Este artigo está licenciado sob a licença Creative Commons Atribuição-NãoComercial-SemDerivação 4.0 Internacional (CC BY-NC-ND 4.0).



Revista de Direito Administrativo
ISSN: 0034-8007
eISSN: 2238-5177

Governos inteligentes? Parâmetros para uma avaliação crítica do uso de sistemas de IA no setor público*

Smart governments? Parameters for a critical assessment of the use of AI systems in the public sector

*Lucas Borges de Carvalho***

RESUMO

Em um futuro próximo, os governos serão "inteligentes", no sentido de que, cada vez mais, processos decisórios no setor público serão conduzidos, direta ou indiretamente, por sistemas de inteligência artificial (IA). Essa transformação, já em curso, proporciona benefícios, mas também impactos negativos, que precisam ser identificados e mitigados. A partir de revisão bibliográfica e de estudos de caso sobre a análise automatizada de benefícios sociais e o uso de sistemas de IA no Judiciário, o artigo sustenta que a adoção de sistemas de IA no setor público demanda

* Artigo recebido em 31 de maio de 2024 e aprovado em 10 de outubro de 2024. DOI: https://doi.org/10.12660/rda.v284.2025.91301.

** Autoridade Nacional de Proteção de Dados (ANPD), Brasília, DF, Brasil. Doutor em direito pela Universidade de Brasília (UnB). Procurador federal. Assessor do Conselho Diretor da ANPD. País: Brasil. E-mail: lucasbcarvalho@gmail.com.

cautela e avaliação crítica. Nesse sentido, é preciso recusar o uso retórico de expressões como "inteligente" e "inovação", em especial quando tem por efeito a legitimação presumida e acrítica do uso desses sistemas e uma confiança cega nos resultados gerados. Assim, a implementação de sistemas de IA no setor público deve se basear em evidências, que atestem os ganhos de eficiência gerados, bem como sua compatibilidade com princípios de justiça, em particular o devido processo legal, a igualdade, a transparência e a supervisão humana eficaz.

PALAVRAS-CHAVE

Inteligência Artificial — inovação — setor público — devido processo legal — supervisão humana.

ABSTRACT

In the near future, governments will be "smart", in the sense that more and more decision-making processes in the public sector will be driven, directly or indirectly, by artificial intelligence (AI) systems. This transformation, which is already underway, provides benefits, but also negative impacts that need to be identified and mitigated. Based on a literature review and case studies on the automated analysis of social benefits and the use of AI systems in the judiciary, the article argues that the adoption of AI systems in the public sector requires caution and critical evaluation. In this sense, we must reject the rhetorical use of terms such as "intelligent" and "innovation", especially when its effect is to legitimize the presumed and uncritical use of these systems and blind trust in the results generated. Therefore, the implementation of AI systems in the public sector must be based on evidence that demonstrates the efficiency gains generated, as well as their compatibility with principles of justice, such as due process, equality, transparency, and effective human oversight.

KEYWORDS

Artificial Intelligence — innovation — public sector — due process — human oversight.

1. Introdução

Nos últimos anos, a transformação digital se intensificou no setor público, promovendo uma verdadeira revolução no modo pelo qual o Estado se relaciona com os cidadãos e na forma pela qual estes acessam serviços públicos. Em uma velocidade sem precedentes, processos analógicos foram substituídos por ferramentas digitais, em particular após a pandemia do Covid-19. Assim, por exemplo, tornou-se usual a possibilidade de apresentar um requerimento administrativo por meio de um aplicativo instalado no celular, sem a necessidade de atendimento presencial ou de autenticar documentos. Nesse e em outros casos, a inovação proporcionada pela digitalização diminui a burocracia, gerando valor e eficiência: decisões mais rápidas e processos simplificados a um custo reduzido para o cidadão e para os entes estatais.

Em um futuro próximo, esse processo de transformação digital deve passar por uma nova e ainda mais revolucionária etapa, que decorre da disseminação do uso de sistemas de inteligência artificial (IA) no setor público. Alimentados por quantidades massivas de dados, esses sistemas automatizados podem realizar previsões, fornecer recomendações ou gerar textos e imagens com o objetivo de auxiliar o processo de tomada de decisão nos mais variados contextos.[1] Dessa forma, uma solicitação administrativa — como o requerimento de um benefício social — pode ser recebida por meios digitais e decidida com o auxílio de um sistema de IA, inclusive sem qualquer intervenção humana durante todo o processo.

Diante desse cenário, o governo será não apenas *digital*, mas também *inteligente*, no sentido de que, cada vez mais, processos decisórios no setor público serão conduzidos, direta ou indiretamente, por sistemas de IA. O problema é que essa transformação, já em curso, não proporciona apenas benefícios, tais como maior padronização e celeridade nas decisões, mas também uma série de riscos — como os de opacidade e discriminação — que precisam ser avaliados e mitigados mediante a adoção de medidas

[1] Conforme a definição da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), um sistema de IA é "um sistema baseado em máquina que infere dos dados que recebe, para objetivos explícitos ou implícitos, como gerar resultados, tais como previsões, conteúdos, recomendações ou decisões que podem influenciar ambientes físicos ou virtuais. Diferentes sistemas de IA possuem distintos níveis de autonomia e adaptabilidade após a sua implantação". RUSSEL, Stuart *et al. Updates to the OECD'S definition of an AI system explained*. nov. 2023. Disponível em: https://bit.ly/3TPeipD. Acesso em: 3 jan. 2024.

apropriadas. Ademais, diante da ausência de uma legislação específica sobre o tema, é necessário lidar com esses desafios com base no arcabouço jurídico vigente, buscando em experiências no Brasil e no exterior elementos que possam orientar a implantação de sistemas de IA no setor público de forma segura e confiável.

Considerando essas questões, o argumento central a ser desenvolvido neste artigo é o de que o adjetivo "inteligente" não pode ser adotado de forma meramente retórica, isto é, apenas com o fim de enaltecer aspectos positivos das novas tecnologias, sob pena de serem ignorados e legitimados seus eventuais impactos negativos. Por isso, é necessário fundamentar e delimitar o uso da expressão "governo inteligente" — e de termos similares, como "inovação" e "eficiência" — a partir de uma reflexão crítica, que considere as evidências e os efeitos reais do uso de sistemas de IA no setor público. Nesse sentido, como será demonstrado, sistemas de IA devem ser avaliados não somente sob o aspecto dos ganhos de eficiência, mas também com base em princípios de justiça, em particular o devido processo legal, a igualdade, a transparência e a supervisão humana eficaz.

A abordagem metodológica proposta se baseia em revisão bibliográfica e em estudos de caso e análises comparativas do uso de sistemas de IA no setor público no Brasil e no exterior. Dessa forma, os parâmetros jurídicos delineados ao longo do artigo são estruturados a partir da análise do processo real de implementação desses sistemas e da sua compatibilidade com as normas jurídicas vigentes. Ao considerar os riscos e benefícios gerados na prática, essa abordagem fornece novas perspectivas para a compreensão dos impactos e dos limites ao uso de sistemas de IA no setor público, além de apontar caminhos para futuros trabalhos.

Assim, na primeira parte, é discutido de que forma escolhas de justiça e eficiência se apresentam no processo de análise automatizada de benefícios sociais, considerando, entre outros documentos, auditorias recentes realizadas por órgãos de controle sobre o tema. Na segunda parte, a discussão se centra na definição de parâmetros de transparência e supervisão humana a partir dos dilemas em torno do uso de sistemas de IA no Poder Judiciário. Ao final, a conclusão sistematiza os argumentos apresentados, propondo uma construção analítica capaz de sustentar uma avaliação crítica do uso de sistemas de IA no setor público.

2. Sistemas de IA e benefícios sociais

No início de 2021, o primeiro-ministro da Holanda, Mark Rutte, e os demais componentes de seu gabinete apresentaram um pedido de renúncia coletiva. Na ocasião, investigações comprovaram a presença de um viés discriminatório em um algoritmo utilizado pelo governo para identificar fraudes no pagamento de um determinado benefício social — uma espécie de reembolso pelos gastos familiares com creche. O sistema de IA em questão realizava uma varredura nas informações dos beneficiários, determinando, na sequência, a suspensão do benefício e a devolução de todos os valores recebidos pelo suposto fraudador.[2]

Ocorre que, entre os critérios utilizados para a identificação de fraude, o sistema considerava relevante se os pais da criança possuíam dupla nacionalidade. Por isso, as decisões automatizadas atingiram, em maior proporção, imigrantes e pessoas de minorias étnicas, obrigando-os a pagarem valores elevados ao governo em decorrência de uma suposta fraude que, em muitos casos, sequer ocorreu. As consequências foram devastadoras para muitas famílias, que se viram em dificuldades financeiras, sem o acesso à creche e sem mecanismos efetivos para contestar a cobrança indevida efetuada pelo governo.[3]

Por um lado, o caso da análise automatizada de benefícios sociais na Holanda indica que o uso de sistemas de IA pode conferir maior eficiência à atividade estatal. Entre outras possibilidades, tais sistemas podem avaliar grandes quantidades de informação em reduzido espaço de tempo, de forma padronizada e com custos reduzidos, auxiliando ou substituindo servidores públicos — no caso em questão, a fim de identificar fraudes no pagamento de benefícios sociais. Nesse contexto, a eficiência decorrente do uso de sistemas de IA no setor público se revela, especialmente, na forma do uso racional dos recursos disponíveis, no estabelecimento de processos simplificados e no proferimento de decisões mais céleres.[4]

---

[2] GEIGER, Gabriel. How a discriminatory algorithm wrongly accused thousands of families of fraud. *Motherboard*, mar. 2021. Disponível em: https://bit.ly/3NRpApC. Acesso em: 4 jan. 2024.

[3] *Ibid*.

[4] Como pondera Artur Flamínio da Silva, "as vantagens de uma utilização de sistemas de IA no dia a dia da Administração são evidentes, destacando-se, logo num primeiro momento, pela relação manifesta com o princípio da boa administração, a importância da aplicação na atividade administrativa de critérios de eficiência, economicidade e celeridade". SILVA, Artur Flamínio da. Inteligência artificial e direito administrativo. In: SILVA, A. F. (org.). *Direito administrativo e tecnologia*. 3. ed. Coimbra: Almedina, 2023. p. 15.

Por outro lado, o mesmo caso demonstra que os eventuais ganhos de eficiência não podem suplantar as exigências de justiça. Afinal, qualquer decisão emitida por uma autoridade pública, inclusive quando baseada em análises automatizadas, deve atender a parâmetros processuais e substanciais mínimos, que assegurem sua legitimidade e conformidade com a legislação vigente. O fato é que o uso de sistemas de IA no setor público deve ser eficiente *e* justo, não se podendo admitir a mera automação das injustiças, com a consequente violação dos princípios do devido processo legal, da igualdade e da transparência.

Sobre o tema, Cathy O'Neil sustenta que as Constituições democráticas tendem a favorecer a justiça, a exemplo de quando demandam o respeito à presunção de inocência, ainda que isso implique o aumento de recursos judiciais ou a demora na execução de uma determinada pena. De forma distinta, sistemas de IA favorecem a eficiência, pois a sua constituição é baseada em informações e análises que podem ser medidas e quantificadas. O problema é que a justiça muitas vezes não pode ser adequadamente calculada, como nos casos de aplicação de conceitos indeterminados e de ponderação entre princípios conflitantes. Por isso, a questão central a ser considerada no uso de sistemas de IA "[…] é se nós, enquanto sociedade, estamos dispostos a sacrificar um pouco de eficiência em prol da justiça".[5]

Dilemas similares surgiram na discussão sobre o uso de sistemas de IA para a concessão de benefícios sociais no Brasil. Com efeito, o Instituto Nacional do Seguro Social (INSS) utiliza sistemas automatizados desde meados dos anos 2000, sempre com o objetivo de conferir maior eficiência aos seus processos e decisões. Ainda assim, uma auditoria recente do Tribunal de Contas da União (TCU) identificou a incapacidade da autarquia de lidar com o quantitativo crescente de recursos administrativos em matéria previdenciária, evidenciando que, no ritmo atual, seriam necessários cerca de 16 anos apenas para o cumprimento dos acórdãos já proferidos pelo Conselho de Recursos da Previdência Social (CRPS). Por isso, o TCU recomendou ao INSS a ampliação do nível de automatização de suas atividades, com a ressalva de que "a IA deve ser usada de forma cuidadosa e transparente, garantindo a proteção dos direitos dos solicitantes e a justiça nas decisões previdenciárias".[6]

---

[5] O'NEIL, Cathy. *Algoritmos de destruição em massa*: como o big data aumenta a desigualdade e ameaça a democracia. Santo André, SP: Rua do Sabão, 2020. p. 150.

[6] BRASIL. TCU. *Acórdão nº 514/2023* — Plenário. Relator ministro Aroldo Cedraz, 22 mar. 2023. p. 21. Disponível em: https://bit.ly/3NR0ZBq. Acesso em: 6 jan. 2024.

Por sua vez, a Controladoria-Geral da União (CGU) realizou uma auditoria específica para avaliar os processos de análise e de concessão automatizadas de benefícios pelo INSS. Entre os achados, vale citar a significativa ampliação de indeferimentos nos casos em que a análise é automatizada, fator apontado como uma das causas para o aumento dos recursos interpostos ao CRPS. Confira-se:

> Enquanto o percentual de indeferimento em análises manuais caiu [de 2021 para 2022] de 53% para 50%, o percentual em análises automáticas cresceu de 41% para 65%, o que pode significar que parte dos indeferimentos que antes era realizada de maneira manual está sendo processada automaticamente […]. Não obstante, o aumento de indeferimentos automáticos teve por consequência um incremento na quantidade de recursos interpostos junto ao Conselho de Recursos da Previdência Social (CRPS).[7]

De forma similar, o TCU identificou "um elevado índice de indeferimentos indevidos por parte do INSS, o que leva a crer que a fase recursal […] sofre também em decorrência da falta de qualidade da análise no reconhecimento inicial de direitos". Como consequência, verifica-se o "aumento da judicialização da concessão de benefícios gerenciados pelo INSS", contribuindo para a ampliação dos custos para os cofres públicos.[8]

Assim, o sistema de IA estabeleceu um padrão que vem sendo chamado de "indeferimento automático" dos requerimentos administrativos.[9] Esse efeito, vale o registro, também foi identificado em decorrência do uso de sistemas automatizados para a concessão de benefícios sociais no estado de Indiana, nos Estados Unidos,[10] o que indica uma possível correlação entre automatização de análises administrativas e ampliação de decisões restritivas de direitos. Tal correlação e seus riscos devem ser objeto de análise em cada caso concreto, em especial mediante a avaliação e o monitoramento de impactos decorrentes do uso desses sistemas no setor público.

---

7 CGU. *Relatório de avaliação*: Instituto Nacional do Seguro Social, exercícios 2021 a 2023. 9 nov. 2023. p. 15-16. Disponível em: https://bit.ly/3tGCRum. Acesso em: 5 jan. 2024.

8 Brasil, *Acórdão nº 514/2023*, *op. cit.*, p. 23.

9 GERCINA, Cristiane. Análise automática de benefícios do INSS por robô falha, diz sindicato. *Folha de S.Paulo*, 4 maio 2022. Disponível em: https://bit.ly/3tIMsAU. Acesso em: 5 jan. 2024.

10 Segundo Virginia Eubanks, o sistema automatizado utilizado para a análise de benefícios sociais no estado de Indiana tornou-se "uma máquina extremamente eficiente para o indeferimento de pedidos. Ao estreitar os requisitos para a concessão de benefícios e aumentar a penalização em caso de não atendimento, o sistema alcançou reduções expressivas no quantitativo de benefícios concedidos". EUBANKS, Virginia. *Automating inequality*: how high-tech tools profile, police, and punish the poor. Nova York: Picador, 2019. p. 81-82.

Nessa linha, a auditoria da CGU destacou que, dado o aumento significativo dos recursos interpostos em face de decisões automatizadas, é necessário "avaliar os riscos associados a essa automatização, assim como medidas que possam eventualmente mitigar esses riscos".[11] A auditoria recomendou a implementação de mecanismos de controle de qualidade das decisões automatizadas e o devido acompanhamento e supervisão do processo, ressaltando a importância da definição de indicadores de avaliação e da reavaliação periódica dessas decisões com vistas a identificar eventuais concessões de benefícios com indicativos de irregularidade.[12] Por sua vez, o TCU identificou que "há deficiências graves na credibilidade dos cadastros utilizados pelo INSS, o que acaba por comprometer as decisões sobre direitos individuais dos cidadãos, especialmente quando tomadas de forma automatizada".[13]

Essas informações sobre o processo real de funcionamento de um sistema de IA demonstram que os ganhos de eficiência decorrentes da inovação e da automatização não podem ser presumidos, o que muitas vezes, efetivamente, tem ocorrido no setor público. É necessário, ao contrário, a avaliação e o monitoramento contínuos, com vistas a produzir evidências que confirmem os efeitos gerados pelo uso de sistemas de IA na redução de custos e na melhoria dos serviços prestados.

Como argumenta Neil Richards, palavras como "inovação" — e, pode-se acrescentar, expressões correlatas como "eficiência" e "inteligente" — são, muitas vezes, utilizadas de forma "seletivamente vaga", no sentido de que tendem a destacar elementos positivos do uso da tecnologia e a encobrir e a desviar a atenção sobre os seus impactos negativos.[14] Conforme destaca o autor:

> Mas aqui vai uma experiência: pegue qualquer frase de uma empresa de tecnologia sobre 'inovação' e substitua a palavra 'inovação' por 'mágica' para ver se o significado da frase muda de alguma forma. Na minha própria experiência, realizando esse jogo muitas vezes na última década, o significado quase nunca é alterado.[15]

---

11 Brasil, *Relatório de avaliação*, *op. cit.*, p. 16.
12 *Ibid.*, p. 35-36.
13 Brasil, *Acórdão nº 514/2023*, *op. cit.*, p. 22.
14 RICHARDS, Neil. *Why privacy matters*. Nova York: Oxford University Press, 2022, p. 178-179.
15 *Ibid.*, p. 177.

Em sentido similar, Virginia Eubanks menciona que, ao focar nas "qualidades inovadoras" e "disruptivas" das novas tecnologias, adotamos um "olhar míope", que mascara o fato de que "ferramentas digitais estão incrustadas em sistemas tradicionais de poder e privilégio", razão pela qual não irão "magicamente mudar a cultura, as políticas e as instituições estabelecidas ao longo dos séculos".[16]

O principal efeito pragmático dessa retórica da inovação é o de legitimar o recurso a sistemas de IA de forma acrítica, como se os ganhos de eficiência decorrentes de seu uso fossem inevitáveis ou surgissem naturalmente como uma espécie de "mágica". É fundamental, por isso, a adoção de uma postura cética em relação aos benefícios gerados por sistemas de IA, inclusive do ponto de vista da eficiência, afastando-se qualquer presunção de efeitos positivos ou confiança cega nos seus resultados sem o amparo em evidências.

Ademais, para além das questões de eficiência, é necessário avaliar os impactos das decisões automatizadas sob o ponto de vista da justiça. Ainda em relação ao INSS, a CGU identificou um aumento significativo do indeferimento do salário-maternidade pelo motivo "não afastamento do trabalho ou da atividade desempenhada", condição legal para a concessão do benefício. De acordo com o exposto no relatório da auditoria realizada pelo órgão de controle:

> Entre janeiro e julho de 2022, os indeferimentos por esse motivo representavam entre 1 e 2% dos indeferimentos totais de Salário Maternidade Urbano no mês. No entanto, a partir de setembro estes já correspondiam a 21,7%, tornando-se a principal causa de indeferimento desta espécie de benefício. Comparando-se com o ano de 2021, verifica-se que a quantidade de indeferimentos pelo não afastamento do trabalho […] subiu de 7.064 para 60.379 em 2022, dos quais 51.883 (ou 85,9%) consistiam em decisões automáticas.[17]

No entanto, identificou-se que a análise automatizada desse critério se baseava apenas na resposta dada pelo segurado a uma pergunta de teor ambíguo sobre seu afastamento do trabalho, na qual "o significado do termo afastamento pode não restar claro ao requerente, que pode erroneamente interpretá-lo como se tratando de um afastamento definitivo, e não temporário, como o legislador quis se referir […]".[18]

---

16 Eubanks, *Automating inequality*, *op. cit.*, p. 178.
17 Brasil, *Relatório de avaliação*, *op. cit.*, p. 18.
18 *Ibid.*, p. 18.

Situação similar foi verificada em relação ao benefício assistencial à pessoa com deficiência. Cerca de 120 mil indeferimentos automáticos foram efetivados durante o ano de 2022 com base, tão somente, na eventual declaração efetuada pelo requerente no Cadastro Único sobre exercício de atividade remunerada. Tal processo decisório, baseado em declaração que pode não refletir a condição real e mais atual do requerente, somente foi modificado no final de dezembro de 2022. Na ocasião, alteração regulamentar passou a exigir que essa informação não fosse considerada de forma isolada para o indeferimento do benefício, proporcionando uma drástica e imediata redução das negativas automáticas sob este motivo.

> Assim, após a publicação da Portaria Conjunta MC/MTP/INSS nº 22, no final de dezembro de 2022, em janeiro de 2023 os indeferimentos motivados por "vínculo em aberto" diminuíram para 112, sendo apenas 28 automáticos. Destaca-se, entretanto, que a Portaria Conjunta foi publicada apenas após mais de 120 mil indeferimentos pelo referido motivo, ocasionando o retorno de cidadãos à fila do INSS, a entrada com recurso junto ao CRPS ou ação judicial para concessão do benefício denegado administrativamente.[19]

Os exemplos do salário-maternidade e do benefício assistencial suscitam três questões essenciais a serem consideradas no uso de sistemas de IA no setor público. A primeira é que o uso desses sistemas tende a ampliar o ônus do cidadão de avaliar se atende às condições necessárias ao deferimento de seu pedido. Daí a importância de considerar a perspectiva das pessoas impactadas[20] e de orientá-las adequadamente sobre o funcionamento de sistemas automatizados. Afinal, uma simples resposta equivocada a um questionamento ou uma informação incompleta pode resultar no indeferimento sumário. Nessa linha, a CGU constatou que a análise automática de benefícios de aposentadoria:

> […] aumentou a responsabilidade sobre o requerente, que deve compreender com clareza as relações previdenciárias apresentadas para que possa editar ou informar precisamente novas relações omitidas no CNIS [Cadastro Nacional de Informações Sociais], sob risco de ter seu requerimento indeferido sumariamente.[21]

---

19 *Ibid.*, p. 20.
20 Eubanks, *Automating inequality*, *op. cit.*, p. 178.
21 Brasil, *Relatório de avaliação*, *op. cit.*, p. 23.

Ao contrário de seres humanos, sistemas de IA não realizam avaliações de teor ponderado, de modo que não são capazes de notar, com base em outros elementos do processo, que a resposta a uma determinada questão se deu em razão da ambiguidade de sua formulação ou diante de um equívoco manifesto do requerente. Por isso, programado para operar a partir de previsões estatísticas e de critérios predeterminados, o sistema tende a ser inflexível com a aplicação de regras e implacável com eventuais erros humanos, com o consequente indeferimento do pedido, levando, muitas vezes, à adoção de decisões excessiva ou equivocadamente restritivas de direitos.

A avaliação feita por humanos, de forma diversa, tem a vantagem de permitir a avaliação ponderada, com maior grau de discricionariedade e baseada em elementos contextuais e de razoabilidade. Tal distinção entre a avaliação humana e a análise automatizada é um dos fatores que explica, no caso do INSS, os reduzidos números de indeferimento do salário-maternidade pelo motivo de "não afastamento do trabalho" quando a análise é efetuada por servidores, o que contrasta com o quantitativo elevado de indeferimentos pelo mesmo motivo quando a análise é realizada por sistemas automatizados.[22]

A segunda questão está relacionada com os limites ao uso de sistemas de IA no setor público. Em algumas situações, ainda que à custa da eficiência de determinada política ou serviço público, pode ser necessário que uma decisão seja tomada exclusivamente por seres humanos. Em outras, pode ser necessária a revisão humana antes (ou depois) da implementação da decisão tomada por um sistema de IA. Esse tipo de avaliação depende, mais uma vez, da ponderação entre critérios de eficiência e de justiça ou, ainda, de escolhas a serem efetuadas pelos gestores públicos — e pela sociedade em geral — a partir dos riscos envolvidos em cada caso. Finalmente, em especial em face de decisões integralmente automatizadas, mecanismos adicionais de controle e de mitigação de riscos devem ser implementados, tais como ferramentas para ampliar a transparência e assegurar a possibilidade de efetiva contestação pelos impactados e de revisão da decisão em caso de erro.

A terceira e última questão se refere ao fato de que grupos sociais vulneráveis tendem a ser os mais impactados por decisões tomadas por

[22] "Ademais, chama a atenção o fato de este motivo apresentar valores médios abaixo de mil quando por meio de análises manuais, contrastando com indeferimentos automáticos superiores a 10.000, o que pode sinalizar a dificuldade de se indeferir por este motivo de maneira manual; ou seja, a dificuldade de se comprovar o não afastamento foi substituída por uma única pergunta, cujo teor deixa dúvida acerca da informação que se quer efetivamente obter." *Ibid.*, p. 19.

sistemas de IA no setor público. Nos casos analisados anteriormente, por exemplo, as decisões de indeferimento automático relativas ao salário-maternidade e ao benefício assistencial atingiram, respectiva e massivamente, mães com crianças recém-nascidas e famílias de baixa renda.

De forma geral, o impacto é maior sobre esses grupos sociais porque, além de uma maior dependência de serviços públicos, como os prestados pelo INSS, há uma maior probabilidade de seus membros serem prejudicados em razão da falta de qualidade dos dados que alimentam o sistema. Por exemplo, pessoas com alto índice de vínculos informais ao longo da vida, como os trabalhadores rurais, terão maior chance de ter seu pedido de aposentadoria não reconhecido pelo sistema do INSS.

Além disso, idosos ou pessoas com baixo índice de escolaridade podem ter maior dificuldade para compreender o funcionamento de sistemas automatizados e dos requisitos aplicáveis ao caso, os quais muitas vezes decorrem de uma legislação complexa, como a previdenciária. Por fim, pessoas desses grupos podem ter dificuldade para entender o conteúdo da decisão automatizada, quais seus direitos ou quais providências adotar no caso e os respectivos prazos, elementos que afetam negativamente as condições efetivas para contestar a determinação administrativa, inclusive em razão da insuficiência de recursos financeiros para contratar um profissional especializado.

Essas três questões — ônus excessivos aos cidadãos, limites ao uso de sistemas de IA no setor público e impactos específicos sobre grupos sociais vulneráveis — devem ser levadas em consideração em todo processo de implementação de sistemas de IA no setor público. Todas elas indicam a existência de riscos que precisam ser devidamente identificados e mitigados com a adoção das medidas apropriadas, a fim de evitar que decisões automatizadas gerem impactos negativos massivos, com a violação de direitos fundamentais dos cidadãos.

3. Sistemas de IA e Judiciário: transparência e supervisão humana eficaz

Em 2013, Eric Loomis foi acusado de envolvimento em um tiroteio, sem vítimas, na cidade de La Crosse, Wisconsin, nos Estados Unidos. Loomis negou qualquer participação no evento, mas admitiu ter dirigido um carro

roubado envolvido no crime, além de ter fugido de um policial. Antes do julgamento, o risco de o acusado cometer novos delitos foi avaliado por um sistema automatizado, conhecido pela sigla Compas. A partir de seus antecedentes e de respostas a um questionário, o sistema atribui uma nota de 1 a 10 ao réu, de forma que, quanto maior o valor atribuído, maiores os riscos de reincidência. Sem acesso à sua nota, dado o sigilo comercial do algoritmo, de propriedade de uma empresa contratada pelo Estado, Loomis soube pela decisão judicial que havia sido classificado como um indivíduo de alto risco para a comunidade, fator levado em conta para a sua condenação.[23]

Em um recurso interposto junto à Suprema Corte do estado de Wisconsin, Loomis alegou que a decisão violava o princípio do devido processo legal, em particular seu direito a uma decisão individualizada e com base em informações corretas. O tribunal negou provimento ao recurso, considerando, entre outros pontos, que o réu pôde verificar suas respostas ao questionário submetido ao sistema e que o próprio juiz do caso não teria conhecimento sobre como a nota de risco havia sido calculada. Ademais, a Corte entendeu que a recomendação automatizada não teria sido determinante para a decisão, pois esta teria se baseado em outros fatores independentes, além do que o juiz poderia discordar da recomendação com base em outras informações disponíveis.[24] Na sequência, um recurso interposto à Suprema Corte dos Estados Unidos não foi admitido, de forma que Loomis permaneceu preso, por cerca de seis anos, até ser liberado em 2019.[25]

O caso do sistema Compas suscita questões de eficiência e de justiça similares às discutidas no tópico anterior. De um lado, a automatização pode conferir maior eficiência ao sistema judicial, na medida em que viabiliza o processamento de grandes quantidades de informação de forma mais célere, com base em critérios e objetivos previamente definidos, podendo resultar, ao final, em menos processos acumulados, a um custo reduzido. Por outro lado, além de avaliados e sustentados em evidências, os eventuais ganhos de eficiência devem ser ponderados com base em princípios de justiça. Assim, assegura-se a legitimidade processual e material de decisões tomadas, direta ou indiretamente, com o auxílio de sistemas de IA.

---

[23] STATE v. LOOMIS. Comment on 881 N.W.2d 749 (Wis. 2016). *Harvard Law Review*, v. 130, n. 5, p. 1.530-1.537, mar. 2017. Disponível em: https://bit.ly/3O2Nvma. Acesso em: 9 jan. 2024.

[24] *Ibid*.

[25] CHESTERMAN, Simon. *We, the robots?* Regulating artificial intelligence and the limits of the law. Cambridge, UK: Cambridge University, 2021. p. 82.

Nessa linha, é importante avaliar a existência de eventuais ônus excessivos conferidos aos cidadãos em razão do uso de sistemas de IA. No caso mencionado, seria essencial fornecer informações e orientações práticas que garantissem ao réu o entendimento adequado sobre os critérios utilizados pelo sistema, bem como sobre as questões apresentadas no formulário e de que maneira suas respostas poderiam impactar a análise de reincidência. Essas informações ampliariam a transparência do sistema de IA, com o empoderamento dos cidadãos no relacionamento direto com o sistema automatizado, tornando, por conseguinte, o processo decisório mais justo e equitativo, em conformidade com o princípio do devido processo legal.

Outro aspecto relevante diz respeito à definição de limites ao uso de sistemas de IA no âmbito judicial. A esse respeito, deve-se considerar que algumas das funções exercidas por juízes sequer podem ser terceirizadas ou delegadas a outros seres humanos, ainda que estes sejam servidores públicos. Como ressalta Simon Chesterman, "se determinadas competências do poder público não podem ser transferidas para terceiros, pela mesma razão não devem ser transferidas para máquinas […]".[26]

Os critérios para avaliar a legitimidade de uma delegação a sistemas de IA podem ser de ordem formal (relacionados às competências decisórias) ou material (referentes ao conteúdo da decisão). Exemplo do primeiro tipo é o disposto no art. 13 da Lei nº 9.784/1999, segundo o qual é vedada a delegação de competências exclusivas dos órgãos, bem como as de editar atos normativos e de decidir recursos. No âmbito judicial, não são delegáveis competências privativas dos tribunais, conforme definidas na Constituição. Do ponto de vista material, vale citar recomendação recente da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco), segundo a qual, em regra, devem ser atribuídas apenas a seres humanos decisões com impacto irreversível ou de difícil reversão, tais como sobre a vida ou a morte de uma pessoa[27] e, pode-se acrescentar, sobre sua liberdade de ir e vir.

Ainda quanto ao caso Loomis, seria possível argumentar que, ao fornecer uma simples "recomendação" sobre o risco de reincidência do réu, o sistema automatizado teria apenas "auxiliado" o juiz em seu processo decisório. Assim, não haveria violação ao devido processo legal e aos limites de

[26] *Ibid.*, p. 110.
[27] UNESCO. *Recomendação sobre a ética da inteligência artificial*. Paris: Unesco, 2022. p. 20-22. Disponível em: https://bit.ly/3ROBnq0. Acesso em: 12 jan. 2024.

delegação de competência, uma vez que o juiz manteve a responsabilidade e o poder decisórios, inclusive o de discordar da máquina.

Esse argumento pressupõe corretamente que a atribuição de responsabilidade a um ser humano é um aspecto central para assegurar a legitimidade de decisões públicas tomadas com o auxílio de sistemas automatizados, especialmente nos casos de maior impacto a direitos fundamentais. Não obstante, a eficácia da supervisão humana não pode ser idealizada, isto é, não se pode presumir que a legitimidade decisória está naturalmente assegurada se uma autoridade puder, em tese, ignorar ou rever as recomendações fornecidas por um sistema de IA.

Isso porque, na prática, seres humanos têm uma tendência a confiar excessivamente nesses sistemas, de modo a acatar suas recomendações de forma acrítica e não reflexiva, o que pode gerar, inclusive, uma espécie de "delegação de fato" da decisão a um sistema de IA, sem as devidas salvaguardas.[28] Como explica Ben Green,

> Um conjunto de pesquisas de longa data mostra que, numa vasta gama de domínios, os sistemas automatizados de apoio à decisão tendem a alterar a tomada de decisões humanas de formas inesperadas e prejudiciais. Diversos estudos demonstram que as pessoas — incluindo especialistas — são suscetíveis ao "viés da automação", ou seja, elas agem com deferência aos sistemas automatizados, reduzindo a quantidade de juízos independentes que exibem ao tomar decisões. O viés de automação pode envolver erros omissivos — deixar de agir porque o sistema automatizado não forneceu um alerta — e erros comissivos — como seguir o conselho de um sistema automatizado, mesmo que este esteja incorreto e apresente evidências contraditórias. Além disso, a automatização de certos aspectos das ações humanas pode tornar as demais mais difíceis e causar a deterioração de determinadas habilidades. Como resultado, os sistemas automatizados podem simplesmente levar a diferentes tipos de erros, em vez de reduzir os erros globais como pretendido. A automação também pode criar um menor senso de controle, responsabilidade e agência moral entre os operadores humanos.[29]

A título de exemplo, vale mencionar que empresas de tecnologia têm apostado no desenvolvimento de carros integralmente autônomos em

---

[28] CALO, Ryan; CITRON, Danielle. The automated administrative state: a crisis of legitimacy. *Emory Law Journal*, v. 70, n. 4, 2021. p. 816. Disponível em: https://bit.ly/4bDTEin. Acesso em: 30 jan. 2024.

[29] GREEN, Ben. The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review*, v. 45, p. 1-22, 2022. p. 7. Disponível em: https://bit.ly/3R8qBeN. Acesso em: 30 abr. 2024.

detrimento de modelos que demandam a presença de um usuário responsável por assumir o controle do carro em situações de perigo. A decisão se baseia em experimentos segundo os quais seres humanos não desempenham bem essa função, pois, além de estarem sujeitos ao viés da automação, possuem um tempo de reação que supera o necessário para emergências ou têm dificuldade para prever os erros produzidos pelos sistemas de IA, já que esses costumam ser distintos daqueles produzidos por humanos.[30]

Considere-se, ainda, que sistemas de IA — sejam carros autônomos ou sistemas de análise de reincidência — são envoltos por uma áurea de objetividade e imparcialidade técnicas, fator que lhes confere legitimidade aos olhos da pessoa responsável por sua supervisão, contribuindo, ademais, para dificultar a compreensão de seu funcionamento e limitar eventuais discordâncias. No caso Loomis, como visto, o juiz do caso não tinha informações adequadas sobre o funcionamento e os critérios utilizados pelo sistema para realizar a análise de risco de reincidência, o que, por si só, constitui um fator que limita a eficácia de seu papel supervisor.

A esse respeito, Miriam Wimmer e Danilo Doneda apontam as principais limitações que se apresentam à efetiva supervisão humana de sistemas de IA:

> Ainda quanto às formas de intervenção humana, um segundo ponto a ser considerado diz respeito à efetiva capacidade humana (intelectual, emocional, motora) de (re)avaliação substantiva de decisões sugeridas por sistemas de IA. Em muitos casos, a própria opacidade do processo decisório de sistemas de IA dificulta a identificação de erros que possam ter sido cometidos. Em outras circunstâncias, o volume de dados tratados por meio de decisões automatizadas é tão elevado que não há condições para que uma avaliação caso a caso seja sequer realizável por um humano. Por fim, é preciso ainda considerar a amplamente difundida visão de que as decisões automatizadas são sempre mais precisas e mais corretas do que as decisões humanas, o que acaba gerando, para o decisor humano, um ônus argumentativo por vezes excessivo ou mesmo intransponível para não adotar a solução sugerida pelo sistema automatizado.[31]

---

[30] DAVE, Paresh. Google ditched autopilot driving feature after test user napped behind wheel. *Reuters*, out. 2017. Disponível em: https://reut.rs/3PbgdPu. Acesso em: 13 jan. 2024; WAGNER, Ben. Liable, but not in control? Ensuring meaningful human agency in automated decision-making systems. *Policy & Internet*, v. 11, p. 104-122, 2019, p. 109-110. Disponível em: https://bit.ly/3X3YIIC. Acesso em: 30 abr. 2024.

[31] WIMMER, Miriam; DONEDA, Danilo. 'Falhas de IA' e a intervenção humana em decisões automatizadas: parâmetros para a legitimação pela humanização. *RDP*, Brasília, v. 18, n. 100, p. 374-406, out./dez. 2021. p. 394. Disponível em: https://bit.ly/3O2M7Qk. Acesso em: 14 jan. 2024.

Em suma, apesar de ser uma medida relevante de proteção a direitos e de legitimação de decisões automatizadas, a supervisão humana de sistemas de IA possui limitações. De fato, "apenas unir humanos e máquinas é ingênuo e simplista", sendo necessária uma "análise mais ampla sobre as funções que os humanos desempenham e deveriam desempenhar no processo".[32] Assim, a fim de evitar a criação de uma falsa sensação de segurança em relação a esses sistemas, a supervisão humana deve ser encarada sob um ponto de vista pragmático, de modo que sua implementação seja acompanhada de medidas que assegurem sua eficácia.

Entre outras medidas, inclusive visando a afastar os riscos decorrentes do viés da automação e a ocorrência de "delegações de fato" a máquinas sem as devidas salvaguardas, deve-se assegurar que as pessoas responsáveis compreendam as capacidades e limitações do sistema e sejam capazes de interpretar suas decisões de forma correta e de efetivamente revê-las ou não as utilizar quando necessário.[33] Se essas garantias não estiverem presentes e não houver evidência de que a supervisão humana é eficaz, o mais adequado é que o sistema de IA não seja utilizado.[34]

O último aspecto a ser analisado sobre o caso Loomis diz respeito ao impacto de sistemas de IA sobre grupos vulneráveis. Por um lado, esse impacto decorre da opacidade dos critérios para a avaliação de risco de reincidência, o que dificulta o escrutínio público sobre o sistema e, no plano individual, amplia os obstáculos para a compreensão dos fundamentos decisórios e para sua efetiva contestação. Por outro lado, um estudo sobre o sistema Compas identificou um viés de discriminação racial nas análises realizadas, isto é, uma maior tendência de classificar réus brancos na faixa de baixo risco e uma taxa duas vezes maior de classificação equivocada de negros na faixa de alto risco de reincidência.[35]

---

[32] SOLOVE, Daniel; MATSUMI, Hideyuki. AI, Algorithms, and Awful Humans. *Fordham Law Review*, v. 92, p. 1.923-1.940, 2024. p. 1938. Disponível em: https://bit.ly/3TVkAEq. Acesso em: 16 jan. 2024.

[33] Nesse sentido é o disposto no art. 14 do Regulamento de Inteligência Artificial ("AI Act"), proposta legislativa recentemente aprovada na União Europeia. UNIÃO EUROPEIA. *Regulamento de Inteligência Artificial*. Bruxelas, 2024. Disponível em: https://bit.ly/3Vl13gT. Acesso em: 30 maio 2024. Redação similar foi adotada no art. 20, parágrafo único, do Projeto de Lei nº 2.338, de 2023, em tramitação no Senado Federal. BRASIL. Senado Federal. *Projeto de Lei nº 2338, de 2023*. Dispõe sobre o uso da Inteligência Artificial. Disponível em: https://bit.ly/3RVNMsf. Acesso em: 13 jan. 2024.

[34] Green, The flaws of policies requiring human oversight of government algorithms, *op. cit*., p. 3.

[35] ANGWIN, Julia. *et al*. Machine bias: there's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*, 2016. Disponível em: https://bit.ly/3JkEEK5. Acesso em: 14 jan. 2024.

Esse viés discriminatório reforça a tendência de impacto negativo sobre grupos sociais vulneráveis, pondo em xeque a legitimidade do sistema automatizado sob o ponto de vista de sua compatibilidade com princípios de justiça, como o do devido processo legal e o da igualdade. As consequências geradas são ainda mais graves por se tratar de um contexto de alto risco — o do sistema prisional — e por envolver processos decisórios em larga escala. Por isso, é imprescindível que sistemas de IA utilizados no setor público sejam objeto de avaliação e monitoramento contínuos, publicizando-se as informações relevantes, inclusive os critérios decisórios, de modo a assegurar o controle social sobre seu uso e evitar que as decisões proferidas reproduzam, ainda que de forma indireta, relações de poder e opressão já existentes na sociedade.[36]

No Brasil, parte dessas questões foi endereçada pela Resolução nº 332, de 21 de agosto de 2020, do Conselho Nacional de Justiça (CNJ), que dispõe sobre o uso de sistemas de IA no Judiciário. Entre outros pontos, a norma prevê que "decisões judiciais apoiadas em ferramentas de Inteligência Artificial devem preservar a igualdade, a não discriminação, a pluralidade e a solidariedade, auxiliando no julgamento justo […]" (art. 7º).[37] Além disso, expressa preocupação com o impacto sobre grupos vulneráveis, ao exigir a prévia homologação de sistemas de IA, "de forma a identificar se preconceitos ou generalizações influenciaram seu desenvolvimento, acarretando tendências discriminatórias no seu funcionamento" (Art. 7º, § 1º).[38]

A Resolução nº 332/2020 estabelece, ainda, regras específicas sobre: (i) transparência, incluindo a obrigação de fornecer "explicação satisfatória" quanto às decisões de sistemas de IA (art. 8º, VI) e a de informar ao usuário externo sobre a "utilização de sistema inteligente nos serviços que lhe forem prestados" (art. 18); e (ii) supervisão humana, garantindo a autonomia dos servidores que utilizam sistemas de IA, mediante a vedação de "qualquer

---

[36] Como ressalta Tarcízio Silva, "os sistemas algorítmicos tomam decisões por nós e sobre nós com frequência cada vez maior. […] Estas decisões trazem impactos em diferentes níveis de imediaticidade e sutileza, podendo modular o comportamento e condutas de seus usuários de forma discreta, na maioria dos casos para reproduzir relações de poder e opressão já existentes na sociedade". SILVA, Tarcízio. Racismo algorítmico em plataformas digitais: microagressões e discriminação em código. In: SILVA, Tarcízio (org.). *Comunidades, algoritmos e ativismos digitais*: olhares afrodiaspóricos. São Paulo: Literarua, 2020. p. 123.

[37] BRASIL. CNJ. *Resolução nº 332, de 21 de agosto de 2020*. Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências. Disponível em: https://bit.ly/3X3oJYl. Acesso em: 30 maio 2024.

[38] *Ibid*.

espécie de vinculação à solução apresentada pela Inteligência Artificial" (art. 17, II).[39]

O regramento do CNJ é relevante por ser uma norma pioneira, que estabelece, de forma centralizada e coordenada, parâmetros para orientar e avaliar o uso eficiente e justo de sistemas de IA no setor público — no caso, no Poder Judiciário — brasileiro. Tal iniciativa pode, inclusive, servir como modelo para regramentos similares, especialmente no âmbito do Poder Executivo. Não obstante, é importante aprofundar a análise e verificar de que forma os parâmetros estabelecidos pelo CNJ estão sendo observados e aplicados na prática pelos tribunais. Nessa linha, pesquisa recente identificou cerca de 64 iniciativas de uso de sistemas de IA por tribunais em todo o país. Entre outras funcionalidades, esses sistemas são utilizados para triagem e classificação de processos, a exemplo da identificação de temas repetitivos ou de repercussão geral; ou, ainda, para elaborar propostas de decisão.[40]

No que concerne às obrigações de transparência previstas na Resolução nº 332/2020, a pesquisa identificou diversos obstáculos para obter informações sobre os sistemas de IA utilizados pelos tribunais. Isso porque o mais comum é "a opção por notícias esparsas", de modo que são divulgados apenas "os objetivos e as aplicações da ferramenta, sem uma consolidação dos dados relacionados aos resultados alcançados e, tampouco, da evolução desses ao longo do tempo".[41] Além disso, "para nenhuma das ferramentas de IA foi identificada a documentação e indicação dos instrumentos de segurança da informação e controle", e "em nenhuma aplicação das ferramentas há o aviso explícito aos usuários externos sobre o uso da IA para o caso que está sendo pesquisado".[42]

A incompletude e a fragmentação das informações disponibilizadas e a ausência de notificação aos usuários externos sobre o uso dos sistemas são fatores que afetam a legitimidade decisória, na medida em que violam o princípio constitucional da publicidade e criam obstáculos injustificados para o controle democrático das decisões proferidas com base em algoritmos.[43]

---

39 *Ibid.*

40 SALOMÃO, Luis; TAUK, Caroline. *Inteligência artificial*: tecnologia aplicada à gestão de conflitos no âmbito do Poder Judiciário brasileiro. 3. ed. Rio de Janeiro: FGV, 2023. p. 76. Disponível em: https://bit.ly/47BmpJJ. Acesso em: 14 jan. 2024.

41 *Ibid.*, p. 79.

42 *Ibid.*, p. 80 e 84.

43 ARAÚJO, Valter; ZULLO, Bruno; TORRES, Maurílio. *Big data*, algoritmos e inteligência artificial na Administração Pública: reflexões para a sua utilização em um ambiente democrático. *Revista de Direito Administrativo & Constitucional*, Belo Horizonte, n. 80, p. 241-261, abr./jun. 2020. p. 254. Disponível em: https://bit.ly/3TfBPip. Acesso em: 3 set. 2024.

Mais especificamente, a falta de transparência pode ampliar: (i) a assimetria de informação entre o poder público e os cidadãos afetados pelo uso desses sistemas; e (ii) o grau de incerteza sobre os potenciais danos e sobre a adequação das medidas de precaução adotadas.[44]

Por isso, é essencial estabelecer padrões a serem observados pelos diversos órgãos, em particular regras objetivas sobre quais informações e documentos devem ser efetivamente disponibilizados. Outra possibilidade é a divulgação centralizada, como uma espécie de "registro público" de iniciativas de sistemas de IA, o qual deve ser objeto de constante atualização, a exemplo do registro público de algoritmos da cidade de Amsterdã, sítio eletrônico no qual são divulgadas informações detalhadas e em formato padronizado sobre os sistemas de IA utilizados pelo governo municipal.[45] Iniciativas como essas podem ampliar o conhecimento e a confiança sobre os sistemas de IA utilizados no setor público e permitir a realização de análises comparativas.

Por fim, no que tange à supervisão humana, a Resolução nº 332/2020 se limita a vedar qualquer tipo de vinculação às decisões apresentadas por sistemas de IA. Especificamente sobre a análise automatizada de reincidência, a norma prevê que os sistemas "não devem indicar conclusão mais prejudicial ao réu do que aquela a que o magistrado chegaria sem a sua utilização" (art. 23, § 2º).[46] Embora sinalizem corretamente que a decisão final deve permanecer sob controle humano, essas diretrizes, isoladamente, são de difícil aplicação prática e podem criar uma falsa sensação de segurança e de legitimidade do uso desses sistemas. Por isso, como mencionado, é fundamental estabelecer regras mais específicas, que assegurem a eficácia do controle humano.

Ademais, "é preciso monitoramento rigoroso na implementação" desses sistemas, notadamente quando "desempenham tarefas de minutar decisões".[47] De fato, a automatização de decisões judiciais sem as devidas salvaguardas, notadamente a supervisão humana eficaz, pode implicar uma indevida transferência de responsabilidades para máquinas e o proferimento

---

[44] Nesse sentido, ver BIONI, Bruno; LUCIANO, Maria. O princípio da precaução na regulação da inteligência artificial: seriam as leis de proteção de dados o seu portal de entrada? In: FRAZÃO, Ana; MULHOLLAND, Caitlin (org.). *Inteligência artificial e direito*: ética, regulação e responsabilidade. 2. ed. São Paulo: Revista dos Tribunais, 2020. p. 210-214.

[45] HOLANDA. Cidade de Amsterdã. *Algorithm register*. 2024. Disponível em: https://algoritmeregister.amsterdam.nl/. Acesso em: 26 jan. 2024.

[46] Brasil, *Resolução nº 332, de 21 de agosto de 2020*, *op. cit.*

[47] Salomão e Tauk, *Inteligência artificial*, *op. cit.*, p. 85.

de decisões injustas e discriminatórias. Isso porque, muitas vezes, sistemas automatizados não levam em consideração as particularidades do caso concreto e não detêm as habilidades necessárias para interpretar princípios e conceitos jurídicos indeterminados. Nessa linha, Bárbara Magalhães destaca que há "o risco de cristalização da jurisprudência", uma vez que, ao se vincular estritamente a decisões anteriores, tais sistemas podem "conduzir a um tratamento igualitário de questões substancialmente diferentes" ou, ainda, "impedir o surgimento de decisões inovadoras".[48]

Por tais razões, como regra geral, sistemas de IA que produzem decisões de autoridades judiciais ou auxiliam sua elaboração devem ser considerados de alto risco. Em tais contextos, como o de análise de reincidência, é insuficiente a determinação vaga e abstrata de ausência de vinculação à decisão do sistema ou de que este não deve propor conclusão mais prejudicial da que um ser humano estabeleceria sem seu auxílio. Como visto no caso Loomis, é preciso ir além, exigindo-se a avaliação dos efetivos ganhos de eficiência gerados e da compatibilidade do sistema de IA com princípios de justiça, bem como a identificação de riscos e de medidas de mitigação apropriadas, em especial diante da possibilidade de: (i) viés discriminatório; e (ii) reprodução acrítica das decisões do sistema pelas pessoas responsáveis por sua supervisão.

4. Conclusão

Nos próximos anos, a transformação digital do setor público será ainda mais intensa e revolucionária do que o verificado com a recente digitalização de procedimentos e serviços. Os governos serão, cada vez mais, "inteligentes", haja vista a ampliação do uso de sistemas de IA nos processos decisórios conduzidos por autoridades públicas. Ao longo deste trabalho, demonstrou-se que, embora este seja um passo desejável e inevitável, a adoção de sistemas de IA no setor público demanda cautela e avaliação crítica, em particular em um contexto de ausência de legislação específica sobre o tema.

---

[48] MAGALHÃES, Bárbara. Desafios da inteligência artificial nas garantias do direito e processo administrativo. In: OLIVEIRA, Sofia; JERÓNIMO, Patrícia (org.). *Liber Amicorum Benedita Mac Crorie*, vol. I. Portugal: UMinho, 2022. p. 259-260. Disponível em: https://bit.ly/4cRFMB3. Acesso em: 3 set. 2024.

Nesse sentido, é preciso recusar o uso retórico de expressões como "inteligente" e "inovação", especialmente quando tem por efeito desviar a atenção sobre impactos negativos, com a consequente legitimação presumida e acrítica do uso desses sistemas e uma confiança cega nos resultados gerados. Para além da retórica, é necessário que a implementação desses sistemas se fundamente em evidências, que atestem, por um lado, os efetivos ganhos de eficiência gerados e, por outro, sua compatibilidade com princípios de justiça, notadamente os do devido processo legal, da igualdade, da transparência e da supervisão humana eficaz.

Com base nos estudos de caso realizados — análise automatizada de benefícios sociais e uso de sistemas de IA no Judiciário —, foi possível demonstrar, ainda, que três questões são essenciais com vistas à adequada avaliação de sistemas automatizados no setor público: ônus excessivos gerados aos cidadãos, limites ao uso de sistemas de IA e impactos específicos sobre grupos sociais vulneráveis. Assim, é fundamental: (i) empoderar o cidadão no relacionamento com sistemas de IA, mediante o fornecimento de informações e orientações práticas sobre seu uso, funcionamento e critérios decisórios adotados; (ii) garantir que a supervisão humana seja eficaz, afastando-se a criação de uma falsa sensação de segurança ou, ainda, a realização de delegações decisórias "de fato" a máquinas sem as devidas salvaguardas; e (iii) avaliar e monitorar de forma contínua o uso de sistemas de IA, a fim de evitar a tomada de decisões injustas, em especial diante de potencial risco discriminatório.

Em suma, os parâmetros apresentados integram uma construção analítica que permite avaliar criticamente o uso de sistemas de IA no setor público, identificando riscos e medidas de mitigação apropriadas a cada caso. É certo que os governos podem e devem ser "inteligentes", "inovadores" e "eficientes". Porém, em nenhuma hipótese devemos aceitar que os encantos ou as promessas das novas tecnologias nos afastem de nosso compromisso com a tradição jurídica democrática, incluindo o respeito incondicional aos direitos e garantias fundamentais.

Referências

ANGWIN, Julia *et al*. Machine bias: there's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*, 2016. Disponível em: https://bit.ly/3JkEEK5. Acesso em: 14 jan. 2024.

ARAÚJO, Valter; ZULLO, Bruno; TORRES, Maurílio. *Big data*, algoritmos e inteligência artificial na administração pública: reflexões para a sua utilização em um ambiente democrático. *Revista de Direito Administrativo & Constitucional*, Belo Horizonte, n. 80, p. 241-261, abr./jun. 2020. Disponível em: https://bit.ly/3TfBPip. Acesso em: 3 set. 2024.

BIONI, Bruno; LUCIANO, Maria. O princípio da precaução na regulação da inteligência artificial: seriam as leis de proteção de dados o seu portal de entrada? In: FRAZÃO, Ana; MULHOLLAND, Caitlin (org.). *Inteligência artificial e direito*: ética, regulação e responsabilidade. 2. ed. São Paulo: Revista dos Tribunais, 2020. p. 205-229.

BRASIL. CGU. *Relatório de avaliação*: Instituto Nacional do Seguro Social, exercícios 2021 a 2023. 9 nov. 2023. Disponível em: https://bit.ly/3tGCRum. Acesso em: 5 jan. 2024.

BRASIL. CNJ. *Resolução nº 332, de 21 de agosto de 2020*. Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências. Disponível em: https://bit.ly/3X3oJYl. Acesso em: 30 maio 2024.

BRASIL. Senado Federal. *Projeto de Lei nº 2338, de 2023*. Dispõe sobre o uso da Inteligência Artificial. Disponível em: https://bit.ly/3RVNMsf. Acesso em: 13 jan. 2024.

BRASIL. TCU. *Acórdão nº 514/2023* — Plenário. Relator ministro Aroldo Cedraz, 22 mar. 2023. Disponível em: https://bit.ly/3NR0ZBq. Acesso em: 6 jan. 2024.

CALO, Ryan; CITRON, Danielle. The automated administrative state: a crisis of legitimacy. *Emory Law Journal*, v. 70, n. 4, p. 797-845, 2021. Disponível em: https://bit.ly/4bDTEin. Acesso em: 30 jan. 2024.

CHESTERMAN, Simon. *We, the robots?* Regulating artificial intelligence and the limits of the law. Cambridge, UK: Cambridge University, 2021.

DAVE, Paresh. Google ditched autopilot driving feature after test user napped behind wheel. *Reuters*, out. 2017. Disponível em: https://reut.rs/3PbgdPu. Acesso em: 13 jan. 2024.

EUBANKS, Virginia. *Automating inequality*: how high-tech tools profile, police, and punish the poor. Nova York: Picador, 2019.

GEIGER, Gabriel. How a discriminatory algorithm wrongly accused thousands of families of fraud. *Motherboard*, mar. 2021. Disponível em: https://bit.ly/3NRpApC. Acesso em: 4 jan. 2024.

GERCINA, Cristiane. Análise automática de benefícios do INSS por robô falha, diz sindicato. *Folha de S.Paulo*, 4 maio 2022. Disponível em: https://bit.ly/3tIMsAU. Acesso em: 5 jan. 2024.

GREEN, Ben. The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review*, v. 45, p. 1-22, 2022. Disponível em: https://bit.ly/3R8qBeN. Acesso em: 30 abr. 2024.

HOLANDA. Cidade de Amsterdã. *Algorithm register*. 2024. Disponível em: https://algoritmeregister.amsterdam.nl/. Acesso em: 26 jan. 2024.

MAGALHÃES, Bárbara. Desafios da inteligência artificial nas garantias do direito e processo administrativo. In: OLIVEIRA, Sofia; JERÓNIMO, Patrícia (org.). *Liber Amicorum Benedita Mac Crorie*, vol. I. Portugal: UMinho, 2022. p. 243-260. Disponível em: https://bit.ly/4cRFMB3. Acesso em: 3 set. 2024.

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

RICHARDS, Neil. *Why privacy matters*. Nova York: Oxford University Press, 2022.

RUSSEL, Stuart; PERSET, Karine; GROBELNIK, Marko. *Updates to the OECD'S definition of an AI system explained*. nov. 2023. Disponível em: https://bit.ly/3TPeipD. Acesso em: 3 jan. 2024.

SALOMÃO, Luis; TAUK, Caroline. *Inteligência artificial*: tecnologia aplicada à gestão de conflitos no âmbito do Poder Judiciário brasileiro. 3. ed. Rio de Janeiro: FGV, 2023. Disponível em: https://bit.ly/47BmpJJ. Acesso em: 14 jan. 2024.

SILVA, Artur Flamínio da. Inteligência artificial e direito administrativo. *In*: SILVA, A. F. (org.). *Direito administrativo e tecnologia*. 3. ed. Coimbra: Almedina, 2023. p. 11-32.

SILVA, Tarcízio. Racismo algorítmico em plataformas digitais: microagressões e discriminação em código. In: SILVA, Tarcízio (org.). *Comunidades, algoritmos e ativismos digitais*: olhares afrodiaspóricos. São Paulo: Literarua, 2020. p. 120-137.

SOLOVE, Daniel; MATSUMI, Hideyuki. AI, algorithms, and awful humans. *Fordham Law Review*, v. 92, p. 1.923-1.940, 2024. Disponível em: https://bit.ly/3TVkAEq. Acesso em: 16 jan. 2024.

STATE v. LOOMIS. Comment on 881 N.W.2d 749 (Wis. 2016). *Harvard Law Review*, v. 130, n. 5, p. 1.530-1.537, mar. 2017. Disponível em: https://bit.ly/3O2Nvma. Acesso em: 9 jan. 2024.

UNESCO. *Recomendação sobre a ética da inteligência artificial*. Paris: Unesco, 2022. Disponível em: https://bit.ly/3ROBnq0. Acesso em: 12 jan. 2024.

UNIÃO EUROPEIA. *Regulamento de Inteligência Artificial*. Bruxelas, 2024. Disponível em: https://bit.ly/3Vl13gT. Acesso em: 30 maio 2024.

WAGNER, Ben. Liable, but not in control? Ensuring meaningful human agency in automated decision-making systems. *Policy & Internet*, v. 11, p. 104-122, 2019. Disponível em: https://bit.ly/3X3YIIC. Acesso em: 30 abr. 2024.

WIMMER, Miriam; DONEDA, Danilo. 'Falhas de IA' e a intervenção humana em decisões automatizadas: parâmetros para a legitimação pela humanização. *RDP*, Brasília, v. 18, n. 100, p. 374-406, out./dez. 2021. Disponível em: https://bit.ly/3O2M7Qk. Acesso em: 14 jan. 2024.