva, precedentes e vieses cognitivos: algumas questões do CPC-2015***. *RBDpro*, v.9, n.50, p.61-81, 2015.

O'NEIL, Cathy. **Algoritmo de destruição em massa: como o big data aumenta a desigualdade e ameaça a democracia**. Santo André: Rua do Sabão, 2020.

OPENAI. *OpenAI* **Red Teaming Network**. Disponível em <https://openai.com/blog/red-teaming-network>. Acesso em 05 fev. 2024.

REQUIÃO, Maurício; COSTA, Diego Carneiro. ***Discriminação algorítmica: ações afirmativas como estratégia de combate.*** *civilistica.com*, ano 11, n. 3, 2022. Disponível em <civilistica.com>. Acesso em 10 fev. 2023.

TABAK, Benjamin Miranda; AMARAL, Pedro Henrique Rincon. Vieses cognitivos e desenhos de políticas públicas. **Revista Brasileira De Políticas Públicas.** V. 8, n. 2, 2018. Disponível em <https://www.arqcom.uniceub.br/RBPP/article/view/5278/3979>. Acesso em 05 fev. 2024.

TAUK, Caroline Somenson; SALOMÃO, Luis Felipe. Inteligência artificial no judiciário brasileiro: estudo empírico sobre algoritmos e discriminação. **Diké (Uesc)**, v.22, n.23, 2023, p.02-32. Disponível em < https://periodicos.uesc.br/index.php/dike/article/view/3819>. Acesso em 01 fev. 2024.

WOJCIECHOWSKI, Paola Bianchi; ROSA, Alexandre Morais da. **Vieses da Justiça: como as heurísticas e vieses operam nas decisões penais e a atuação contraintuitiva.** 2.ed. Florianópolis: Emais, 2021.

Como citar:
MAURÍCIO REQUIÃO, De Sant'Ana. Inteligência artificial, vieses cognitivos e decisões judiciais. **Revista do Programa de Pós-Graduação em Direito da UFBA – Journal of the Graduate Program in Law at UFBA**, Salvador, v. 34, n.2, p. 1-18, Jul/Dez - 2024. DOI: (endereço doDOI desse artigo).

Originais recebido em: 24/09/2024.
Texto aprovado em: 05/10/2024.