



LISIANE PENZ NEGRUNI

**ENTRE A OPACIDADE E A EXPLICABILIDADE: NOVOS CONTORNOS
JURÍDICOS PARA A MOTIVAÇÃO DAS DECISÕES ADMINISTRATIVAS
ALGORÍTMICAS**

Dissertação com vista à obtenção do grau
de Mestre em Direito na especialidade de
Direito e Tecnologia.

Orientador:

Doutor Francisco Pereira Coutinho

Setembro 2022

UNIVERSIDADE NOVA DE LISBOA

LISIANE PENZ NEGRUNI

**ENTRE A OPACIDADE E A EXPLICABILIDADE: NOVOS CONTORNOS
JURÍDICOS PARA A MOTIVAÇÃO DAS DECISÕES ADMINISTRATIVAS
ALGORÍTMICAS**

Dissertação com vista à obtenção do grau
de Mestre em Direito na especialidade de
Direito e Tecnologia.

Orientador:

Doutor Francisco Pereira Coutinho

Lisboa,

Setembro 2022

Declaração Antiplágio

Declaro por minha honra que o trabalho que apresento é original e que todas as minhas citações estão corretamente identificadas. Tenho consciência de que a utilização de elementos alheios não identificados constitui uma grave falta ética e disciplinar.

Dedicatória

Ao meu marido Roberto, que me deu asas para voar, vencer meus medos, cruzar o oceano e retornar ao nosso lar.

À minha filha Stella, por quem dou tudo.

Dedico a vocês essa obra com todo o meu amor.

Agradecimentos

Agradeço aos meus pais, que, mesmo nos momentos de dificuldade, nunca mediram esforços para que eu tivesse acesso aos livros. À minha mãe, que me acompanhou em Portugal durante a parte letiva do mestrado e tornou possível a realização desse sonho. Ao meu pai, que sempre incentivou meus estudos e, em especial, a conclusão desta dissertação. Agradeço ao meu grande amigo Tiago, que muito contribuiu nos debates sobre o tema, além de ter sido um zeloso revisor deste trabalho.

Agradeço ao meu orientador, Professor Francisco, por ser tão acessível, dedicado e amigo. Foi uma honra ser sua aluna e orientanda.

Agradeço à Procuradoria-Geral do Estado do Rio Grande do Sul, que vem oportunizando a realização das minhas aspirações rumo à construção de um setor público inovador, desburocratizado e conectado às necessidades do cidadão.

Breves Notas

Em relação ao modo de citar, a referência neste trabalho segue o padrão OSCOLA (*Oxford University Standard for the Citation of Legal Authorities*).

Quanto ao número de caracteres, declara-se que o corpo da dissertação, incluindo espaços e notas, ocupa um total de 152.821 caracteres.

A presente dissertação foi redigida em língua portuguesa na variante brasileira.

Lista de Abreviaturas

CNJ	Conselho Nacional de Justiça
COMPAS	Correctional Offender Management Profiling for Alternative Sanctions
EBIA	Estratégia Brasileira de Inteligência Artificial
IA	Inteligência Artificial
LGPD	Lei Geral de Proteção de Dados Pessoais brasileira
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
RGPD	Regulamento Geral de Proteção de Dados Pessoais europeu

Resumo

O trabalho tem por objetivo estabelecer novos contornos para o dever de motivação dos atos administrativos individuais, considerando o cenário de digitalização da Administração Pública e a utilização de algoritmos para a produção de decisões administrativas. A partir do conteúdo jurídico do dever de motivação, enquanto corolário do princípio da legalidade, sustenta-se que sua observância, num contexto de decisões administrativas algorítmicas, consiste em garantir que o administrado seja capaz de compreender e desafiar o processo de tomada de decisão, ou, dito de outra forma, contra-argumentar o algoritmo. Demonstra-se que, tecnicamente, há meios de assegurar maior nível de explicabilidade aos resultados fornecidos pelos sistemas algorítmicos e que, juridicamente, um direito à explicação decorre de uma interpretação holística da LGPD. Conclui-se que, para fazer uso de decisões administrativas algorítmicas, a Administração Pública deverá resguardar a explicabilidade como requisito da motivação, de forma a determinar que a escolha do algoritmo observe padrões exigentes de acessibilidade e compreensibilidade, o que impõe um olhar cuidadoso sobre as escolhas humanas feitas em todas as fases de concepção e implementação do modelo (*explicability by design*). Recomenda-se a regulação do tema no ordenamento jurídico brasileiro, em especial diante da incerteza normativa em relação aos tipos de decisões administrativas que podem ser automatizadas, do grau de automatização apropriado em cada caso e da falta de clareza sobre a necessidade de intervenção humana no processo de revisão dessas decisões.

Palavras-chave: Administração Pública – opacidade – explicabilidade – decisões algorítmicas – motivação – direito à explicação.

Abstract

The article aims to establish the new contours for the duty of motivation of individual administrative acts, considering the scenario of digitalization of the Public Administration and the use of algorithms to produce administrative decisions. From the legal content of the duty of motivation, as a corollary of the principle of legality, it is argued that its observance, in a context of algorithmic administrative decisions, consists in ensuring that the citizen is able to understand and challenge the decision-making process, or, in other words, counter-argument the algorithm. It is shown that, technically, there are means to ensure a higher level of explicability to the results provided by algorithmic systems and that, legally, a right to explanation stems from a global interpretation of the LGPD. It is concluded that, in order to make use of algorithmic administrative decisions, the Public Administration should

safeguard explainability as a requirement of the motivation, in order to determine that the choice of algorithm observes demanding standards of accessibility and interpretability, which imposes a careful look at the human choices made at all stages of design and implementation of the model (explainability by design). It is recommended the regulation of the topic in the Brazilian legal system, especially in face of the normative uncertainty regarding the types of administrative decisions that can be automated, the degree of automation appropriate in each case and the lack of clarity on the need for human intervention in the review process of these decisions.

Keywords: Public administration – opacity – explainability – algorithmic decision-making – right to explanation.

Introdução

Ao cogitar o uso de algoritmos como apoio ou para a própria produção de decisões administrativas, há uma série de desafios a enfrentar rumo ao desenvolvimento do devido processo legal tecnológico. No âmbito do Direito Administrativo brasileiro, um dos cânones que convém ser revisitado é o dever de motivação dos atos administrativos.

Concebido com a finalidade substancial de favorecer o controle dos atos diante da inerente falibilidade humana, o dever de motivação é uma garantia essencial dos administrados, cujo escopo de incidência se ampliou com a evolução do Estado de Direito. A seu turno, as decisões algorítmicas apresentam o potencial de extirpar parte dos erros intrínsecos ao julgamento humano, contribuindo com a aplicação isonômica das regras jurídicas e com a concretização do princípio da eficiência da Administração Pública.

Ocorre que, por outro lado, o uso de algoritmos nas decisões dá origem a novas formas de falibilidade para as quais urge atentar.¹ Por trás do manto da objetividade matemática, há uma série de escolhas humanas embutidas no processo de automatização de decisões que também podem gerar imperfeições e culminar, por exemplo, em viés discriminatório. O enviesamento causa profundas preocupações aos juristas e tem sido objeto de muitos escritos. Ao lado dessa inquietação andam também a opacidade algorítmica e a crescente atenção em assegurar que se possa compreender – ou explicar – o resultado produzido por meio do processamento operado por algoritmos.

Com efeito, essa preocupação perpassa a esfera jurídica e exige que se penetre, ainda que com inevitável generalização, em aspectos computacionais e estatísticos. No âmbito técnico, é preciso entender, basicamente, como os algoritmos operam, como se origina o problema da opacidade algorítmica e se há meios para mitigá-la.

No âmbito jurídico, o cuidado em assegurar algum nível de compreensibilidade ao aspecto algorítmico das decisões individualizadas automatizadas foi alvo tanto do Regulamento Geral de Proteção de Dados Pessoais europeu,² quanto da legislação brasileira sobre a matéria nele inspirada.³ Nessa linha, compreender o escopo da explicabilidade que deriva dessa nova legislação trará luz para a discussão que ora se apresenta, em especial, por duas razões. A primeira é que as disposições da LGPD incidem sobre o tratamento de dados

¹ O conceito de algoritmo será detalhado no item 2.1 deste trabalho.

² Regulamento (EU) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016 relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (RGPD) (2016) OJ L119/1.

³ Lei brasileira nº 13.709 de 14 de agosto de 2018 (LGPD) (2018).

personais realizado pela Administração Pública, o que acontece na grande maioria dos casos, senão em todos, em que há um ato administrativo individual. A segunda é que a interpretação conferida pela doutrina a respeito do direito à explicação que decorre da proteção de dados pessoais pode iluminar a compreensão do grau de explicabilidade que deve ser conferido às decisões administrativas algorítmicas, em cotejo com o dever de motivação.

Além disso, quando se trata do uso de inteligência artificial, diversos documentos internacionais, e também a própria Estratégia Brasileira de Inteligência Artificial por eles guiada, recomendam medidas destinadas ao chamado desenvolvimento ético, que incluem a transparência e a explicabilidade dos resultados produzidos.⁴ Embora inexistam disposições normativas específicas a respeito do ecossistema que deve ser assegurado para a implementação de decisões algorítmicas em âmbito administrativo no direito brasileiro, encontra-se regulamentação aplicável ao Poder Judiciário que é capaz de alvejar os caminhos que se deve seguir também a nível administrativo.

Nesse contexto, o tema central deste trabalho é compreender os pontos de atrito que a opacidade algorítmica apresenta em relação ao dever de motivação das decisões administrativas. Partindo da premissa de que a automatização das decisões algorítmicas é, em tese, possível no ordenamento jurídico brasileiro, e até desejável em alguma medida, o desafio proposto é pensar em como garantir um nível de motivação suficiente e infundir confiança nas decisões administrativas automatizadas, permitindo que os cidadãos possam efetivamente contra-argumentá-las.

No ponto, cabe, desde logo, fazer uma delimitação de escopo. É assente que a Administração Pública manifesta-se, na maior parte das vezes, por meio da expedição de atos administrativos.⁵ Quando esses atos produzem efeitos jurídicos no caso concreto, ou seja, possuem destinatários individualizados, diz-se, em contraposição aos atos gerais (também chamados de normativos), que se trata de atos administrativos individuais, também

⁴ Ministério da Ciência, Tecnologia e Inovações Secretaria de Empreendedorismo e Inovação, 'Estratégia Brasileira de Inteligência Artificial' (2021) <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-dia-gramacao_4-979_2021.pdf> acesso em 10 set 2022 (EBIA). Esclarece-se que a EBIA foi instituída por meio da Portaria GM nº 4.617/2021, publicada no Diário Oficial da União em 12 abril de 2021.

⁵ Adota-se o seguinte conceito de ato administrativo em sentido estrito: "declaração do Estado ou de quem o represente, que produz efeitos jurídicos imediatos, com observância da lei, sob regime jurídico de direito público e sujeita a controle pelo Poder Judiciário." Maria Sylvania Zanella Di Pietro, *Direito Administrativo* (22ª edição, Atlas 2008) 198.

conhecidos como atos concretos.⁶ Esses últimos podem ser impugnados diretamente pelos interessados quanto à legalidade, tanto na via administrativa como na judicial.⁷

Nessa linha, esclarece-se que o estudo feito neste trabalho tem como foco os atos administrativos individuais, os quais exprimem uma exteriorização da vontade de agentes da Administração Pública e contêm, portanto, uma decisão.⁸ É nesse sentido que, para esta obra, se refere ao termo “decisão administrativa”.⁹ Limita-se, portanto, o seu objeto aos sistemas de decisões algorítmicas utilizados para a produção de atos administrativos individuais de forma automatizada ou semiautomatizada.

Além disso, o estudo tem como foco o direito brasileiro, mas utilizará o direito europeu como elemento de direito comparado. No tocante a este, optou-se por realizar um estudo comparativo entre o RGPD e a LGPD no que tange, em particular, ao direito à explicação das decisões automatizadas, tendo por propósito trazer subsídios para a interpretação da legislação brasileira, mais recente e largamente inspirada pelo regulamento europeu.

Para dar maior concretude à discussão, será analisado um caso prático que demonstra a capacidade desafiadora do tema. Na falta de jurisprudência brasileira ou europeia sobre o ponto, elegeu-se o caso americano *State of Wisconsin v. Eric Loomis*, referente ao conhecido software COMPAS, utilizado em algumas jurisdições dos EUA para análise de risco de reincidência criminal.¹⁰ O intento não é analisar o direito americano em si, mas, a partir da análise desse caso, obter elementos para compreender, no mundo real, as dificuldades adicionais que se impõem ao cidadão que intenciona argumentar contra uma decisão tomada e motivada por um julgador humano com suporte em sistema algorítmico.

Assim, no primeiro capítulo, inicia-se o trabalho com a exposição do panorama geral sobre a tendência de automatização das atividades desempenhadas pela Administração Pública. Na sequência, será analisado o dever de motivação dos atos administrativos e serão apresentadas as novas formas de falibilidade que derivam do uso de algoritmos nas decisões.

⁶ José dos Santos Carvalho Filho, *Manual de Direito Administrativo* (32ª edição, Atlas 2018) 134-135. No mesmo sentido, Maria Sylvia Zanella Di Pietro, *Direito Administrativo* (22ª edição, Atlas 2008) 223.

⁷ Carvalho Filho (n 6) 135.

⁸ Antônio Carlos de Araújo Cintra, *Motivo e motivação do ato administrativo* (RT 1979) 106-107.

⁹ Aqui cabe o alerta feito por José dos Santos Carvalho Filho para a confusão que o termo “decisões administrativas” pode causar, uma vez que tanto pode-se entender que se trata de sinônimo de “atos administrativos”, como se pode interpretar no sentido de que se cuida de atos administrativos decisórios, proferidos em processos administrativos em que haja conflito de interesses. Para o fim deste trabalho, o termo decisão administrativa faz referência a decisões contidas em atos administrativos individuais, sejam eles expedidos no âmbito de processo administrativo ou não. Carvalho Filho (n 6) 21.

¹⁰ *State of Wisconsin v Eric L Loomis* (2016) 881 NW2d (Supreme Court of Wisconsin).

No segundo capítulo, o problema da explicação algorítmica será aprofundado, com olhar mais técnico sobre a opacidade, definindo-se o escopo do direito à explicação decorrente dos regimes de proteção de dados pessoais no contexto brasileiro e europeu.

No capítulo final, pretende-se estabelecer os novos contornos do direito à motivação das decisões administrativas na era digital, especialmente diante da análise do referido caso prático e dos principais documentos que recomendam o desenvolvimento e o uso ético da inteligência artificial.

1. Os algoritmos na Administração Pública Digital

1.1. A automatização dos serviços públicos

Verifica-se um interesse crescente dos governos pelo desenvolvimento e pelo uso de algoritmos voltados a aprimorar a entrega dos serviços públicos, haja vista as vantagens em termos de eficiência, redução de custos, acesso e cruzamento de informações. Esse interesse é evidenciado pelo número de países que desenvolveram estratégias digitais, dando passos rumo ao governo digital.¹¹ Não por acaso o imperativo denominado *digital by default* é consagrado nas estratégias digitais de diversos países, como Estados Unidos, Reino Unido,¹² Austrália¹³ e em documentos da União Europeia.¹⁴

A adoção de um governo digital requer a integração do uso das novas tecnologias digitais (como internet das coisas, computação em nuvem, *big data*, *analytics*, mídias sociais e inteligência artificial) entre os esforços para a modernização do setor público, de forma a redesenhar profundamente os processos e as entregas da Administração Pública em torno das expectativas e necessidades reais, complexas e dinâmicas dos administrados. Essas iniciativas, inspiradas na experiência do setor privado, vão muito além da mera transposição

¹¹ Adota-se aqui o termo “governo digital” por entender-se que é uma evolução na linha do tempo e um novo paradigma em relação ao “governo eletrônico” (e-government). Enquanto este é entendido como “a utilização pelos governos de tecnologias da informação e comunicação (TIC), e particularmente a Internet, como instrumento para alcançar uma melhor governança”, aquele representa um novo estágio de maturidade no uso das tecnologias digitais pelos governos, provendo uma visão de abertura, inovação e modernização do setor público, rumo a uma profunda transformação dos processos e dos serviços públicos. Nesse sentido, vide Organização para a Cooperação e Desenvolvimento Econômico, ‘Recommendation of the Council on Digital Government Strategies’ (2014) 1, 2 <<https://www.oecd.org/gov/digital-government/recommendation-on-digital-government-strategies.htm>> acesso em 10 jun 2022. Para mais informações sobre a evolução do governo eletrônico para o digital, vide Bouchaïb Bounabat, ‘From e-Government to digital Government: Stakes and Evolution Models’ (2017) 10 Electronic Journal of Information Technology 1.

¹² UK Cabinet Office, ‘Government Digital Strategy: action 6 - digital by default service standard’ (2013) <<https://www.gov.uk/government/publications/government-digital-strategy-action-6>> acesso em 12 jul 2022.

¹³ Government of South Australia, ‘Digital by Default Declaration’ (2014) <https://www.dit.sa.gov.au/__data/assets/pdf_file/0005/264695/Digital-by-Default-Declaration.pdf> acesso em 07 set 2022.

¹⁴ European Commission, ‘EU eGovernment Action Plan 2016-2020’ (Communication) COM(2016) 179 final.

dos serviços e do *modus operandi* analógico para o espaço virtual, mas visam efetivamente transformar a experiência dos administrados e a entrega dos serviços públicos. Nesse sentido, o foco na pura automação e na redução de custos mostra-se já ultrapassado.¹⁵ À Administração Pública cabe voltar-se para um funcionamento mais horizontal, organizado em rede, buscando o compartilhamento de informações, a integração entre sistemas, a transparência, a cooperação entre os cidadãos, os agentes do setor privado e os diversos órgãos administrativos, em uma relação ressignificada de efetiva colaboração e participação nas atividades administrativas.¹⁶

De fato, a integração das novas tecnologias ao cotidiano das pessoas e dos negócios permite transformar a relação entre esses agentes e a Administração Pública, viabilizando um novo grau de interação entre as esferas pública e privada.¹⁷ A comunicação pautada na conectividade oportuniza níveis de engajamento, participação e colaboração impossíveis até então, permitindo que os administrados participem ativamente, *e.g.*, do desenho das políticas públicas, da eleição de prioridades e da própria construção de soluções para os problemas sociais.¹⁸

Esse novo contexto muda a expectativa do cidadão acerca da sua relação com o governo. O cidadão da era digital espera obter, no serviços públicos, as mesmas facilidades de acesso, tempo de resposta e qualidade que encontra nos serviços privados.¹⁹ É, portanto, naturalmente mais conectado e exigente de um Estado mais aparelhado e responsivo, capaz de acompanhar as mudanças incessantes e aceleradas da sociedade do século XXI.²⁰ Além disso, esse cidadão espera estar mais próximo da Administração Pública por meio do uso da tecnologia, comunicar-se ativamente com ela no espaço virtual e participar efetivamente da construção das soluções para as suas necessidades. Por essa razão, diz-se que o governo digital deve englobar não apenas práticas para facilitar o acesso e melhorar a própria

¹⁵ Bounabat (n 11) 15.

¹⁶ José Fernando Ferreira Brega, 'Governo Eletrônico e Direito Administrativo' (Tese de doutorado, Universidade de São Paulo 2012) 34-35.

¹⁷ Exemplo disso é o uso das mídias sociais pelos governos, usualmente conhecido como Governo 2.0, uma evolução do e-government em relação à interatividade e à comunicação. Nesse sentido, vide Gohar Feroz Khan, "Social Media-based Government Explained: Utilization Model, Implementation Scenarios, and Relationships" in Imed Boughzala, Marijn Janssen e Saïd Assar, *Case Studies in e-Government 2.0: Changing Citizen Relationships* (Springer 2015) 15.

¹⁸ Organização para a Cooperação e Desenvolvimento Econômico, 'Recommendation of the Council on Digital Government Strategies' (2014) 1, 2 <<https://www.oecd.org/gov/digital-government/recommendation-on-digital-government-strategies.htm>> acesso em 10 jun 2022.

¹⁹ Bounabat (n 11) 12.

²⁰ José Luiz de Moura Faleiros Júnior, *Administração Pública Digital: proposições para o aperfeiçoamento do Regime Jurídico Administrativo na sociedade da informação* (Foco 2020) 81.

prestação dos serviços públicos, mas também deve ter como objetivo mais amplo melhorar as relações entre as instituições governamentais e os cidadãos.²¹

Considerando as novas expectativas dos administrados, o processo de transformação digital da Administração Pública torna-se não só interessante, mas, acima de tudo, absolutamente necessário na sociedade de informação. Se a integração da tecnologia ao cotidiano das pessoas estende a comunicação, as atividades e a própria prestação de serviços para o ambiente digital, como a Administração Pública assegurará eficiência sem estar nesse espaço?²² Então, nesse sentido, a transformação digital da Administração Pública é meio necessário para a concretização do princípio da eficiência, consagrado no artigo 37, *caput*, da Constituição da República Federativa do Brasil de 1988.

Ademais, esse aumento da expectativa em relação à entrega de valor público na era digital eleva também o patamar necessário para o ganho e a manutenção da confiança do cidadão nos governos, que é um dos pilares mais importantes para sustentar a legitimidade e a sustentabilidade dos sistemas políticos.²³ Por conseguinte, os governos são pressionados a modernizar a estrutura e as atividades administrativas com foco nas exigências e necessidades dos cidadãos, e não mais para os processos internos e sua lógica intrínseca.²⁴ Não é exagero afirmar, nessa linha, que “o processo de adaptação às mudanças produzidas pelo desenvolvimento tecnológico no campo da informação e da comunicação constitui um verdadeiro dever das Administrações Públicas no novo contexto da sociedade de informação do século XXI”.²⁵ Diz-se, então, que chegará o momento em que todos os atos e processos administrativos serão realizados diretamente na memória dos computadores, dispensando, finalmente, papéis e deslocamentos físicos, rumo a uma Administração Pública totalmente digital.²⁶

²¹ *ibid* 258.

²² Eficiência aqui pode ser compreendida como o “modo de organizar, estruturar, disciplinar a Administração Pública” com o objetivo de “alcançar os melhores resultados na prestação do serviço público”. Nesse sentido, Di Pietro (5) 82.

²³ Organização para a Cooperação e Desenvolvimento Econômico, ‘Trust in Government: Evidence, Policies and Decision-making’ (Meeting of the Council at Ministerial Level 2013) 1, 69 <[http://www2.oecd.org/oecdinfo/info.aspx?app=OLIScoteEN&Ref=C/MIN\(2013\)4/FINAL](http://www2.oecd.org/oecdinfo/info.aspx?app=OLIScoteEN&Ref=C/MIN(2013)4/FINAL)> acesso em 05 jul 2022. Em sentido contrário, vide Frank Bannister e Regina Connolly, ‘Trust and transformational government: A proposed framework for research’ (2011) 28(2) *Government Information Quarterly* 137.

²⁴ *ibid*.

²⁵ Brega (n 16) 39.

²⁶ *ibid* 42. A título de exemplo, aponta-se que o governo da Estônia disponibiliza 99% dos serviços públicos de forma digital, exigindo presença física apenas para casamento e divórcio. Ana Laura Stachewski, ‘Como a Estônia se tornou uma nação digital e um berço de negócios inovadores’ (Revista Pequenas Empresas & Grandes Negócios, 16 ago 2022) <<https://revistapegn.globo.com/Startups/noticia/2022/08/como-estonia-se-tornou-uma-nacao-digital-e-um-berco-de-negocios-inovadores.html>> acesso em 5 set 2022.

Assim, no compasso de desenvolvimento de governos digitais, uma das iniciativas possíveis é o uso de algoritmos para a tomada de decisões administrativas.²⁷ Apesar de os algoritmos não serem uma invenção recente, o seu uso em sistemas para apoiar a tomada de decisões vem crescendo tanto no setor privado como no setor público. De uma forma simplificada, esses sistemas de decisões automatizadas baseiam-se, fundamentalmente, na análise de grande quantidade de dados para derivar informações consideradas relevantes para a tomada de decisão.²⁸

No setor público, os exemplos de aplicação vão desde sistemas de avaliações de professores,²⁹ de previsão de risco de reincidência em processos criminais,³⁰ de distribuição de vagas em escolas, de concessão de benefícios de seguridade social, de processamento automático de declarações tributárias, de detecção de riscos de fraudes a procedimentos licitatórios e de corrupção.³¹ Há, ainda, aplicabilidades para o uso de sistemas de decisões automatizadas em outras áreas do setor público, como na segurança pública (e.g., policiamento preditivo), na ciberdefesa e na salvaguarda da democracia e da soberania nacional.³²

Não há dúvidas de que o uso desses sistemas apresenta vantagens, especialmente em relação à eficiência, com a redução de custos e do tempo de resposta da Administração.³³ Todavia, além disso, se bem concebidos e empregados, eles podem aumentar a transparência e a *accountability*. Sob essa perspectiva, a automatização das decisões administrativas poderia encorajar uma discussão ampla sobre os critérios usados pelo sistema, a lógica envolvida e as expectativas da sociedade em relação aos resultados, particularmente em termos de igualdade, justiça e não-discriminação.³⁴

²⁷ O conceito de algoritmo será detalhado no item 2.1 deste trabalho.

²⁸ Claude Castelluccia e Daniel Le Métayer, 'Understanding algorithmic decision-making: Opportunities and challenges' (European Parliamentary Research Service, 2019) 1.

²⁹ Rashida Richardson, 'Defining and Demystifying Automated Decision Systems' 81(3) Maryland Law Review 785, 808.

³⁰ Como, por exemplo, o software COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) utilizado em algumas jurisdições dos Estados Unidos e que será utilizado como exemplo no item 3.1 deste trabalho.

³¹ Jonas Valente, 'Órgãos públicos usam inteligência artificial para combater corrupção' (Agência Brasil, 3 ago 2018) <<https://agencia.brasil.ebc.com.br/geral/noticia/2018-08/orgaos-publicos-usam-inteligencia-artificial-para-combater-corrupcao>> acesso em 23 jul 2022.

³² A aplicabilidade dos sistemas de decisões automatizadas no setor público pode ter como objetivo a melhoria dos serviços públicos, o incremento da segurança pública ou o resguardo da soberania nacional. Castelluccia e Le Métayer (n 29) 19-20.

³³ A Constituição da República Federativa do Brasil, no seu Artigo 5º, LXXVIII, assim estabelece: "a todos, no âmbito judicial e administrativo, são assegurados a razoável duração do processo e os meios que garantam a celeridade de sua tramitação".

³⁴ Castelluccia e Le Métayer (n 29) 11.

Outra vantagem seria auxiliar na redução ou até na eliminação do erro humano, gerando decisões mais corretas, precisas e objetivas.³⁵ Especialmente no caso do uso de *machine-learning*, as vantagens apontadas são a capacidade de fazer previsões extraordinariamente complexas sem tanta dependência do conhecimento humano prévio, com maior precisão preditiva e eficiência estatística.³⁶ Sob essa perspectiva, a automatização das decisões poderia reduzir o ruído inerente às decisões humanas.³⁷

Por outro lado, o desafio regulatório para as decisões algorítmicas é dos mais relevantes das próximas décadas. E dentre os diversos riscos que podem ser apontados para a implementação de decisões algorítmicas no âmbito da Administração Pública, o presente trabalho destaca a preocupação em assegurar que os administrados afetados pelas decisões possam compreendê-las e desafiá-las, ou, dito de outra forma, que possam contrargumentar o algoritmo, especialmente diante da sua inerente opacidade.

Cabe ponderar que, até então, os seres humanos não haviam criado máquinas que produzissem comportamentos total ou em parte inexplicáveis por seus criadores.³⁸ Essa constatação, aliada ao fato de que o processo de tomada de decisão humana também apresenta alguma opacidade, provoca uma reflexão interessante a respeito dos diferentes graus de confiança depositado nesses dois tipos de decisão.

A capacidade de julgamento humano é, naturalmente, digna de confiança, reforçada pela argumentação, ainda que nem sempre se consiga – ou até mesmo seja desejável – explicar todo o *iter* percorrido pelo agente humano no processo decisório (e.g., os seres humanos executam a tarefa de reconhecimento visual de forma imediata e subconsciente). Ao contrário, a preocupação com a explicabilidade das decisões das máquinas deriva da falta de confiança sobre o processo de classificação algorítmica, essencialmente, porque pode operar de forma muito diferente do raciocínio humano, sem que se possa estabelecer uma linear relação de causa e efeito entre os valores de entrada e o resultado respectivo.

³⁵ Cary Coglianese, ‘Robot Regulators could eliminate human error’ (*The Regulatory Review*, 16 mai 2016) <<https://www.theregview.org/2016/05/16/coglianese-robot-regulators-eliminate-error/?msclkid=0b239623a94411ec803bebc5f6522e9a>> acesso em 07 jul 2022.

³⁶ Cary Coglianese e David Lehr, ‘Regulating by Robot: Administrative Decision Making in the Machine-Learning Era’ (2017) 105 *The Georgetown Law Journal* 1147, 1159.

³⁷ Nesse contexto, ruído deve ser compreendido como “uma variabilidade indesejada nos julgamentos sobre um mesmo problema”, conforme Daniel Kahneman, Olivier Sibony e Cass R. Sunstein, *Ruído: uma falha no julgamento humano* (Cássio de Arantes Leite tr, Objetiva 2021) 40.

³⁸ Will Knight, ‘The Dark Secret at the Heart of AI’ (2017) 120 *MIT Technology Review* <<https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/>> acesso em 07 jul 2022.

Por essa razão, aumentar a confiança no processo de decisão algorítmica é fundamental para se decida usar esses sistemas e, em especial, para que se possa cogitar a sua utilização no âmbito da Administração Pública, o que exige assegurar compreensão e controle sobre as novas formas de falibilidade que decorrem do uso de algoritmos nessas decisões. Nesse sentido, como será desenvolvido na próxima seção, o dever de motivação é elemento indispensável para que esses atos possam ser alvo de questionamento pelos afetados e, portanto, para que seja viabilizado o exercício do direito de revisão administrativa e judicial constitucionalmente previsto.

1.2. Motivação e controle da falibilidade humana

O agir da Administração Pública é pautado pela legalidade, o que justifica a existência de requisitos para o ato administrativo. Dentre eles, está o motivo, elemento que deve estar sempre presente e que constitui pressuposto de validade de qualquer ato administrativo.³⁹ De acordo com Celso Antônio Bandeira de Mello, “motivo é o pressuposto de fato que autoriza ou exige a prática ato. É, pois, a situação do mundo empírico que deve ser tomada em conta para a prática do ato. Logo, é externo ao ato. Inclusive o antecede.” Assim, toda vontade emitida por agente da Administração decorre da verificação de razões de fato ou de direito, que correspondem, portanto, ao motivo do ato administrativo.⁴⁰

Por sua vez, a motivação, segundo o autor, “é a exposição dos motivos, fundamentação na qual são enunciados (a) a regra de direito habilitante, (b) os fatos em que o agente se estribou para decidir e, muitas vezes, obrigatoriamente, (c) a enunciação da relação de pertinência lógica entre os fatos ocorridos e o ato praticado”.⁴¹ Ou, ainda, motivação é a “justificativa do pronunciamento tomado”.⁴²

No ordenamento jurídico brasileiro, discute-se se o dever de motivação dos atos administrativos tem ou não assento constitucional, em face do escopo restrito do Artigo 93, X, da Constituição da República, que se situa no capítulo constitucional destinado apenas ao Poder Judiciário.⁴³

³⁹ Carvalho Filho (6) 86.

⁴⁰ *ibid* 118.

⁴¹ Celso Antônio Bandeira de Mello, *Curso de Direito Administrativo* (26ª edição, Malheiros 2009) 392-395.

⁴² José Cretella Júnior, *Curso de Direito Administrativo* (Forense 1986) 310.

⁴³ Artigo 93. Lei complementar, de iniciativa do Supremo Tribunal Federal, disporá sobre o Estatuto da Magistratura, observados os seguintes princípios: (...) X - as decisões administrativas dos tribunais serão motivadas e em sessão pública, sendo as disciplinares tomadas pelo voto da maioria absoluta de seus membros.

A nível infraconstitucional, o Artigo 2º da Lei nº 9.784/99 assegura a motivação como princípio ao estabelecer que “a Administração Pública obedecerá, dentre outros, aos princípios da legalidade, finalidade, motivação, razoabilidade, proporcionalidade, moralidade, ampla defesa, contraditório, segurança jurídica, interesse público e eficiência”. Por sua vez, o Artigo 50 da aludida lei estabelece expressamente que os atos administrativos “deverão ser motivados, com indicação dos fatos e dos fundamentos jurídicos” quando, dentre outras hipóteses, “neguem, limitem ou afetem direitos ou interesses”, “imponham ou agravem deveres, encargos ou sanções” e “decidam recursos administrativos”.⁴⁴

Tendo em vista que a referida legislação estabelece, explicitamente, os casos em que há obrigatoriedade de motivação, há autores que sustentam, *a contrario sensu*, a inexistência de uma obrigação generalizada de motivação de todos os atos administrativos, uma vez que aqueles não listados no aludido dispositivo prescindiriam de motivação para sua validade.⁴⁵ A vertente majoritária, porém, deriva a motivação da própria ordem constitucional e interpreta o disposto no Artigo 50º da Lei nº 9.784/99 como concretização do dever constitucional de motivar, entendendo que o rol nele previsto não é taxativo.⁴⁶

Com efeito, o movimento histórico de evolução do Direito Administrativo caminhou no sentido do robustecimento do dever de motivação, para que seja regra nos atos administrativos, sejam eles de natureza vinculada ou discricionária. A razão de ser dessa evolução coincide com a ressignificação do próprio papel da Administração Pública no Estado Democrático de Direito, de que resulta o direito fundamental à boa administração.⁴⁷ Nesse sentido, a exigência de que as decisões administrativas sejam motivadas, com suporte em razões objetivas e consistentes, é fruto da transição para o Direito Administrativo dialógico, em oposição ao autocrático, e revela a importância de que a Administração Pública zele pela estabilidade e pela ordem nas relações jurídicas.⁴⁸

⁴⁴ Artigo 50. Os atos administrativos deverão ser motivados, com indicação dos fatos e dos fundamentos jurídicos, quando: I - neguem, limitem ou afetem direitos ou interesses; II - imponham ou agravem deveres, encargos ou sanções; III - decidam processos administrativos de concurso ou seleção pública; IV - dispensem ou declarem a inexigibilidade de processo licitatório; V - decidam recursos administrativos; VI - decorram de reexame de ofício; VII - deixem de aplicar jurisprudência firmada sobre a questão ou discrepem de pareceres, laudos, propostas e relatórios oficiais; VIII - importem anulação, revogação, suspensão ou convalidação de ato administrativo.

⁴⁵ Carvalho Filho (n 6) 120.

⁴⁶ Juarez Freitas, *O controle dos atos administrativos e os princípios fundamentais* (Malheiros 2004) 260; Di Pietro (n 5) 81.

⁴⁷ Juarez Freitas, *Discricionariedade administrativa e o direito fundamental à boa administração pública* (Malheiros 2007).

⁴⁸ Freitas (n 47) 63.

Há quem sustente que o fundamento constitucional da obrigação de motivar encontra assento na própria cidadania, estabelecida como fundamento da República no artigo 1º, II, de Constituição. Afinal, se a Carta Constitucional estabelece que “todo o poder emana do povo”,⁴⁹ o mínimo que se pode assegurar no Estado de Direito é que o cidadão, titular último do poder, não esteja sujeito a decisões arbitrárias, imotivadas e sigilosas tomadas por quem deveria servi-lo.⁵⁰ Assim, a motivação tem a finalidade de democratizar o exercício da função administrativa, contribuindo para que se contenha a atividade do administrador dentro do regime jurídico-administrativo, inspirando confiança e certeza da legalidade de suas decisões.⁵¹

Entende-se, assim, que reconhecer o dever de motivação como regra significa dar a devida importância à análise do motivo (e, portanto, à necessidade de declinação explícita deste pelo administrador) como elemento de controle dos atos administrativos. A propósito, a Constituição da República assegura, em seu Artigo 5º, XXXV, que “a lei não excluirá da apreciação do Poder Judiciário lesão ou ameaça a direito”. Então, até para viabilizar a submissão do ato administrativo ao controle jurisdicional as suas razões fundantes precisam ser conhecidas.⁵²

Assim, ainda que exista divergência doutrinária a respeito da obrigatoriedade generalizada ou não de motivação dos atos administrativos, não se discute que o administrador deve ser incentivado a declinar a justificativa do ato sempre que possível, a fim de assegurar transparência e, portanto, maior controle aos atos da Administração.⁵³ Nesse sentido, entende-se que a motivação deve estar presente em todos os atos administrativos, à exceção dos de mero expediente, dos autodecifráveis por sua singeleza e daqueles em que a própria Constituição dispensar a obrigatoriedade de motivar (e.g., nomeação ou exoneração de cargos em comissão).⁵⁴

Na sequência, cabe tecer algumas considerações a respeito de como deve ser a motivação. Novamente, o Artigo 50 da Lei nº 9.784/99, em seus parágrafos, confere algumas diretrizes. Primeiro, estabelece que “a motivação deve ser explícita, clara e congruente”. O dispositivo autoriza inclusive a utilização de “meio mecânico que reproduza os fundamentos das

⁴⁹ Artigo 1º, parágrafo único, da Constituição da República Federativa do Brasil (1988).

⁵⁰ Bandeira de Mello (n 42) 113.

⁵¹ Florivaldo Dutra de Araújo, *Motivação e controle do ato administrativo* (2ª edição, Del Rey 2005) 109-110.

⁵² Celso Antônio Bandeira de Mello, *Discrecionalidade e controle jurisdicional* (2ª edição, Malheiros 2003) 99.

⁵³ Carvalho Filho (n 6) 121.

⁵⁴ Freitas (n 48) 47.

decisões” quando se tratar da “solução de vários assuntos da mesma natureza” e “desde que não prejudique direito ou garantia dos interessados”. Entende-se, portanto, que a exposição dos motivos não pode ser genérica ou indefinida,⁵⁵ devendo demonstrar que foram sopesadas as peculiaridades e circunstâncias do caso concreto, de modo que referências a expressões como “interesse público” são insuficientes.⁵⁶

Além disso, tradicionalmente na doutrina administrativista, exige-se que a motivação seja apresentada anteriormente ou concomitantemente à prática do ato administrativo. Essa preocupação se justifica para impedir que o administrador público possa, após a expedição do ato, fabricar, criar ou forjar motivos para burlar eventual impugnação.⁵⁷ Nesse sentido, há jurisprudência recente do Superior Tribunal de Justiça brasileiro a afirmar que “não se deve admitir como legítima, portanto, a prática imotivada de um ato que, ao ser contestado na via judicial ou administrativa, faça com que o gestor construa algum motivo que dê ensejo à validade do ato administrativo”.⁵⁸

A depender do grau de vinculação do poder conferido pela lei à Administração Pública, a tarefa de oferecer motivação suficiente poderá exigir maior ou menor esforço argumentativo. Assim, para os atos administrativos vinculados, em que a lei estabelece todos os elementos da situação de fato, verificada a sua ocorrência no plano empírico, cabe ao administrador agir de tal ou qual forma,⁵⁹ de modo que o exame de legalidade consistirá substancialmente no confronto do motivo do ato com o motivo legalmente previsto.⁶⁰ Ainda que sempre haja janela para interpretação, nesses casos, o espaço para subjetivismo é reduzido, visto que a lei não dá opções ao administrador.⁶¹ Então, a motivação (entendida como exposição dos motivos) dos atos vinculados visará a dar conhecimento acerca da interpretação conferida pelo agente à lei e a tornar possível a verificação da subsunção do ato administrativo à situação fática que o tenha justificado.⁶²

⁵⁵ Carvalho Filho (n 6) 122.

⁵⁶ Cintra (n 8) 128.

⁵⁷ Bandeira de Mello (n 42) 112.

⁵⁸ Superior Tribunal de Justiça, AgInt no AgInt no AREsp (1108757/PI) Rel. Ministro NAPOLEÃO NUNES MAIA FILHO, julgado em 30/11/2020 (DJe 03/12/2020) <disponível em <https://scon.stj.jus.br/SCON/GetInteiroTeorDoAcordao?num_registro=201701238930&dt_publicacao=03/12/2020> acesso em 10 jul 2022.

⁵⁹ Maria Sylvia Zanella Di Pietro, *Discricionariedade administrativa na Constituição de 1988* (3ª edição, Atlas 2012) 61.

⁶⁰ Carvalho Filho (n 6) 118-119. Nesse sentido também Celso Antônio Bandeira de Mello, para quem “naqueles em que há aplicação quase automática da lei, por não existir campo para interferência de juízos subjetivos do administrador, a simples menção do fato e da regra de Direito aplicanda pode ser suficiente, por estar implícita a motivação”. Bandeira de Mello (n 42) 112.

⁶¹ Araújo (n 52) 112.

⁶² *ibid.*

Diferentemente, nos casos em que haja discricionariedade administrativa, a lei, ao regular o comportamento da Administração Pública, confere maior margem de liberdade ao agente público, para que possa optar por uma dentre as várias soluções possíveis, todas válidas perante o direito.⁶³ Essa análise deve ser feita, por certo, mediante avaliação de critérios de conveniência e oportunidade dentro das balizas do regime jurídico-administrativo, sob pena de ser arbitrária. O objetivo de conferir essa margem de discricionariedade é justamente permitir que o administrador possa adotar a melhor solução⁶⁴ levando em conta as particularidades de cada caso concreto e tendo flexibilidade para responder às necessidades sociais de cada momento.⁶⁵

Soa lógico que, quanto maior for o espaço de escolha conferido à Administração Pública, maior deve ser o ônus argumentativo de justificar a solução adotada. Tendo em vista que o grau de subjetivismo é maior, a fundamentação exigida há de ser naturalmente mais robusta para que seja entendida como suficiente.⁶⁶ Assim, quando a lei conferir maior margem de discricionariedade à conduta administrativa, na motivação, o agente público deve, em acréscimo, expor as razões pelas quais escolheu aquele caminho dentre outros possíveis.⁶⁷ Dito de outra forma, exigir-se-á, nesses casos, a indicação da “pertinência lógica entre o supedâneo fático e a medida tomada, de maneira a se poder compreender sua idoneidade para lograr a finalidade legal”.⁶⁸

Além disso, impera reconhecer que essa aferição de subsunção do motivo previsto para o ato na hipótese normativa à realidade empírica é mais simples de ser feita nos casos em que a regra de direito caracteriza o motivo utilizando conceitos precisos, determinados (e.g., aposentadoria por idade aos 65 anos). Todavia, quando utiliza conceitos vagos ou indeterminados (e.g., interesse público, boa fé, probidade), essa aferição será mais complexa por haver maior margem de imprecisão e, portanto, maior subjetivismo.⁶⁹

⁶³ Di Pietro (n 60) 61-62.

⁶⁴ Bandeira de Mello (n 53) 35. Refere o autor: “é exatamente porque a norma legal só quer a solução ótima, perfeita, adequada às circunstâncias concretas, que, ante o caráter polifacético, multifário, dos fatos da vida, se vê compelida a outorgar ao administrador – que é quem se confronta com a realidade dos fatos segundo seu colorido próprio – certa margem de liberdade para que este, sopesando as circunstâncias, possa dar verdadeira satisfação à finalidade legal.”

⁶⁵ Di Pietro (n 60) 63.

⁶⁶ Segundo Florivaldo Dutra de Araújo, suficiência significa que “a Administração Pública deve dar ideia completa do processo lógico e jurídico percorrido até a decisão”. Araújo (n52) 122.

⁶⁷ Nesse sentido, Cintra (n 8) 128.

⁶⁸ Bandeira de Mello (n 53) 99.

⁶⁹ *ibid* 90-92. Para aprofundamento da discussão sobre a relação entre os conceitos jurídicos indeterminados e a discricionariedade administrativa, com análise sobre ordenamentos jurídicos estrangeiros, vide Di Pietro (n 60) 90-119.

Por fim, cabe ponderar que a exigência de motivação parte do pressuposto da intrínseca e inevitável falibilidade humana⁷⁰ e da necessidade de conhecer e compreender as justificativas de uma decisão administrativa para a interpretação e para o controle do ato, incluindo a submissão ao crivo jurisdicional, se necessário.⁷¹ O dever de motivação, tal como foi concebido e fortalecido ao longo do tempo, tem como finalidade a democratização do exercício da função administrativa, impondo, de imediato, um freio para o próprio agente tomador de decisão, e viabilizando, a seguir, as demais formas de controle do ato, desde internas até as mais amplas e difusas, como a opinião pública.⁷²

Diante do exposto, mostra-se essencial assegurar que esse tão importante dever, consagrado e fortalecido com a evolução do Direito Administrativo, não esmoreça perante o cenário de algoritmização das atividades administrativas. Para tanto, é preciso compreender as novas formas de falibilidade que decorrem do contexto tecnológico, o que será feito nas linhas que seguem.

1.3. Motivação e controle da falibilidade algorítmica

Como visto, a exigência de motivação parte do pressuposto da inerente imperfeição humana⁷³ e da necessidade de conhecer e compreender as justificativas de uma decisão administrativa para assegurar o seu controle.⁷⁴ Por um lado, o uso de sistemas de decisões automatizadas pode reduzir o erro humano, o que, sem dúvida, é um propósito mais do que desejável. Todavia, em contrapartida, ao transpor a tomada de decisões administrativas para o ambiente digital, incorporam-se outros riscos.

Sob esse prisma, destacam-se, nas linhas a seguir, quatro aspectos que podem contribuir para a falibilidade das decisões algorítmicas: a falta de compreensão de que há decisões humanas no processo de construção de algoritmos e bases de dados; a transposição de vieses pré-existentes para os algoritmos; a necessidade de interpretação dos resultados combinada com a tendência humana de superestimar a sua objetividade; e a opacidade, que é tema central neste estudo.

⁷⁰ Bandeira de Mello (n 53) 102.

⁷¹ *ibid* 99.

⁷² Araújo (n 52) 110.

⁷³ Bandeira de Mello (n 53) 102.

⁷⁴ Araújo (n 52) 111. O autor bem ressalta que “além dos controles jurisdicional, administrativo e parlamentar, existe o controle pela opinião pública, que pode se perfazer não só depois de exauridos os efeitos do ato, mas em todo o processo de formação da vontade administrativa, dependendo dos mecanismos que compõem os canais de ligação entre Estado e sociedade”.

Bases de dados e algoritmos são criados por seres humanos e, portanto, exigem escolhas. Por essa razão, Cathy O’Neil escreveu que “models are opinions embedded in mathematics”.⁷⁵ De fato, a criação de modelos estatísticos envolve decisões humanas ao longo de todo o processo, *inter alia*, na definição do objetivo perseguido pelo modelo, na tradução desse objetivo em uma variável específica de resultado (que, de forma simplificada, representa a expressão matemática do objetivo do algoritmo), na seleção das bases de dados e na forma como eles serão coletados, preparados e tratados, nos ajustes de parâmetros, no desenho dos algoritmos em si e na interpretação dos resultados.⁷⁶ Assim, por trás do manto da neutralidade e da objetividade do resultado matemático, há um processo composto por escolhas humanas – tanto na construção das bases de dados como no desenho e na execução do próprio algoritmo – que se incorpora ao modelo, influenciando o resultado.

Como a acurácia depende muito – ainda que não exclusivamente – da base de dados, a utilização de um conjunto de dados escasso ou de baixa qualidade, por exemplo, comprometerá os índices de acerto do modelo, fenômeno que pode ser resumido na emblemática expressão “garbage in, garbage out”. Com efeito, os dados utilizados devem ser grandes em número e em variedade, visando reproduzir a representatividade do mundo real, uma vez que, em geral, “algoritmos de aprendizado de máquina são tão bons quanto os dados que são fornecidos para eles analisarem”.⁷⁷

Ademais, a insuficiência ou a contaminação da base de dados pode também levar ao enviesamento e a resultados discriminatórios, o que acontece quando o algoritmo fornece resultados sistematicamente diferentes para determinados indivíduos ou grupos de pessoas com base em atributos particulares que não poderiam levar a um tratamento diferenciado (como gênero, raça ou religião).⁷⁸

O enviesamento pode derivar de uma série de razões, incluindo as escolhas humanas feitas na seleção, coleta e preparação dos dados utilizados para o treinamento do algoritmo.⁷⁹ Assim, se, por exemplo, a base de dados de treinamento for enviesada em relação a um certo grupo, o algoritmo replicará, nos resultados, esse preconceito humano que está embutido na

⁷⁵ Cathy O’Neil, *Weapons of Math Destruction* (Broadway Books 2017) 21.

⁷⁶ Aleš Završnik, ‘Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings’ (2019) *European Journal of Criminology* 1; Jenna Burrell, ‘How the machine “thinks”: Understanding opacity in machine learning algorithms’ (2016) *Big Data & Society* 1.

⁷⁷ Harry Surden, ‘Machine Learning and Law’ 89 *Washington Law Review* (2014) 87, 106.

⁷⁸ European Union Agency for Fundamental Rights, ‘Big data: discrimination in data-supported decision making’ (Publications Office 2018) 1, 5.

⁷⁹ *ibid.*

base de dados. Isso porque os dados não são “limpos” de circunstâncias sociais, culturais e econômicas, nem dos preconceitos inerentes à própria linguagem humana.⁸⁰

Da mesma forma, as correlações indiretas feitas a partir de *proxies* podem levar a resultados discriminatórios. Se a base de dados contiver, ainda que indiretamente, informações relativas a características protegidas (gênero, etnia, religião), o resultado pode ser discriminatório por levar em consideração essas características que não deveriam ser sopesadas. Para citar um exemplo, verificou-se que um programa para avaliar candidatos a vagas de emprego utilizado nas décadas de 1970 e 1980 por um hospital em Londres apresentava resultados enviesados em detrimento de minorias e mulheres por inferência dessa informação a partir dos sobrenomes e do lugar de nascimento, diminuindo as chances de serem selecionadas para a entrevista.⁸¹

No que tange ao desenho do algoritmo em si, importa perceber que há uma escolha de valores embutida nos cálculos matemáticos (e.g., quando se decide dar mais peso à faixa etária do que ao estado civil em um modelo preditivo de reincidência criminal). Por essa razão, os profissionais jurídicos precisam estar mais bem habilitados a compreender, analisar e questionar esses valores – e as escolhas – que integram o desenho do próprio algoritmo.⁸² Como não existe, matematicamente, uma definição única de justiça,⁸³ uma das questões mais problemáticas é justamente compreender e definir a teoria de justiça que estará implicitamente incorporada no algoritmo, especialmente quando se pensa, por exemplo, no uso de algoritmos por tribunais.⁸⁴

As aludidas considerações são chave para a atividade humana de interpretar o resultado fornecido pelo sistema de decisão automatizada. Primeiro porque, na interpretação, é de suma relevância estar ciente de que as correlações entre as variáveis não implicam causalidade. Essa sentença, aliás, é frequentemente repetida por estatísticos, cujo significado pode ser assim resumido: o fato de duas variáveis exibirem um padrão semelhante de

⁸⁰ Završnik (n 77) 12. Outro exemplo de estudo conduzido nesse sentido é o seguinte: Aylin Caliskan, Joanna J Bryson e Arvind Narayanan, 'Semantics derived automatically from language corpora contain human-like biases' (2017) 356 (6334) *Science* 183.

⁸¹ Stella Lowry e Gordon Macpherson, 'A blot on the profession' (1988) 296(6623) *British medical journal* 657.

⁸² Završnik (n 77) 9.

⁸³ Anne L Washington, 'How to Argue with an Algorithm: Lessons from the COMPAS ProPublica Debate' (2019) 17(1) *Colorado Technology Law Journal* 131, 150.

⁸⁴ Angèle Christin, Alex Rosenblat e Danah Boyd, 'Courts and Predictive Algorithms' (2015) *Data & Civil Rights: A New Era of Policing and Justice* <http://www.datacivilrights.org/pubs/2015-1027/Courts_and_Predictive_Algorithms.pdf> acesso em 10 set 2022.

ocorrências não significa que tenham entre si uma relação de causa e efeito.⁸⁵ Em suma, com resultados derivados do aprendizado de máquina, as relações causais entre entradas e saídas podem simplesmente não existir, por mais intuitivas que tais relações possam parecer à superfície. Se um algoritmo de aprendizagem por máquina tende a prever que indivíduos mais velhos cometem menos crimes do que indivíduos mais novos, por exemplo, não se pode afirmar, com base nesse processo, que a idade mais avançada causa qualquer redução na propensão para cometer crimes.⁸⁶

Segundo porque se deve reconhecer que, ainda que se busque a objetividade estatística, a atividade de interpretação do resultado envolve algum grau de subjetividade.⁸⁷ Nesse sentido, assim como vieses podem estar embutidos dentro do modelo estatístico – a partir de uma base de dados insuficiente, por exemplo –, eles também podem surgir na atividade humana de interpretação do resultado, visto que a visão de mundo do intérprete exercerá influência na sua análise.⁸⁸

Terceiro porque, em geral, o ser humano tende a superestimar a acurácia, a objetividade e a confiabilidade das recomendações ou informações fornecidas por sistemas automatizados. Por falta de tempo, contexto ou habilidade técnica para analisar e compreender o resultado algorítmico, ou até mesmo em um impulso involuntário de reduzir o peso da responsabilidade da decisão, existe uma propensão natural dos tomadores de decisão em seguir a recomendação do sistema automatizado.⁸⁹ Esse traço é também parcialmente explicado pela propensão humana de adotar o caminho com menor esforço cognitivo no processo decisório.⁹⁰

Por fim, cabe reconhecer que o próprio algoritmo pode tomar um caminho não previsto, apesar das boas intenções de seus criadores. Então, há um risco de que, inadvertidamente, se produzam decisões equivocadas, “aprendidas” a partir de correlações falsas no processamento do conjunto de dados de treinamento (e.g., verificou-se que um algoritmo treinado para reconhecer lobos e cachorros da raça husky, apesar de apresentar bom nível

⁸⁵ Zach Bobbitt, ‘Correlation Does Not Imply Causation: 5 Real-World Examples’ (18 ago 2021) <<https://www.statology.org/correlation-does-not-imply-causation-examples/>> acesso em 15 ago 2022.

⁸⁶ Coglianese e Lehr (n 37) 1159.

⁸⁷ ‘Correlation vs Causation’ (*The Decision Lab*) <<https://thedeclaration.com/reference-guide/philosophy/correlation-vs-causation>> acesso em 10 set 2022.

⁸⁸ Završnik (n 77) 10.

⁸⁹ Frederik Zuiderveen Borgesius, ‘Discrimination, Artificial Intelligence and Algorithm Decision-Making’ (2018) Council of Europe 11-12.

⁹⁰ Matthew Grissinger, ‘Understanding Human Over-Reliance On Technology’ (2019) 44(6) PT 375.

de acurácia, embasava suas classificações na presença ou não de neve ao fundo da imagem).⁹¹

Talvez o maior desafio esteja na dificuldade (ou até impossibilidade) de entender o caminho que levou àquele resultado. Essa questão será detalhada no capítulo a seguir. Todavia, por ora, cabe dizer que a literatura aponta para o fato de que o receptor do resultado (*output*) da decisão algorítmica, em geral, não tem condições concretas de dizer como ou por que uma específica classificação foi derivada dos *inputs*.⁹² Como os algoritmos usados em aplicações de *machine-learning* e inteligência artificial são cada vez mais complexos e, além disso, há cada vez mais dados sendo utilizados, torna-se uma tarefa muito desafiadora precisar qual informação influenciou o resultado e em que medida.⁹³

Nesse contexto, por um lado, reconhece-se que, especialmente em decisões que afetam a vida das pessoas, é necessário assegurar transparência, no sentido de que o código e os dados utilizados para a construção do algoritmo estejam acessíveis para auditamento.⁹⁴ No entanto, a transparência encontra óbices relacionados à inerente opacidade de métodos de *machine-learning* e redes neurais (que são *black-boxed by default*⁹⁵), bem como ao fato de que transparência sem explicabilidade não resolve o problema, como será aprofundado a seguir neste trabalho.

Os aludidos pontos apresentam-se como desafio ao controle das decisões administrativas algorítmicas, visto que elas estarão expostas a essas novas formas de falibilidade, que precisam ser evitadas e controladas (verificadas e corrigidas, caso ocorram). É inegável, portanto, a necessidade de conferir grau de explicabilidade suficiente a essas decisões. Trata-se inclusive de tema palpitante a nível internacional.

No Canadá, a diretiva sobre as decisões automatizadas (Directive on Automated Decision-Making)⁹⁶ estabelece o dever de explicação, garantindo que os cidadãos sejam informados *ex ante* de que a decisão será tomada total ou parcialmente com o uso de um sistema automatizado, bem como que, após a decisão, os afetados recebam explicação significativa sobre como e por que a decisão foi tomada. A diretiva divide os tipos de decisões

⁹¹ Marco Tulio Ribeiro, Sameer Singh e Carlos Guestrin, 'Why Should I Trust You?: Explaining the Predictions of Any Classifier' (2016) arXiv <<https://arxiv.org/abs/1602.04938>> acesso em 10 set 2022.

⁹² Jenna Burrell, 'How the machine "thinks": Understanding opacity in machine learning algorithms' (2016) *Big Data & Society* 1, 1.

⁹³ European Union Agency for Fundamental Rights (n 80) 6.

⁹⁴ *ibid.*

⁹⁵ Završnik (n 77) 9.

⁹⁶ Government of Canada, Directive on Automated Decision-Making (2019) <<https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592§ion=html>> acesso em 10 set 2022.

administrativas automatizadas em quatro níveis, conforme o grau de impacto aos administrados, aumentando o grau de exigência à medida que cresce o risco e a irreversibilidade dos efeitos da decisão para os cidadãos.

Na França, a *Loi pour une République numérique* (Loi 2016-1321, de 07/10/2016),⁹⁷ em conjunto com o Decreto nº 2017-330, estabeleceu critérios para o direito à explicação no âmbito das decisões da Administração Pública, que incluem indicar o grau e o modo de contribuição do algoritmo para a decisão.

No Brasil, por sua vez, o Conselho da Justiça Federal, na Plenária da I Jornada de Direito Administrativo, aprovou o seguinte enunciado: “a decisão administrativa robótica deve ser suficientemente motivada, sendo a sua opacidade motivo de invalidação”.⁹⁸

Ademais, em se tratando do uso da inteligência artificial, a recomendação da Organização para a Cooperação e o Desenvolvimento Econômico (OCDE), da qual o Brasil é signatário, estabeleceu princípios e diretrizes para o desenvolvimento responsável da inteligência artificial.⁹⁹ Dentre os elementos constantes no documento destaca-se a necessidade de comprometimento com a transparência e com o fornecimento de informações que permitam “(i) promover a compreensão geral sobre sistemas de IA; (ii) tornar as pessoas cientes quanto às suas interações com sistemas de IA; (iii) permitir que aqueles afetados por um sistema de IA compreendam os resultados produzidos; e (iv) permitir que aqueles adversamente afetados por um sistema de IA possam contestar seu resultado”.

Também nesse sentido, a Estratégia Brasileira de Inteligência Artificial destaca a importância de que as decisões tomadas por sistemas automatizados sejam passíveis de explicação e interpretação. Todavia, também reconhece que a divulgação de informações em excesso sobre o funcionamento de um algoritmo ou um processo de inteligência artificial pode confundir ou não ser inteligível para os cidadãos interessados, ou ainda revelar indevidamente segredos comerciais, industriais e de propriedade intelectual. Além disso, admite que, algumas vezes, a explicação em nível detalhado sobre uma decisão ou resultado específico sequer é viável tecnicamente diante dos algoritmos de “sistemas fechados”. No

⁹⁷ France, LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique (2019) <<https://www.legifrance.gouv.fr/loda/id/JORFTEXT000033202746/>> acesso em 10 set 2022.

⁹⁸ ‘I Jornada de Direito Administrativo aprova 40 enunciados’ (*Conselho da Justiça Federal*, 10 ago 2022) <<https://www.cjf.jus.br/cjf/noticias/2020/08-agosto/i-jornada-de-direito-administrativo-aprova-40-enunciados>> acesso em 10 ago 2022.

⁹⁹ Organização para a Cooperação e Desenvolvimento Econômico, ‘Recommendation of the Council on Artificial Intelligence’ (2019) C/MIN(2019)3/FINAL <<https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>> acesso em 05 jul 2022

documento, entende-se que “o foco da explicabilidade deve se voltar ao fornecimento de informações significativas que permitam interpretar o sistema”.¹⁰⁰

Do exposto até aqui, conclui-se que a motivação, como tradicionalmente concebida, serve como importante forma de controle da falibilidade humana na tomada de decisões, visando a permitir o controle de conformidade dos atos da Administração com o regime jurídico-administrativo, cujos pilares são os princípios da supremacia do interesse público sobre o privado e a indisponibilidade do interesse público pela Administração.¹⁰¹ Todavia, quando se cogita o uso de sistemas de decisões automatizadas no âmbito administrativo, essa exigida motivação deve ser qualificada, porque ela deverá acrescentar a explicabilidade da decisão algorítmica como forma de garantir também o controle das novas formas de falibilidade que surgem em decorrência do uso dessa tecnologia.

Com esse propósito, as linhas a seguir destinam-se a aprofundar a análise sobre o problema da explicação algorítmica, com duplo olhar: primeiramente, sob o aspecto técnico no que tange à opacidade e, a seguir, com enfoque jurídico sobre o direito à explicação que decorre da legislação de proteção de dados pessoais.

2. A decisão algorítmica: entre o dever de motivação e a explicação possível

2.1. Os desafios impostos pelos sistemas de decisões algorítmicas

Algoritmos são uma sequência de instruções, regras e cálculos executados por um computador em uma ordem específica para gerar um resultado, geralmente uma resposta a um problema indicado.¹⁰² Assim, o algoritmo descreve um procedimento computacional bem caracterizado que toma algum valor como entrada (dados de entrada ou *input*) e produz, ao final da execução das etapas previamente definidas, um resultado (dados de saída ou *output*).¹⁰³ De um modo geral, a grande vantagem da utilização de algoritmos é a capacidade que possuem de fazer classificações e categorizações com eficiência.¹⁰⁴

Algoritmos podem atuar de forma isolada ou de forma combinada com outros algoritmos para resolver problemas mais complexos.¹⁰⁵ Além disso, podem ser codificados

¹⁰⁰ EBIA (n 3) 21.

¹⁰¹ Bandeira de Mello (n 53) 56.

¹⁰² Fernanda de Carvalho Lage, *Manual de Inteligência Artificial no Direito Brasileiro* (JusPodivm 2021) 37.

¹⁰³ *ibid* 38.

¹⁰⁴ Karen M Richmond, ‘AI, Machine Learning, and International Criminal Investigations: The lessons from forensic science’ iCourts Working Paper Series 222/2020, 7 <<https://ssrn.com/abstract=3727899>> acesso em 10 ago 2022.

¹⁰⁵ *ibid*.

manualmente por programadores (algoritmos convencionais) ou gerados automaticamente a partir de dados, como é o caso dos algoritmos derivados da aprendizagem de máquina (*machine-learning algorithms*). Um algoritmo de *machine-learning* é, assim, um algoritmo que, a partir de dados, gera outro algoritmo automaticamente, também conhecido como modelo.¹⁰⁶

Nesse contexto, entende-se por sistema de decisão algorítmica uma classe específica de algoritmos desenvolvida com o objetivo de apoiar o processo de tomada de decisão.¹⁰⁷ O envolvimento humano nessas decisões algorítmicas é variável, de modo que os sistemas podem servir como suporte para uma decisão humana ao final (gerando, então, decisões semi-automatizadas¹⁰⁸), ou podem gerar decisões totalmente automatizadas,¹⁰⁹ produzidas pelos algoritmos sem qualquer participação humana.¹¹⁰

Embora nem todos os sistemas de decisões algorítmicas sejam fundados em técnicas de inteligência artificial (das quais *machine-learning* é exemplo),¹¹¹ frequentemente, o uso de sistemas de decisões automatizadas pela Administração Pública é abordado dentro dessa temática. Nessa linha, nas estratégias nacionais específicas para o uso da inteligência artificial na transformação do setor público, o uso desta como apoio aos processos de tomada de decisão costuma ser trazido como um dos pontos principais.

Apesar de existirem muitas definições de inteligência artificial, para o escopo deste estudo, é suficiente dizer que o seu principal objetivo é permitir que as máquinas se comportem de forma inteligente, desempenhando funções semelhantes às humanas. Há variados métodos utilizados para que os computadores possam aprender e operar, ou seja, atuar de forma inteligente. Dentre eles, destaca-se o aprendizado de máquina (*machine-learning*).

Assim, define-se como aprendizado de máquina (*machine-learning*) a ciência de fazer com que os computadores possam aprender e agir de forma análoga aos seres humanos, além de aperfeiçoar autonomamente o seu aprendizado ao longo do tempo, por meio de dados, observações e interações com o mundo real.¹¹² Em suma, o aprendizado de máquina é baseado em algoritmos que conseguem autonomamente melhorar o desempenho com o

¹⁰⁶ Castelluccia Le Métayer (n 29) 4; David Lehr e Paul Ohm, 'Playing with the Data: What Legal Scholars Should Learn About Machine Learning' (2017) 51 Davis University of California 653, 671.

¹⁰⁷ Castelluccia Le Métayer (n 29) 3.

¹⁰⁸ Um exemplo seria a decisão de um funcionário de um banco sobre conceder ou não empréstimo a um consumidor com base em uma análise de crédito feita previamente por um sistema de inteligência artificial.

¹⁰⁹ Filtros de spam feitos por um provedor de email; sugestões de compras por sites de compras online.

¹¹⁰ Castelluccia Le Métayer (n 29) 3.

¹¹¹ *ibid.*

¹¹² Daniel Faggella, 'What is Machine Learning?' (*Emerj*, 26 fev 2020) <<https://emerj.com/ai-glossary-terms/what-is-machine-learning/>> acesso em 23 jun 2022.

passar do tempo, por meio do treinamento com métodos de análise de dados e modelagem analítica.¹¹³ Desse modo, o aprendizado não acontece por programação explícita, ou seja, por meio de um programador que escreve um código com comandos para a resolução de um problema determinado.¹¹⁴ O aprendizado acontece, substancialmente, a partir da experiência obtida pela própria máquina com a análise de grande quantidade de dados, que permite a identificação de correlações (também são utilizados os termos “padrões” ou “relações”) entre as variáveis de um conjunto de dados e a consequente realização de previsões ou classificações sobre algum aspecto do mundo real.¹¹⁵

Importante notar que o conhecimento é adquirido com base na generalização e com dados históricos,¹¹⁶ de modo que “o objetivo do aprendizado de máquina é entender a estrutura dos dados e integrá-los a modelos que possam ser entendidos e usados por todos”.¹¹⁷ Assim, de forma simplificada, esse treinamento com dados históricos resulta na criação de um modelo cujo objetivo é habilitá-lo para, com acurácia, fazer previsões a partir do processamento de novos dados de entrada. Ou seja, ao processar dados diferentes dos que foram utilizados para o treinamento (dados, portanto, não conhecidos até então), espera-se que o modelo apresente uma performance acurada e satisfatória.¹¹⁸

No tocante ao método empregado para que a máquina possa aprender, destacam-se dois principais tipos: aprendizado supervisionado e aprendizado não supervisionado. No primeiro, utiliza-se um conjunto de dados conhecido para treinar ou “supervisionar” o algoritmo, incluindo as entradas e as saídas desejadas.¹¹⁹ Em linhas gerais, nessa modalidade, é fornecido à máquina um conjunto de dados de treinamento, que contém, basicamente, exemplos de situações (entrada) e, para cada exemplo, um resultado rotulado (saída). Cabe ao algoritmo, então, identificar padrões nessas entradas e respectivas saídas para conseguir estabelecer um modelo capaz de, ao processar novos dados, prever o resultado futuro, a partir do aprendizado obtido com a base de dados de treinamento.¹²⁰

¹¹³ Lage (n 106) 70.

¹¹⁴ Dorian Pyle e Cristina San José, ‘An executive’s guide to machine learning’ (*McKinsey Quarterly*, 1 jun 2015) <<https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/an-executives-guide-to-machine-learning>> acesso em 23 jun 2022.

¹¹⁵ Lehr e Ohm (n 110) 671.

¹¹⁶ Faggella (n 117).

¹¹⁷ Lage (n 106) 71.

¹¹⁸ Lehr e Ohm (n 110) 684.

¹¹⁹ Juliana Delua, ‘Supervised vs. Unsupervised Learning: What’s the Difference?’ (*IBM Cloud*, 12 mar 2021) <<https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning>> acesso em 23 jun 2022.

¹²⁰ Lage (n 106) 77.

No aprendizado não supervisionado, não há treinamento prévio com um conjunto de dados selecionado e estruturado. O algoritmo trabalha de forma autônoma para descobrir a estrutura inerente a um conjunto de dados não rotulado.¹²¹ Ou seja, diferentemente do que acontece no aprendizado supervisionado, aqui são conhecidos apenas os dados de entrada, não os respectivos resultados, cabendo ao algoritmo identificar padrões nesses dados disponíveis e classificá-los conforme as características que compartilham.¹²² A principal utilidade é obter *insights* a partir de um grande volume de dados, como, por exemplo, para detectar anomalias (e.g., fraudes em compras *online*), fazer recomendações de produtos em *websites* e analisar imagens médicas.

A escolha entre qual dos métodos utilizar é feita de acordo com o resultado que se espera da máquina. Tende a haver mais precisão com o aprendizado supervisionado, mas, em contrapartida, exige-se maior esforço humano inicial para estruturar o conjunto de dados de treinamento de forma adequada.¹²³ Todavia, o aprendizado não supervisionado também exige envolvimento humano para validar as variáveis de saída, ou seja, os resultados encontrados pelo algoritmo, a fim de que o modelo adquira maior acurácia.

Como subconjunto do aprendizado de máquina, tem-se o aprendizado profundo (*deep learning*), que apresenta maior grau de complexidade. Trata-se, essencialmente, de uma rede neural artificial organizada em três ou mais camadas interconectadas por nós, que tenta simular o comportamento do cérebro humano, especialmente a interconexão entre nossos neurônios, de forma a aprender a partir de um grande volume de dados.¹²⁴ A primeira camada (camada de entrada ou *input layer*) recebe os dados que serão ponderados por pesos, vieses ou limites (*threshold*). Se o resultado obtido em uma camada atingir o valor mínimo definido para a ativação do nó, ele será propagado para a frente nessa rede (fenômeno conhecido como *forward propagation*), ou seja, para a próxima camada oculta¹²⁵, onde será considerado como sinal de entrada para o cálculo, e assim sucessivamente, até chegar à camada de saída (*output layer*), que fornecerá o resultado final da rede.¹²⁶

¹²¹ Delua (n 124).

¹²² Lage (n 106) 78.

¹²³ Delua (n 124).

¹²⁴ ‘What is Deep Learning?’ (*IBM Cloud Education*, 1 mai 2020) < <https://www.ibm.com/cloud/learn/deep-learning> > acesso em 23 jun 2022.

¹²⁵ Eda Kavlakoglu, ‘AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What’s the Difference?’ (*IBM Cloud*, 27 mai 2020) < <https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks> > acesso em 23 jun 2022.

¹²⁶ Fernanda de Carvalho Lage, *Manual de Inteligência Artificial no Direito Brasileiro* (JusPodivm 2021) 88.

A interconexão entre esses nós, que refinam e otimizam a previsão ou categorização nessa sequência,¹²⁷ é estruturada de forma a que o impacto do peso atribuído ao *input* em cada camada afeta o resultado final, gerando um efeito cascata.¹²⁸ Além disso, há um processo conhecido como retropropagação (*back propagation*), que utiliza algoritmos para calcular erros nas previsões e ajustar os pesos e vieses da função em cada camada, para afinar os cálculos dos neurônios individuais no sentido inverso, permitindo o aprendizado da rede e incrementado a precisão do modelo.¹²⁹

O aprendizado de máquina profundo distingue-se do tradicional pelo tipo de dados que utiliza e pela forma como aprende, uma vez que elimina parte do pré-processamento que é tipicamente envolvido no aprendizado de máquina tradicional. Isso porque, neste, a definição das variáveis independentes ou atributos (*features*) é feita mediante intervenção humana, ao passo que, no aprendizado profundo, essa extração é automatizada, o que é de grande utilidade para o processamento de bases de dados não estruturadas e de grande volume.¹³⁰

O aprendizado profundo representa a vanguarda no desenvolvimento da inteligência artificial e contribuiu fortemente para que o uso dela se propagasse. Por exemplo, a habilidade de identificar palavras verbalizadas quase tão bem como um ser humano é algo tão complexo que não poderia ser codificado manualmente por programadores. Assim, o desenvolvimento do aprendizado profundo deu poderes extraordinários aos computadores,¹³¹ viabilizando que realizem tarefas que não poderiam ser ensinadas a eles pelo ser humano via codificação manual.

O objetivo das linhas anteriores nesta seção foi expor, basicamente, o funcionamento algorítmico para que se possa avançar em relação aos desafios que se apresentam em termos de explicabilidade. De fato, a preocupação em compreender as razões determinantes para o resultado fornecido por uma máquina passou a ter maior relevância nos últimos anos, quando os algoritmos foram incorporados em decisões que afetam de forma significativa a vida das pessoas, como naquelas que envolvem concessão de crédito, emprego, seguro e tratamento médico. Assim, o anseio de tornar a inteligência artificial mais compreensível, já muito

¹²⁷ ‘What is Deep Learning?’ (n 129).

¹²⁸ Kavlakoglu (n 130).

¹²⁹ ‘What is Deep Learning?’ (n 129).

¹³⁰ Kavlakoglu (n 130).

¹³¹ Knight (n 39).

relevante na atualidade, tende a crescer nos próximos anos com o avanço do desenvolvimento da tecnologia e com a sua penetração ainda maior no cotidiano humano.¹³² Ocorre que qualquer tecnologia de aprendizado de máquina é inerentemente mais opaca, mesmo para cientistas da computação, do que sistemas codificados manualmente por programadores. Isso porque, como o aprendizado é feito de forma autônoma, ou seja, desenvolvido pelos próprios algoritmos a partir da análise de dados, mesmo os engenheiros que constróem o programa não conseguem explicar completamente o seu comportamento.¹³³ Embora isso não signifique que todas as técnicas de inteligência artificial serão igualmente inexplicáveis, pela sua natureza, o aprendizado profundo de máquina é particularmente uma caixa preta ainda mais escura.¹³⁴ Nesses casos em particular, a explicação que se deseja obter pode ser impossível mesmo para sistemas com uma aparência simples, como *websites* que usam o aprendizado profundo para, por exemplo, apresentar anúncios personalizados. E, quanto maior for a rede neural (*e.g.*, com milhares de neurônios artificiais em cada uma das centenas de camadas), mais difícil se torna a tarefa de tornar compreensível esse processo. Assim, por um lado, a explosão no desenvolvimento da inteligência artificial ocorreu, essencialmente, a partir do aumento da capacidade de processamento dos computadores e da disponibilidade de grandes volumes de dados (fenômeno conhecido como *big data*), que viabilizou o aprimoramento de técnicas de desenvolvimento algorítmico independentes da programação manual e explícita. Os computadores adquiriram, então, capacidades indiscutivelmente úteis para a sociedade, fazendo categorizações e classificações que, além de propiciarem ganho em eficiência na execução de atividades, fornecem *insights* que não seriam facilmente alcançáveis ao olhar puramente humano (*e.g.*, prever o surgimento de doenças a partir da identificação de padrões em imensas quantidades de dados). Por essa razão, diz-se que o aprendizado de máquina apresenta maior utilidade justamente àqueles tipos de problemas para os quais a codificação manual de uma lógica explícita de tomada de decisão não funciona bem – ou seria praticamente impossível.¹³⁵

Por outro lado, esses sistemas, por natureza, apresentam maior opacidade, visto que “aprendem” à sua própria maneira, de forma autônoma, divergindo das normas de aprendizado humano. Então, por exemplo, ao realizar a tarefa de reconhecer números escritos à mão, a máquina não decompõe o dígito em subtarefas que seriam facilmente

¹³² *ibid.*

¹³³ *ibid.*

¹³⁴ *ibid.*

¹³⁵ Burrell (n 94) 6.

inteligíveis para os humanos, tais como a identificação de uma barra horizontal, uma forma oval fechada, uma barra diagonal etc.¹³⁶ E, quanto mais complexas ficam as tarefas destinadas aos computadores, mais misteriosa se revela para os humanos a capacidade perceptiva das máquinas.

Ainda que o acesso à lógica interna do algoritmo (e.g., acessar o peso atribuído a cada variável independente) possa permitir verificar a existência de formas de discriminações codificadas (e.g., constatar que a raça lá consta como atributo), a explicação do porquê de determinado resultado (decisão de classificação) exige um processo de raciocínio interpretativo humano sobre um processo matemático de otimização estatística.¹³⁷ No entanto, inegavelmente, há meios de tornar os algoritmos mais explicáveis, e avançam os estudos que visam a produzir modelos algorítmicos que se preocupam com a explicabilidade.

Nesse sentido, o tópico seguinte, com olhar mais técnico, propõe-se a esclarecer até que ponto os algoritmos podem ser explicáveis, de forma confiável e inteligível.

2.2. O problema da opacidade algorítmica

Cumprе agora tecer considerações mais detalhadas acerca da opacidade. Para melhor compreensão do tema, adota-se a classificação de Burrell, que divide a opacidade em três tipos: intencional, inerente e técnica.¹³⁸

Opacidade intencional é aquela propositalmente desenhada dentro do sistema como uma forma de proteção ao software proprietário, essencialmente para manter uma posição competitiva no mercado e assegurar a proteção do segredo comercial. Além de resguardar os interesses comerciais e prevenir a concorrência, essa opacidade intencional está presente em aplicações de *machine learning* que precisam esconder a lógica interna dos algoritmos computacionais para atingirem a sua finalidade, como é o caso de filtros de *spam* e detectores de fraude.¹³⁹

A **opacidade técnica** é decorrente do elevado grau de especialização e conhecimento técnico necessário para conceber sistemas computacionais. A capacidade de ler e escrever código informático requer uma fluência avançada em linguagens de programação, além de

¹³⁶ *ibid.*

¹³⁷ *ibid* 9.

¹³⁸ *ibid.*

¹³⁹ *ibid* 4.

familiaridade com a engenharia de software, o que torna os algoritmos incompreensíveis para a maioria da população.¹⁴⁰ Mesmo entre os desenvolvedores, a tarefa de compreender códigos escritos por outras pessoas pode ser muito desafiadora se não houver observância a boas práticas de programação (e.g., inserção de comentários destinados a outras pessoas que lerão o código posteriormente à escrita; escrita do código da forma mais simples e direta possível).¹⁴¹

Por sua vez, a **opacidade inerente** aparece como uma função das características internas e da dinâmica operacional dos sistemas algorítmicos mais complexos (integrados), em especial aqueles que são multicomponentes, ou seja, formados por várias partes interligadas.¹⁴² Além das dificuldades para compreender a lógica do sistema em si, no caso do aprendizado de máquina, existe o desafio adicional de compreender como o algoritmo se comporta em ação, o que depende não só da interpretação do código, mas também, substancialmente, de observar como ele, na prática, interage com os dados. Vale lembrar que, como visto no tópico anterior, a lógica interna da tomada de decisão algorítmica se modifica conforme ele “aprende” com a base de dados.¹⁴³ Assim, a interação prática entre o algoritmo e a base de dados é o ponto crucial da opacidade inerente aos sistemas de aprendizado de máquina.¹⁴⁴

Como antídoto aos efeitos adversos da opacidade, a transparência algorítmica, entendida como o seu oposto, vem sendo apontada como instrumento-chave para viabilizar a compreensão de como funcionam esses sistemas, limitando as consequências desfavoráveis da “black box” na sociedade.¹⁴⁵ Apesar do termo ser utilizado de diversas formas na literatura, em geral, a transparência é definida de forma mais restrita, correspondendo à disponibilização do código do sistema, da sua documentação, dos parâmetros utilizados, bem como, quando envolver aprendizado de máquina, da base de dados de treinamento.¹⁴⁶ Todavia, embora seja um requisito importante, a transparência assim definida não garante, por si só, que se possa compreender a lógica do funcionamento algorítmico. Isso porque o

¹⁴⁰ *ibid.*

¹⁴¹ *ibid.*

¹⁴² *ibid.* 5.

¹⁴³ *ibid.*

¹⁴⁴ *ibid.*

¹⁴⁵ Nesse sentido, Frank Pasquale e Danielle Keats Citron, ‘The scored society: due process for automated predictions’ (2014) 89 *Washington Law Review* 1.

¹⁴⁶ Castelluccia e Le Métayer (n 29) 6. No documento “Orientações Éticas para uma IA de confiança”, a transparência é inserida como um dos sete requisitos para a concretização da IA de confiança, enquanto a explicabilidade é considerada princípio base. European Commission, *Orientações éticas para uma IA de confiança* (Publications Office 2019) < <https://data.europa.eu/doi/10.2759/2686> > acesso em 10 jul 2022.

ponto-chave da opacidade inerente é a interação prática entre o algoritmo e os dados, de forma que o mero acesso possibilita enxergar, mas não necessariamente compreender o processo decisório.¹⁴⁷ Por isso, a reivindicação por transparência não deve partir da equivocada premissa de que o conhecimento se revela ao simplesmente enxergar o que há dentro desses sistemas, mas sim da noção de que resulta de interações complexas entre agentes humanos e não-humanos.¹⁴⁸ Em suma, transparência deve envolver não só acessibilidade (no sentido de permitir ver), mas também compreensibilidade (no sentido de possibilitar entender).¹⁴⁹

Por essa razão, há uma segunda corrente na literatura que se refere ao conceito de explicabilidade, o qual, de forma mais abrangente, exigiria a disponibilização de informações adicionais ao sistema de decisão automatizada em si¹⁵⁰ e estaria mais próximo à noção de interpretabilidade.¹⁵¹ Em mineração de dados e *machine-learning*, interpretabilidade é definida como a habilidade de explicar ou dar sentido de forma compreensível para os humanos.¹⁵² Nesse contexto, pode-se afirmar que explicabilidade seria a capacidade dos algoritmos em fornecer razões para as suas previsões.¹⁵³

Em termos técnicos, são crescentes os estudos que visam a assegurar explicabilidade aos algoritmos derivados do aprendizado de máquina,¹⁵⁴ seja por meio do desenho de modelos ao mesmo tempo precisos e explicáveis, seja por meio do desenvolvimento de métodos para compreender o funcionamento de sistemas algorítmicos do tipo “caixa preta”.

De forma ampla, os diversos métodos existentes podem ser divididos em duas categorias: engenharia reversa ou explicabilidade por desenho (*explicability by design*).¹⁵⁵ No primeiro caso, somente é viável interpretar as razões da previsão a partir da análise de *inputs* e *outputs*, de modo que o objetivo é analisar os resultados fornecidos por um modelo existente e perceber as relações entre eles para, então, reconstruir uma explicação.¹⁵⁶ No segundo, a

¹⁴⁷ Mike Ananny e Kate Crawford, ‘Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability’ (2018) 20(3) *New Media and Society* 973.

¹⁴⁸ *ibid* 981.

¹⁴⁹ Brent Daniel Mittelstadt e outros, ‘The ethics of algorithms: Mapping the debate’ (2016) *Big Data & Society* 1.

¹⁵⁰ Castelluccia e Le Métayer (n 29) 26.

¹⁵¹ Riccardo Guidotti e outros, ‘A Survey of methods for explaining black box models’ (2019) 51(5) *ACM Computing Surveys* 6.

¹⁵² *ibid* 5.

¹⁵³ Lehr e Ohm (n 110) 706.

¹⁵⁴ Para uma excelente revisão da literatura nesse sentido, vide Riccardo Guidotti e outros, ‘A Survey of methods for explaining black box models’ (2019) 51(5) *ACM Computing Surveys* 1.

¹⁵⁵ Guidotti (n 156) 11.

¹⁵⁶ *ibid* 19.

tarefa é assegurar a explicabilidade desde a concepção do sistema algorítmico, no seu próprio desenho, mediante o desenvolvimento de um modelo que forneça resultados acurados e suas respectivas explicações.

Os métodos de engenharia reversa são os mais utilizados para solucionar os problemas da ausência de explicabilidade dos sistemas algorítmicos opacos já existentes, como nos casos em que não se tem acesso ao código nem ao conjunto de dados de treinamento do modelo (por exemplo, quando há opacidade intencional). Nessas situações, a compreensão dos resultados ocorre a partir da análise das relações entre eles, que permitirá, então, a depender do método utilizado, prover uma explicação global (sobre a lógica do sistema) e/ou local (sobre o resultado em si).

No primeiro caso, busca-se compreender toda a lógica de funcionamento do modelo algorítmico, bem como o processo que o leva a prever os diferentes resultados. O uso da engenharia reversa para obter a explicabilidade global funciona, basicamente, por meio da criação de um modelo que imita o comportamento da “caixa preta” e que, ao mesmo tempo, é transparente e compreensível por seres humanos.¹⁵⁷ A maior parte dos estudos existentes nesse sentido representa a explicação por meio de uma árvore de decisão ou de um conjunto de regras.¹⁵⁸

Em muitos casos, porém, é difícil obter a explicação global de um sistema algorítmico, de forma que há outros tantos estudos que buscam explicar apenas localmente o comportamento do modelo. Dito de outra forma, o objetivo não é explicar toda a lógica de funcionamento da “caixa preta”, e sim conceber um método capaz de identificar as razões pelas quais o modelo forneceu aquele resultado específico. Essa categoria de métodos tem sido a mais estudada nos últimos anos.¹⁵⁹ Em um exemplo de fácil compreensão, é utilizada para prover algum nível de explicação para o reconhecimento de imagens pelos algoritmos de aprendizado profundo de máquina (*deep learning*), o que é feito pela sinalização das partes da imagem que foram consideradas preponderantes pelo sistema ao realizar a tarefa.

Essas explicações costumam ser apresentadas na forma de texto ou artefatos visuais que provejam compreensão qualitativa da relação entre os *inputs* (por exemplo, palavras em texto, pedaços de uma imagem) e o resultado da previsão feita pelo modelo (*output*).¹⁶⁰ Por exemplo, em um diagnóstico algorítmico de gripe, o resultado positivo poderia ser acrescido

¹⁵⁷ *ibid* 13-14.

¹⁵⁸ *ibid* 13-14.

¹⁵⁹ *ibid* 29.

¹⁶⁰ Ribeiro, Singh e Guestrin (n 93).

da marcação em verde dos fatores preponderantes à classificação, como os sintomas “febre” e “dor de cabeça”; por sua vez, a inexistência do sintoma “cansaço” no paciente seria assinalada em vermelho, pois iria contra o resultado positivo.¹⁶¹ Nessa hipótese, o objetivo geral seria fornecer elementos interpretáveis ao usuário humano (no caso do exemplo, o médico), para que ele possa tomar a sua decisão a partir da explicação fornecida pelo algoritmo – ou seja, avaliar se a explicação fornecida pelo modelo para aquele resultado faz sentido e, por conseguinte, decidir se deve ou não segui-lo.¹⁶²

Ainda que existam trabalhos de construção de métodos agnósticos, ou seja, capazes de fornecer explicações para qualquer modelo, independentemente do tipo de algoritmo (seja, por exemplo, uma rede neural ou uma árvore de decisão) ou de dados (*e.g.*, imagens, textos ou tabelas) utilizados,¹⁶³ muitos métodos são direcionados especificamente para alguns tipos de algoritmos e, portanto, não funcionam em qualquer modelo. Assim, a doutrina alerta que, para fazer uso de métodos que assegurem um maior nível de explicabilidade, é importante fazer boas escolhas nas fases de seleção do modelo e do seu treinamento.¹⁶⁴ Nesse sentido, tendo em vista as fases que envolvem o processo de aprendizado de máquina, o problema da opacidade e da falta de explicabilidade é gerado na escolha do tipo e no desenvolvimento do algoritmo, não na fase de coleta dos dados nem na de especificação das variáveis, diferentemente do que ocorre com o problema do enviesamento.¹⁶⁵

Assim, a explicabilidade global e local pode ser assegurada desde a concepção do sistema algorítmico, no seu próprio desenho (*explainability by design*). Nessa linha, mostra-se possível a redução da opacidade na origem de duas formas: (a) ou pela utilização de uma técnica algorítmica que, pelo seu próprio desenho, assegure inteligibilidade e acurácia suficientes, ou (b) pela criação de um algoritmo que, além dos resultados nominais, gere uma explicação confiável e compreensível para esses resultados.¹⁶⁶

No primeiro caso, a escolha de classes de algoritmos mais simples, cujo processo de otimização é menos complexo, pode ser mais indicada quando se quer assegurar maior grau de interpretabilidade - ainda que se perca um pouco em acurácia nos resultados.¹⁶⁷ Na segunda hipótese, em alguns tipos de algoritmos derivados do aprendizado de máquina, um

¹⁶¹ Ibid.

¹⁶² *ibid.*

¹⁶³ Guidotti (n 156) 18.

¹⁶⁴ Lehr e Ohm (n 110) 708.

¹⁶⁵ *ibid* 668.

¹⁶⁶ Castelluccia e Le Métayer (n 29) 51.

¹⁶⁷ Lehr e Ohm (n 110) 692.

analista de dados pode gerar representações gráficas que mostrem quão importantes foram as diferentes variáveis de entrada na geração do resultado, e como mudanças nos seus valores influenciam esse resultado.¹⁶⁸

Assim, considerando os estudos existentes, pode-se afirmar que, tecnicamente, há meios de tornar as decisões algorítmicas mais compreensíveis aos seres humanos, e avançam os trabalhos que pretendem produzir modelos preocupados com a explicabilidade. Juridicamente, o direito à explicação será aprofundado no tópico a seguir, mas, por ora, está demonstrado que, em termos técnicos, as escolhas humanas que envolvem o processo de desenvolvimento de um modelo podem estar pautadas pela interpretabilidade. E, embora esse não seja o elemento mais importante em todos os casos – já que o valor preponderante para decisões em outros níveis pode ser a precisão –, para as decisões da Administração Pública que afetam a esfera de direitos dos cidadãos, com certeza, o é.

A corroborar a afirmação, vale ressaltar que a EBIA recomenda, no tocante ao uso da inteligência artificial no setor público, que, em licitações e contratos administrativos voltados à aquisição de produtos e serviços de Inteligência Artificial, sejam considerados critérios voltados não apenas à eficiência técnica, mas também relativos à incorporação de princípios éticos relacionados à transparência, à equidade e à não-discriminação.¹⁶⁹

2.3. O direito à explicação no RGPD e na LGPD

A preocupação com a opacidade de sistemas de decisões algorítmicas resultou em recentes normativas sobre o tema, em especial, na matéria de proteção de dados pessoais. Considerando que o RGPD foi precursor e inspirou, em grande medida a legislação brasileira, mais recente, sobre o tema, optou-se por iniciar a análise a partir do Regulamento europeu para trazer subsídios para a interpretação da legislação brasileira, cuja doutrina acerca do tema é menos extensa.

No âmbito do RGPD, é possível encontrar bases para o direito à explicação em três diferentes conjuntos de normas: direito à informação (Artigos 13 e 14, comentados pelos Considerandos 60-62); direito de acesso (Artigo 15, comentado pelo Considerando 63); e nas salvaguardas contra o uso de decisões automatizadas previstas pelo Artigo 22(3), comentadas pelo Considerando 71.

¹⁶⁸ *ibid.*

¹⁶⁹ EBIA (n 3) 45.

Nessa linha, os Artigos 13(2)(f), 14(2)(g) e 15(1)(h) garantem o direito do titular dos dados pessoais a, respectivamente, receber informações e obter a acesso no que diz respeito à existência de decisões automatizadas, incluindo “informações úteis relativas à lógica subjacente, bem como a importância e as consequências previstas de tal tratamento para o titular dos dados”.¹⁷⁰

Por sua vez, no Artigo 22, como regra, estabelece-se a proibição de decisões tomadas exclusivamente com base no tratamento automatizado que produzam efeitos na esfera jurídica do titular dos dados ou que o afetem significativamente de forma similar.¹⁷¹ Todavia, há exceções a essa regra, dentre as quais se destaca se a decisão for “autorizada pelo direito da União ou do Estado-Membro a que o responsável pelo tratamento estiver sujeito, e na qual estejam igualmente previstas medidas adequadas para salvaguardar os direitos e liberdades e os legítimos interesses do titular dos dados”.

De qualquer sorte, entre as salvaguardas em face dessas decisões tomadas exclusivamente com base no tratamento automatizado, consta, no Artigo 22(3), expressa menção ao direito de, no mínimo, “obter intervenção humana por parte do responsável, manifestar seu ponto de vista e contestar a decisão”. Esse dispositivo do Regulamento relaciona-se com o Considerando 71, cujo texto expressamente acrescenta, dentre as garantias adequadas ao titular, o direito de “obter uma explicação sobre a decisão tomada na sequência dessa avaliação e de contestar a decisão”.

Especialmente porque a referência a um direito de obter explicação consta apenas de forma expressa no preâmbulo do Regulamento, cujas disposições têm relevância interpretativa, mas não caráter mandatório, somada às dificuldades práticas de explicar a decisão algorítmica e à necessidade de proteger os segredos comercial e industrial,¹⁷² há um amplo debate a respeito da existência e da extensão do direito à explicação em face das decisões individuais automatizadas com base no RGPD.

Nessa linha, parte da doutrina entende que o Regulamento garante apenas um direito à informação sobre a lógica envolvida (funcionalidade do sistema), sobre o significado e as pretendidas consequências do tratamento de dados, mas não um direito amplo de explicação

¹⁷⁰ Cabe salientar que, na versão em inglês, utilizou-se a expressão “meaningful information” e, em espanhol, “información significativa”.

¹⁷¹ Article 29 Working Party, ‘Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679’ WP251 (rev 6 February 2018).

¹⁷² O Considerando 63 do RGPD esclarece que o direito de acesso “não deverá prejudicar os direitos ou as liberdades de terceiros, incluindo o segredo comercial ou a propriedade intelectual e, particularmente, o direito de autor que protege o *software*.”

que envolva revelar as razões específicas e as circunstâncias individuais de uma determinada decisão automatizada depois de produzida (*ex post*).¹⁷³ Parte da literatura entende que o RGPD, tal como está, garante apenas um direito à informação *ex ante* que, por si só e na sua literalidade, é muito limitado e não atende ao que se espera em termos de um direito geral à explicação das circunstâncias particulares das decisões automatizadas.¹⁷⁴

De fato, na literalidade, o escopo de aplicação do Artigo 22(3) é estrito. As salvaguardas aplicam-se para as decisões tomadas “exclusivamente” com base no tratamento automatizado, que produzam efeitos na esfera jurídica do titular dos dados ou que o afetem significativamente de forma similar. A doutrina europeia também debate qual o grau de envolvimento humano que seria suficiente para considerar que a decisão não é exclusivamente automatizada e, assim, excluir a aplicação do Artigo 22(3). Sobre o ponto, o Grupo de Trabalho do Artigo 29 para a Proteção de Dados já esclareceu que essa intervenção humana precisa ser relevante e executada por alguém com autoridade e competência para alterar a decisão.¹⁷⁵

Por outro lado, há quem entenda que a discussão sobre a existência de um direito à explicação não está adstrita à análise do Artigo 22 do RGPD, de modo que esse direito decorreria de uma interpretação conjunta de outros dispositivos do Regulamento (em especial dos Artigos 13 e 15), bem como do próprio princípio da transparência.¹⁷⁶ Nesse sentido, sustentam que as salvaguardas do Artigo 22 reforçam o direito à explicação decorrente da interpretação holística da norma.¹⁷⁷ Além disso, a explicação das razões da decisão específica constitui pressuposto para o exercício de outros direitos previstos expressamente no RGPD, como o direito de oposição.

Nessa linha, há, ainda, quem defenda, a partir de uma interpretação sistemática dos Artigos 13-15 e 22 do RGPD, a existência de um direito à legibilidade da arquitetura e da

¹⁷³ Sandra Wachter, Brent Mittelstadt e Luciano Floridi, ‘Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation’ (2017) 7(2) *International Data Privacy Law* 76.

¹⁷⁴ Para bons argumentos em sentido contrário, defendendo a existência de um direito de explicação que inclua as circunstâncias específicas da decisão tanto com base no Artigo 15 quanto com base no Artigo 22 do RGPD, em uma interpretação em consonância com os princípios que norteiam o direito à proteção de dados pessoais, vide Tiago Sérgio Cabral, ‘AI and the Right to Explanation: Three Legal Bases under the GDPR’ in Dara Hallinan, Ronald Leenes e Paul De Hert (eds), *Data Protection and Privacy: Data Protection and Artificial Intelligence* (Hart Publishing 2021).

¹⁷⁵ WP251(n 176) 21.

¹⁷⁶ Margot E Kaminski, ‘The Right to Explanation, Explained’ (2019) 34 *Berkeley Technology Law Journal* 189; Andrew D Selbst e Julia Powles, ‘Meaningful information and the right to explanation’ (2017) 7(4) *International Data Privacy Law* 233.

¹⁷⁷ *ibid.*

implementação do processamento algorítmico, que garante que os indivíduos possam, de forma autônoma, compreender a funcionalidade e o impacto contextual dos algoritmos nas decisões tomadas a seu respeito.¹⁷⁸

No Brasil, a nova Lei Geral de Proteção de Dados Pessoais, muito inspirada pelo Regulamento europeu, dedicou um capítulo inteiro ao tratamento de dados pessoais pelo poder público. Em síntese, a legislação brasileira estabelece que o tratamento de dados pessoais pela Administração Pública “deverá ser realizado para o atendimento de sua finalidade pública, na persecução do interesse público, com o objetivo de executar as competências legais ou cumprir as atribuições legais do serviço público”.¹⁷⁹ Como condição para o tratamento de dados pessoais pelas pessoas jurídicas de direito público está a disponibilização de informações “claras e atualizadas” sobre os “procedimentos” e as “práticas” utilizadas para a execução dessas atividades, preferencialmente em nos respectivos sítios eletrônicos.¹⁸⁰

Diferentemente do RGPD, a nova LGPD brasileira, no seu Artigo 20, apesar de ter previsto que “o titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses”, não garantiu que essa revisão seja feita mediante intervenção humana.¹⁸¹ O trecho da disposição que previa que a revisão fosse feita por pessoa natural, inicialmente previsto no projeto de lei aprovado, sofreu veto pelo Presidente da República, que foi mantido pelo Congresso Nacional.¹⁸² O tema da intervenção humana será retomado mais adiante neste tópico.

Como se vê, ao comparar-se o Artigo 22 do RGPD com o Artigo 20 da LGPD, percebe-se, primeiramente, que esta não previu, como regra, a proibição de decisões exclusivamente automatizadas que produzam efeitos na esfera jurídica do titular dos dados ou que o afetem significativamente de forma similar. Seguindo um caminho diferente, a LGPD permite essas

¹⁷⁸ Gianclaudio Malgieri e Giovanni Comandé, ‘Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation’ (2017) 7(4) International Data Privacy Law 243.

¹⁷⁹ LGPD, Artigo 23, *caput*.

¹⁸⁰ LGPD, Artigo 23, I.

¹⁸¹ Art. 20. O titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, incluídas as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade. (Redação dada pela Lei nº 13.853, de 2019).

¹⁸² O argumento para o veto é o de que o dispositivo contrariaria o interesse público, já que inviabilizaria os atuais planos de negócios de muitas empresas, impactaria na análise de risco de crédito e de novos modelos de negócios instituições financeiras, “gerando efeito negativo na oferta de crédito aos consumidores, tanto no que diz respeito à qualidade das garantias, ao volume de crédito contratado e à composição de preços, com reflexos, ainda, nos índices de inflação e na condução da política monetária.” Inteiro teor da mensagem de veto disponível em <http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2019/Msg/VEP/VEP-288.htm> acesso em 07 set 2022.

decisões automatizadas como regra e, assim, assegura ao titular dos dados o direito de solicitar a sua revisão. Para tanto, estabelece que “o controlador deverá fornecer, sempre que solicitadas, informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada, observados os segredos comercial e industrial.”

Por informação “clara” entende-se aquela fornecida de modo acessível e que favoreça a compreensão por todos os indivíduos, sejam *experts* ou leigos. A utilização do adjetivo “adequada” denota que a informação deve ser apropriada para atingir uma finalidade, que é viabilizar o exercício de outros direitos pelo titular dos dados, em especial, o direito à revisão da decisão automatizada.¹⁸³ Por sua vez, os termos “critérios” e “procedimentos” trazem maior especificidade em relação aos empregados no RGPD, aparentando albergar, respectivamente, os parâmetros utilizados para a tomada de decisão e o modo como o algoritmo opera.¹⁸⁴

Aqui parece não haver dúvidas de que há um direito de explicação *ex post*, sobre a decisão específica produzida, devendo ser informados os critérios e os procedimentos utilizados para a decisão automatizada. Caso essas informações não sejam fornecidas pelo controlador com fundamento na observância do segredo comercial e industrial, a lei permite à autoridade nacional de proteção de dados realizar auditoria “para verificação de aspectos discriminatórios em tratamento automatizado de dados pessoais”.¹⁸⁵

O dispositivo assegura a existência de um procedimento para que o titular possa comunicar à ANPD que teve negado o acesso às informações necessárias para permitir o exercício do seu direito de revisão da decisão automatizada, cabendo àquela decidir sobre a realização da auditoria.¹⁸⁶ Não se trata de direito subjetivo do titular dos dados, já que a realização de auditoria ou a determinação de sua realização é uma atribuição da autoridade nacional destinada a fiscalizar e aplicar sanções quando verificar o tratamento de dados pessoais em descumprimento à legislação.¹⁸⁷

¹⁸³ Eduardo Magrani, Christian Perrone e Carlos Affonso Souza, ‘O direito à explicação entre a experiência europeia e a sua positivação na LGPD’ in Danilo Doneda e outros (coord), *Tratado de proteção de dados pessoais* (Forense 2021) 265.

¹⁸⁴ *ibid.*

¹⁸⁵ Apesar de conveniente a previsão, cabe a crítica de que sua aplicação prática exigiria que a ANPD tivesse capacidade de execução rápida, o que dependeria de um grande número de pessoal especializado ou da execução desse trabalho em parceria com outras instituições (e.g., auditorias independentes). Além disso, como já demonstrado ao longo desse trabalho, o aspecto discriminatório é apenas uma das formas de falibilidade da decisão algorítmica.

¹⁸⁶ Magrani, Perrone e Souza (n 189) 268.

¹⁸⁷ LGPD, Artigo 55-J, IV e XVI.

Assim, apesar da limitação de escopo do Artigo 20 – também dirigido às decisões “unicamente” automatizadas tal como no RGPD –, é possível extrair o direito à explicação de uma interpretação sistemática da LGPD brasileira, uma vez que, assim como o RGPD, ela está fundada em um conjunto de normas que assegura robusta proteção aos direitos dos indivíduos. Nesse sentido, incluem-se, entre seus princípios (Artigo 6º), o acesso à informação, a transparência, a prevenção de danos, a não discriminação, bem como a responsabilização e a prestação de contas.

Dessa forma, o direito à revisão previsto no Artigo 20 não é a única garantia ou direito que se aplica a uma decisão automatizada que envolva o tratamento de dados pessoais. O direito à explicação decorre do próprio princípio da transparência, do direito de acesso à informação e é, no mínimo, pressuposto para o exercício de outros direitos, em particular, o direito a requerer a revisão dessas decisões.¹⁸⁸ Por conseguinte, embora com escopo um pouco diverso do Artigo 22 do RGPD, o Artigo 20 da LGPD é entendido como a consagração da diretriz da explicabilidade e do vinculado princípio da motivação decisória algorítmica no direito brasileiro.¹⁸⁹

Na sequência, cabem tecer algumas considerações específicas sobre a intervenção humana e sobre as decisões semiautomatizadas.

Assegurar a intervenção humana na revisão das decisões automatizadas é importante e, inclusive, recomendável, pois reforça a confiança dos usuários no processo, tornando-o mais plural, acessível e transparente.¹⁹⁰ Nesse sentido, a já referida Estratégia Brasileira de Inteligência Artificial insere como ações estratégicas 1) a criação de parâmetros sobre a intervenção humana em contextos em que o resultado de uma decisão automatizada implica um “alto risco de dano” para o indivíduo, além do 2) incentivo à exploração e ao desenvolvimento de mecanismos de revisão apropriados em diferentes contextos de utilização de IA por organizações privadas e por órgãos públicos.¹⁹¹

No entanto, se os sistemas algorítmicos não forem explicáveis em si, assegurar, formalmente, a intervenção humana pode ser insuficiente a depender do contexto.¹⁹² Afinal, pouco adianta manter o ser humano no processo de tomada de decisão se o seu papel for

¹⁸⁸ Magrani, Perrone e Souza (n 189) 262.

¹⁸⁹ Juarez Freitas e Thomas Bellini Freitas, *Direito e Inteligência Artificial: em defesa do humano* (Forum 2021) 103.

¹⁹⁰ Magrani, Perrone e Souza (n 189) 267.

¹⁹¹ EBIA (n 3) 23.

¹⁹² Para um erro de reconhecimento facial, por exemplo, a intervenção humana pode ser suficiente para permitir o ingresso no país de alguém equivocadamente não reconhecido pela máquina.

receber um resultado incompreensível gerado pelo modelo. Nesse caso, ele pouco poderá fazer além de validar ou dar uma motivação cosmética para esse resultado, o que poderia dar uma enganosa sensação de conforto. No entanto, talvez o maior dos desconfortos seja a inversão de papéis, em que o trabalho do ser humano se transforma em mero “carimbador” do resultado indecifrável produzido pela máquina.¹⁹³ E essa preocupação também se aplica, com igual importância, às decisões individualizadas semiautomatizadas. Se o ser humano não puder compreender as razões que levaram o algoritmo a uma determinada conclusão, como poderá verdadeiramente decidir se acata ou não esse resultado?

A explicabilidade dos sistemas algorítmicos direciona-se não só ao indivíduo que é afetado pela decisão, mas também ao agente que a toma com suporte no resultado fornecido pelo modelo (caso da decisão semiautomatizada). Esse ser humano, em primeiro lugar, deve compreender as razões para aquele resultado. Ademais, e não menos importante, a Administração Pública também precisa ter acesso e compreensibilidade sobre o sistema algorítmico, para que possa promover ajustes, controlá-lo e aprimorá-lo por meio de intervenções humanas não só ao final do processo, diante de um resultado específico, mas, também e substancialmente, nas fases anteriores de desenvolvimento. Ainda, enquanto controladora, a Administração Pública deverá estar pronta para cumprir os seus deveres relacionados à *accountability*.¹⁹⁴

Nesse contexto, o capítulo a seguir iniciará com a análise de um caso prático que bem demonstra a importância de assegurar a explicabilidade como qualificadora da motivação da decisão algorítmica. Na sequência, pretende-se estabelecer os novos contornos do direito à motivação das decisões administrativas na era digital.

3. Os novos contornos do direito à motivação na Administração Pública Digital

3.1. O cidadão, o algoritmo e o contra-argumento

Propõe-se nesta seção a examinar um caso prático para auxiliar a compreensão do que foi escrito nas linhas acima. Apesar de derivar do Poder Judiciário, entende-se que a análise é apta a proporcionar interessantes reflexões que também se aplicam ao processo de tomada de decisão inerente ao ato administrativo individual. O exemplo é útil para demonstrar, de forma mais concreta, as dificuldades enfrentadas para que seja assegurada, ao jurisdicionado

¹⁹³ Lehr e Ohm (n 110)716.

¹⁹⁴ LGPD, Artigos 29 e 32.

(e ao cidadão), o direito de conhecer os fundamentos e de apresentar contra-argumentos a uma decisão algorítmica.

O software COMPAS, utilizado em algumas jurisdições dos EUA, consiste em uma ferramenta de análise de risco de reincidência criminal desenvolvida por uma empresa privada, denominada Northpointe, Inc. (renomeada para Equivant). Em termos gerais, o software utiliza a análise de dados para estabelecer uma pontuação de risco para cada acusado, prevendo, então, a probabilidade de reincidência baseada na comparação das informações com os dados de um grupo de pessoas com histórico similar.¹⁹⁵ Os algoritmos do COMPAS usam como *input* as respostas para um questionário de 137 itens – parte das respostas oferecida pelo próprio acusado, parte delas extraídas do histórico criminal –, estabelecendo uma pontuação de 1 a 10, dividida em três categorias: pontuações de 1 a 4 são consideradas como “baixo risco” de reincidência criminal; de 5 a 7, “médio risco”; de 8 a 10, “alto risco”.¹⁹⁶

Para além do intrigante debate a respeito do viés discriminatório dos resultados fornecidos por esse sistema,¹⁹⁷ após estudos que concluíram que ele classificaria de forma desproporcional os acusados negros como tendo um risco mais elevado de reincidência, interessa para o fim deste trabalho a decisão proferida pela Suprema Corte de Wisconsin em *State of Wisconsin v. Eric Loomis*.¹⁹⁸

No caso, Loomis recorreu ao referido tribunal contra a decisão da corte regional, que o condenou utilizando, entre outros elementos, o relatório fornecido pelo software COMPAS, que o classificara na categoria de altíssimo risco de reincidência. Em síntese, Loomis alegou que a decisão da corte regional, ao utilizar o relatório produzido pelo software COMPAS, infringira o seu direito ao devido processo legal por três razões: 1) por violar o direito de ser julgado com base em informações corretas, em parte porque a natureza proprietária do software COMPAS impossibilita a avaliação da acurácia das informações consideradas para a produção do relatório de análise de risco de reincidência; 2) por violar o direito a uma

¹⁹⁵ *State of Wisconsin v Eric L Loomis* (n 10) para 15.

¹⁹⁶ O questionário completo está disponível em <<https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html>> acesso em 24 de abril de 2022.

¹⁹⁷ Julia Angwin e outros, 'Machine Bias' (*ProPublica*, 23 de maio de 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> acesso em 07 de setembro de 2022; Jeff Larson e outros, 'How We Analyzed the COMPAS Recidivism Algorithm' (*ProPublica*, 23 de maio de 2016) <<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>> acesso em 07 de setembro de 2022.

¹⁹⁸ *State of Wisconsin v Eric L Loomis* (n 10).

decisão individualizada e 3) por indevidamente utilizar o gênero como atributo nas avaliações.¹⁹⁹

Ao apreciar o caso, a Suprema Corte de Wisconsin afirmou a legalidade do uso software COMPAS para embasar decisões judiciais no âmbito criminal.²⁰⁰ Quanto às questões alegadas por Loomis, utilizou dois principais argumentos para rebatê-las: 1) que o relatório de análise de risco do COMPAS havia sido fornecido para as duas partes; e 2) que o acusado teve a possibilidade de determinar a acurácia do resultado da sua análise de risco uma vez que poderia ter corrigido as suas próprias respostas fornecidas no questionário. Sobre a utilização do gênero como variável no algoritmo, o tribunal entendeu que a diferenciação entre homens e mulheres é necessária para aumentar a precisão das previsões fornecidas pelo sistema. Todavia, referiu que o acusado não se desincumbiu do ônus de demonstrar que a decisão recorrida efetivamente considerou o gênero como um fator para o julgamento.

Esse exemplo revela a profundidade do problema da opacidade, visto que a complexidade da decisão algorítmica parece ter sido subestimada no julgamento. A uma porque a mera disponibilização do relatório produzido pelo software não garante que o acusado possa entender por que a ferramenta o incluiu em determinada categoria (ou seja, o porquê e o como do *output*), já que nenhuma informação sobre o funcionamento do sistema foi disponibilizada (no caso, o algoritmo é protegido por propriedade intelectual). Desconhecesse, no caso, qual tipo de algoritmo foi escolhido (*e.g.*, árvore de decisão, rede neural); também não se tem conhecimento mínimo acerca de como as variáveis influenciam nos resultados (*e.g.*, se o fato de ser casado importa, se o bairro onde reside importa, se a origem étnica importa).

A duas porque, ainda que o acusado pudesse modificar algumas de suas respostas ao questionário que serve de *input*, isso não garante por si só que o resultado (*output*) seja justo devido à complexidade das correlações entre bases de dados. Além disso, os dados de *input* não são sopesados igualmente no processamento algorítmico, ou seja, há itens que recebem mais peso que outros – e esse sopesamento é desconhecido porque integra o próprio desenho do algoritmo, protegido pela propriedade intelectual nesse caso. Nesse sentido, sem

¹⁹⁹ *State of Wisconsin v Eric L Loomis* (n 10) para 34.

²⁰⁰ Apesar de ter reconhecido a legalidade do uso do software COMPAS para embasar decisões judiciais no âmbito criminal, constaram recomendações expressas na decisão da Suprema Corte de Wisconsin sobre os cuidados que os julgadores devem ter ao fazer uso dos relatórios fornecidos pela ferramenta. Em especial, foram ressaltadas a impossibilidade de embasar a decisão apenas nesse elemento e a proibição de que o relatório seja utilizado para determinação ou não do encarceramento e da extensão da pena.

indicação de como as respostas foram avaliadas, não é possível realmente desafiar o *score* preditivo apenas com a revisão das respostas ao questionário.²⁰¹

A três, no tocante à questão do gênero, como visto, a Corte referiu que o acusado não comprovou que a decisão recorrida realmente levara em conta o gênero como uma variável no julgamento. No entanto, cabe a pergunta: sem acesso mínimo à explicabilidade algorítmica, como Loomis poderia ter comprovado a influência do gênero sobre o resultado? Demonstra-se, então, a possibilidade prática de uma decisão judicial declinar motivos (na hipótese, houve extensa motivação da decisão pelo órgão julgador) que, na realidade, são insuficientes para que a pessoa afetada possa efetivamente desafiá-la no que tange ao aspecto algorítmico. O exemplo é útil para perceber que o problema não afeta apenas as decisões integralmente automatizadas, mas também tem importância para as decisões semiautomatizadas, como a do caso. Isso porque a tradicional motivação, concebida para o controle da falibilidade humana, não assegura que o indivíduo afetado possa contra-argumentar a possível falha algorítmica existente na decisão. E é justamente essa a importância fundamental de garantir também a explicabilidade, como qualificadora da motivação da decisão algorítmica.

Ao fim e ao cabo, Loomis, como qualquer outra pessoa significativamente afetada por uma decisão (seja ela judicial ou administrativa), quer, essencialmente, poder exercer o controle sobre o julgamento em dois principais aspectos: 1) se as informações levadas em conta para a decisão individualizada são corretas e 2) se o tratamento dado foi isonômico, em comparação com outros indivíduos em situação análoga.

Em relação ao primeiro ponto (informações individualizadas e corretas), a questão de fundo nos argumentos de Loomis resume-se à necessidade de conhecer as informações que foram consideradas para o cálculo da probabilidade de reincidência pelo sistema, a fim de permitir o controle sobre o seu próprio julgamento. Como visto, a finalidade do uso do sistema COMPAS nos Estados Unidos é a promoção da segurança pública e a redução da reincidência criminal; portanto, o escopo algorítmico é obter uma finalidade específica: a previsão mais precisa acerca da probabilidade de reincidência criminal.

Tendo em vista essa finalidade, é razoável pensar que assegurar a explicabilidade ao COMPAS exigiria conhecer minimamente as variáveis consideradas pelo modelo e o peso atribuído a elas. No entanto, a disponibilização dessas informações revelaria a teoria de justiça incorporada pelo sistema, ou seja, como o objetivo abstrato (prever e reduzir a

²⁰¹ Washington (n 85) 146.

probabilidade de reincidência criminal) foi transformado em um resultado mensurável.²⁰² Estaria descortinada uma discussão importantíssima – e sem resposta em nível puramente humano – sobre os critérios considerados preponderantes para prever a probabilidade de uma pessoa reincidir criminalmente.

Com efeito, em muitos aspectos, a legislação estabelece parâmetros, mas deixa propositalmente um espaço de discricionariedade a ser preenchido pelo juiz no caso concreto, segundo critérios subjetivos que reservam lugar à intuição. Nessa linha, no caso *Loomis*, aquela Suprema Corte de Wisconsin reconhece que várias determinações ao longo da justiça criminal são reservadas à intuição do julgador humano e, portanto, não consideram “fatos testados” de nenhuma espécie (e.g., em quais casos determinar tratamento ao acusado, qual o tipo e o tempo de duração da supervisão).²⁰³

Mesmo na decisão integralmente humana que decreta a prisão preventiva – que é das que afetam mais gravemente a vida do cidadão, visto que o leva ao encarceramento precoce, antes da conclusão de todo o rito do processo penal –, o julgador humano não é obrigado a enfrentar absolutamente todos os argumentos trazidos pelas partes, nem precisa declinar todos os critérios utilizados. Para essa decisão, embora o ordenamento jurídico brasileiro exija uma motivação robusta e concreta, os critérios são propositalmente abertos (e.g., “garantia da ordem pública”, “conveniência da instrução criminal”, “perigo gerado pelo estado de liberdade do imputado”), de modo que a suficiência da motivação, cuja análise é subjetiva, não exige uma explicação do tipo matemática de todos os fatores que foram efetivamente considerados pelo juiz com indicação da respectiva preponderância.²⁰⁴

Nesse sentido, cabe reconhecer que a motivação de uma decisão humana, seja judicial ou administrativa, manifesta um discurso justificativo com o qual o agente intenciona persuadir o particular interessado ou a coletividade de que a decisão tomada atende à legalidade.²⁰⁵ Assim, não é revelado todo o *iter* lógico ou psicológico percorrido pelo agente humano ao longo do processo decisório. Alguns elementos do raciocínio são expostos na motivação, mas ela não arrola, por exemplo, todas as hipóteses rejeitadas na formação da decisão nem elementos inconscientes que participaram daquele processo.²⁰⁶

²⁰² Para uma análise profunda sobre esse tema, vide Lehr e Ohm (n 110) 672.

²⁰³ *State of Wisconsin v Eric L Loomis* (n 10) para 40.

²⁰⁴ Código de Processo Penal brasileiro, Decreto-Lei nº 3.689, de 3 de outubro de 1941, Artigos 312 a 315.

²⁰⁵ Cintra (n 8) 108-109.

²⁰⁶ *ibid.*

Assim, especialmente para os casos em que há um espaço destinado à discricionariedade (ou, até mesmo, à intuição), um questionamento pertinente seria o seguinte: ao exigir do sistema algorítmico que arrole todos os fatores considerados e identifique o respectivo nível de preponderância em cada decisão, não se estaria impondo um nível de explicação (e motivação) maior da máquina do que o frequentemente fornecido pelo ser humano nas decisões judiciais ou administrativas? Parece que, em um sistema de decisão algorítmica, se todos os critérios e o nível de influência de cada um deles fosse revelado, paradoxalmente, passaria a ser mais fácil auditar a decisão do algoritmo do que a decisão humana motivada. Cabe retomar, então, o segundo ponto acima referido (isonomia no tratamento). O aspecto do tratamento isonômico está no centro do debate sobre as decisões automatizadas, especialmente em relação ao dever de não discriminação. Ocorre que, mesmo para os julgamentos integralmente humanos, não se tem, ao menos no Brasil, uma mensuração do grau de isonomia observado pelas decisões em casos semelhantes, especialmente no que tange a aspectos que a lei confere margem de discricionariedade ao julgador, como, *e.g.*, para a fixação do *quantum* de pena ou do valor de uma multa ou indenização.

Diante dessa realidade, o uso de dados e o processo de digitalização poderiam auxiliar significativamente na redução do ruído, que é a variabilidade inexplicável no julgamento humano, que conduz, portanto, a tratamentos não isonômicos, frequentemente não monitorados.²⁰⁷ Sob essa perspectiva, poder-se-ia cogitar que auditar o julgamento humano também seria uma medida pertinente, assim como auditar o julgamento algorítmico. A afirmação é contra-intuitiva e talvez até gere algum desconforto no leitor, mas, pela relevância, sugere-se, desde logo, seja objeto de futuros estudos.

3.2. As decisões algorítmicas e a exigência de explicabilidade no contexto normativo brasileiro

No ordenamento jurídico brasileiro, ainda não há uma legislação com abrangência nacional que regule a utilização de sistemas de decisões algorítmicas no setor público.²⁰⁸ O primeiro documento editado nesse sentido, de caráter infralegal, foi a Resolução nº 332/2020 do

²⁰⁷ Kahneman, Sibony e Sunstein (n 38) 40.

²⁰⁸ Está em trâmite perante o Congresso Nacional projeto de lei cujo objetivo é estabelecer o marco legal da inteligência artificial no Brasil (PL 21/2020). O texto atual do projeto vem sofrendo muitas críticas de especialistas pela sua insuficiência e pela ausência de participação pública na sua construção. Para uma análise geral sobre o tema, vide: Giovanni Santa Rosa, 'Lei da inteligência artificial no Brasil: entenda o projeto aprovado na Câmara' (*Tecnoblog* 2022) <<https://tecnoblog.net/especiais/lei-da-inteligencia-artificial-no-brasil-entenda-o-projeto-aprovado-na-camara/>> acesso em 13 jul 2022.

Conselho Nacional de Justiça (CNJ), dirigida ao uso da inteligência artificial no Poder Judiciário.²⁰⁹ A elaboração da normativa teve como razão de ser a proliferação de iniciativas que envolvem o uso de inteligência artificial nos diferentes tribunais do País – há 111 projetos desenvolvidos ou em desenvolvimento²¹⁰ – e a consequente necessidade de regulação do ecossistema já existente, que abrange a maioria dos tribunais brasileiros. O documento, declaradamente, inspira-se na Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seus ambientes, tendo em vista a ausência, até então, de normas nacionais específicas quanto à governança e aos parâmetros éticos para o desenvolvimento e uso da inteligência artificial.²¹¹

A leitura do documento deixa claro que se permite a utilização de “sistemas inteligentes” no Poder Judiciário brasileiro para a produção de decisões semiautomatizadas, reconhecendo que o uso da inteligência artificial pode oferecer agilidade e coerência no processo de tomada de decisão. O texto salienta que a proposta de decisão gerada pelo sistema algorítmico tem sempre caráter não vinculante, de forma que é submetida à análise da autoridade competente (jugador humano). Não há espaço, portanto, para decisões judiciais totalmente automatizadas que envolvam o uso de inteligência artificial, já que, segundo a regulação aplicável, o papel dos sistemas de decisões judiciais algorítmicas restringe-se a apoiar o juiz humano no processo de julgamento (como, por exemplo, pelo fornecimento de proposta de texto de decisão ao juiz com base na jurisprudência; pela análise semântica de peças processuais para triagem de processos²¹²).

²⁰⁹ CNJ, Resolução nº 332/2020 (2020) <<https://atos.cnj.jus.br/atos/detalhar/3429>> acesso em 7 jun 2022.

²¹⁰ Vanessa Maeji, ‘Justiça 4.0: Inteligência Artificial está presente na maioria dos tribunais brasileiros’ (*Conselho Nacional de Justiça* 2022) <<https://www.cnj.jus.br/justica-4-0-inteligencia-artificial-esta-presente-na-maioria-dos-tribunais-brasileiros/>> acesso em 10 set 2022. Para um painel detalhado sobre as iniciativas existentes: <https://paineisanalytics.cnj.jus.br/single/?appid=9e4f18ac-e253-4893-8ca1-b81d8af59ff6&sheet=b8267e5a-1f1f-41a7-90ff-d7a2f4ed34ea&lang=pt-BR&theme=IA_PJ&opt=ctxmenu,cursel&select=language,BR> acesso em 13 jul 2022.

²¹¹ Comissão Europeia para a Eficácia da Justiça, Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu ambiente (2018) <<https://rm.coe.int/carta-etica-traduzida-para-portugues-revista/168093b7e0>> acesso em 10 jun 2020.

²¹² Essa é uma das funcionalidades do sistema Sócrates, implementado no Superior Tribunal de Justiça. “O modelo aplicado seleciona para exibição, numa “nuvem de palavras”, as trinta palavras mais frequentes da peça, objetivando fornecer ao usuário uma ideia do conteúdo do recurso, antes mesmo da leitura do seu inteiro teor. Esta “nuvem de palavras” dispõe os termos em tamanhos diferentes de fonte; quanto maior a fonte, maior a frequência da palavra na petição de recurso especial.” Guilherme Veiga Chaves e Elizabeth Veiga, ‘A inteligência artificial na formação dos precedentes do STJ: Sistema Sócrates 2.0’ (*Portal Migalhas*, 31 mai 2021) <<https://www.migalhas.com.br/depeso/346278/a-inteligencia-artificial-na-formacao-dos-precedentes-do-stj>> acesso em 20 jul 2022.

No tocante ao controle pelo usuário,²¹³ a referida Resolução estabelece que os modelos utilizados deverão possibilitar “a revisão da proposta de decisão e dos dados utilizados para sua elaboração, sem que haja qualquer espécie de vinculação à solução apresentada pela Inteligência Artificial.”²¹⁴ Estabelece também o dever de informação aos usuários sobre a utilização de sistema inteligente nos serviços que forem a eles prestados, o que deve ser feito “em linguagem clara e precisa”.²¹⁵

Outro importante ponto constante na Resolução é a previsão de que os modelos de inteligência artificial deverão observar, como critério preponderante para definir a técnica utilizada, “a explicação dos passos que conduziram ao resultado”. A explicabilidade está aqui prevista, portanto, no *caput* do artigo 19, que é complementado pelo parágrafo único, esclarecendo que esses sistemas “deverão permitir a supervisão do magistrado competente”. É nítida a preocupação em assegurar a transparência em suas duas faces: acessibilidade e compreensibilidade.²¹⁶

Nessa linha, o artigo 24 determina que os modelos de inteligência artificial devem adotar, preferencialmente, *softwares* de código aberto que permitam maior transparência. E, como medidas destinadas a assegurar a transparência, incluem-se, entre outras, o “fornecimento de explicação satisfatória e passível de auditoria por autoridade humana quanto a qualquer proposta de decisão apresentada pelo modelo de Inteligência Artificial, especialmente quando essa for de natureza judicial”, bem como a “apresentação dos mecanismos de auditoria e certificação de boas práticas.”²¹⁷

Ainda, interessante notar a disposição que estabelece que os dados utilizados no processo de treinamento do modelo sejam obtidos de fontes seguras – preferencialmente governamentais – e mantidos em cópia para cada versão de modelo desenvolvida,²¹⁸ o que tem por objetivo buscar a precisão e a auditabilidade do sistema desde a sua concepção, por meio da preservação da base de dados de treinamento. Ademais, ao assegurar a diversidade e a interdisciplinariedade na composição das equipes para pesquisa, desenvolvimento e

²¹³ O Artigo 3º, inciso IV, define usuário como “pessoa que utiliza o sistema inteligente e que tem direito ao seu controle, conforme sua posição endógena ou exógena ao Poder Judiciário, pode ser um usuário interno ou um usuário externo”.

²¹⁴ Artigo 17, II.

²¹⁵ Artigo 18.

²¹⁶ Nesse sentido, vide Brent Daniel Mittelstadt e outros, ‘The ethics of algorithms: Mapping the debate’ (2016) *Big Data & Society* 1.

²¹⁷ Artigo 8º.

²¹⁸ Artigos 13 a 16.

implantação das soluções que utilizem IA,²¹⁹ impõe-se importante medida voltada à redução dos riscos de enviesamento algorítmico.

Por fim, a Resolução enfatiza a utilidade da inteligência artificial para auxiliar no tratamento isonômico aos jurisdicionados, para além do tema da não discriminação. Por exemplo, no artigo 1º, afirma que o conhecimento associado à inteligência artificial pode “aprofundar maior compreensão entre a lei e o agir humano”. O Artigo 2º esclarece que o uso da inteligência artificial pelo Poder Judiciário visa a promover a “prestação equitativa da jurisdição”. Por sua vez, o Artigo 5º explicita que a utilização dos modelos deverá “colaborar para que o Poder Judiciário respeite a igualdade de tratamento aos casos absolutamente iguais”.

Na sequência, no ano de 2021, foi publicada a Estratégia Brasileira de Inteligência Artificial, já citada neste trabalho. Em síntese, no tocante à explicabilidade, esse documento 1) recomenda a intervenção humana para decisões que envolvam alto risco para os indivíduos; 2) afirma ser desejável que os sistemas sejam passíveis de explicação e interpretação, por meio do fornecimento de informações significativas que permitam interpretar o sistema; 3) reconhece a importância de que os usuários tenham ciência das suas interações com sistemas de IA.²²⁰

Dentre as ações estratégicas do eixo “Legislação, Regulação e Uso Ético”, destacam-se para o fim deste estudo:²²¹ 1) “criar parâmetros sobre a intervenção humana em contextos de IA em que o resultado de uma decisão automatizada implica um alto risco de dano para o indivíduo”; 2) “estabelecer como requisito técnico em licitações que os proponentes ofereçam soluções compatíveis com a promoção de uma IA ética”; 3) “incentivar a exploração e o desenvolvimento de mecanismos de revisão apropriados em diferentes contextos de utilização de IA por organizações privadas e por órgãos públicos”; 4) “criar e implementar melhores práticas ou códigos de conduta com relação à coleta, implantação e uso de dados, incentivando as organizações a melhorar sua rastreabilidade, resguardando os direitos legais”; 5) “estimular ações de transparência e de divulgação responsável quanto ao uso de sistemas de IA, e promover a observância, por tais sistemas, de direitos humanos, de valores democráticos e da diversidade”.

²¹⁹ Artigo 20.

²²⁰ EBIA (n 3) 21.

²²¹ *ibid* 23.

No eixo “setor público”, destacam-se as seguintes recomendações: 1) “estabelecer mecanismos para célere apuração de denúncias e reclamações sobre violações de direitos em decisões realizadas por sistemas de IA”; 2) “estabelecer valores éticos para uso da IA na Administração Pública Federal”; e 3) “realizar análise de impacto nos casos de uso da IA que afetem diretamente o cidadão ou o servidor público”.²²²

Diferentemente do ecossistema estruturado no âmbito do Poder Judiciário, na esfera do Poder Executivo não há o acompanhamento dos projetos que envolvam sistemas de decisões algorítmicas na Administração Pública por um órgão centralizado.²²³ São conhecidos alguns projetos esparsos, citados inclusive na EBIA publicada em 2021, mas, desde então, assim como no Poder Judiciário, outras tantas iniciativas que envolvem o uso de inteligência artificial devem estar em desenvolvimento nos mais distintos órgãos – municipais, estaduais e federais –, sem que se tenha a concentração da governança.²²⁴

No âmbito administrativo, aponta-se que o uso da inteligência artificial apresenta utilidade tanto para as tarefas relacionadas ao processo de tomada de decisão quanto para as relacionadas à predição.²²⁵ No campo da decisão, seria possível cogitar a automação integral da decisão para os atos administrativos vinculados, visto que neles basta a constatação material da presença dos elementos definidos em lei (e.g., deferimento de aposentadoria por idade). Já para os atos discricionários, haveria um espaço de decisão dentro do qual, a princípio, a máquina não poderia avançar, pois dependeria de uma opção estratégica ou valorativa do ser humano.²²⁶

No entanto, o tema da automação da discricionariedade é controverso e está em desenvolvimento. Há quem entenda que o fato de um processo organizacional exigir o exercício da discricionariedade em determinada etapa não inviabiliza, por si só, a sua automatização. Isso porque não é clara e, diante do desenvolvimento da tecnologia, não

²²² *ibid* 45.

²²³ Esse ecossistema estruturado no âmbito do Poder Judiciário é reforçado pela Portaria nº 271/2020 do Conselho Nacional de Justiça. CNJ, Portaria nº 271/2020 (2020) <<https://atos.cnj.jus.br/atos/detalhar/3613>> acesso em 10 set 2022.

²²⁴ EBIA (n 3) 42-43.

²²⁵ Vanice Lírio do Valle, ‘Inteligência artificial incorporada à Administração Pública: mitos e desafios teóricos’ (2020) 81 *Revista de Direito Administrativo e Constitucional* 179, 187.

²²⁶ *ibid*. Nesse sentido também, Administrative Review Council, ‘Report 46 - Automated Assistance in Administrative Decision Making 2004’ (Commonwealth of Australia 2004) <<https://www.ag.gov.au/legal-system/publications/report-46-automated-assistance-administrative-decision-making-2004>> acesso em 10 set 2022.

parece ser definitiva a afirmação de que todos os atos discricionários precisam ser exercidos pessoalmente por um agente humano tomador de decisão.²²⁷

Nessa linha, poder-se-ia cogitar a hipótese de, dentro de determinada temática, programar o exercício automatizado da discricionariedade para beneficiar os administrados (e.g., em determinado período, deferir todos os pedidos de férias de servidores públicos).²²⁸ Outro caminho possível seria proibir o uso de decisões automatizadas em algumas matérias, como fez a legislação francesa, que, na esfera administrativa, vedou o uso de sistemas de decisões algorítmicas em grau de recurso.²²⁹ No entanto, para os demais tipos de decisões administrativas, o ordenamento jurídico francês permite, como regra, a automatização parcial ou total, desde que uma série de condições sejam respeitadas, incluindo a notificação explícita ao administrado, bem como a possibilidade de explicação, em detalhes e de forma inteligível, acerca do processamento algorítmico a quem for por ele afetado.²³⁰

Ocorre que, no Brasil, inexistente regulamentação específica que abranja a Administração Pública como um todo. A citada Resolução nº 332/2020 do CNJ, que bem esmiúça o tema, tem âmbito de aplicação restrito ao Poder Judiciário, ante a competência do Conselho Nacional de Justiça prevista no artigo 103-B, parágrafo 4º, da Constituição da República.²³¹ Por sua vez, a EBIA contém diretrizes e recomendações gerais, mas carece de maior detalhamento.

Ainda assim, considerando que o uso de sistemas algorítmicos no Poder Judiciário brasileiro é permitido, soaria lógico argumentar que também o seja para a Administração Pública, visto que o grau de afetação ou risco para os administrados é igual ou até menor do que para os jurisdicionados, diante da possibilidade de submissão do ato administrativo ao crivo judicial. Assim, ao menos no que tange às decisões semiautomatizadas aparenta não haver dúvidas

²²⁷ Commonwealth Ombudsman, Automated Decision-Making Better Practice Guide (2020). <https://www.ombudsman.gov.au/__data/assets/pdf_file/0030/109596/OMB1188-Automated-Decision-Making-Report_Final-A1898885.pdf> acesso em 14 jul 2022.

²²⁸ Ibid.

²²⁹ Gianclaudio Malgieri, 'Automated Decision-Making in the EU Member States: The Right to Explanation and Other "Suitable Safeguards" for Algorithmic Decisions in the EU National Legislations' (2019) 35(5) Computer Law & Security Review 1, 13.

²³⁰ Ibid.

²³¹ No ordenamento jurídico brasileiro, o Conselho Nacional de Justiça é um órgão que integra o Poder Judiciário (Artigo 92, I-A, da Constituição da República) e é presidido pelo Presidente do Supremo Tribunal Federal (Artigo 103-B, §1º). Dentre suas atribuições (Artigo 103-B, §4º), está zelar pelo cumprimento dos princípios da administração pública no âmbito do Poder Judiciário e "apreciar, de ofício ou mediante provocação, a legalidade dos atos administrativos praticados por membros ou órgãos do Poder Judiciário, podendo desconstituí-los, revê-los ou fixar prazo para que se adotem as providências necessárias ao exato cumprimento da lei, sem prejuízo da competência do Tribunal de Contas da União".

de que, no sistema jurídico brasileiro, o processo de tomada de decisão administrativa pelo agente humano poderia, em tese, ser apoiado pelo uso de “sistemas inteligentes”.

No tocante à possibilidade de decisões administrativas totalmente automatizadas, a lacuna no ordenamento jurídico brasileiro é ainda maior, porquanto, como visto, no âmbito do Poder Judiciário, são admitidas apenas decisões semiautomatizadas. Além disso, o regime jurídico-administrativo atual parte do pressuposto de que as decisões administrativas são tomadas por agentes humanos, estabelecendo todo um aparato de controle ao ato administrativo, dentro do qual a motivação, como visto, assume grande importância. Então, para a produção de atos administrativos individuais de forma totalmente automatizada, é imprescindível a criação de mecanismos robustos capazes de assegurar o necessário controle desses atos pelo administrado afetado.

Por fim, outra lacuna a suprir seria identificar quais tipos de decisão podem ser delegados total ou parcialmente à máquina. Sobre esse ponto, a Resolução nº 332/2020 do CNJ apesar de declarar que, sobre a aplicação em matéria penal, o uso de sistemas inteligentes não deve ser incentivado, especialmente para modelos de decisões preditivas, permite o uso de soluções destinadas à automação e ao oferecimento de subsídios para “cálculo de penas, prescrição, verificação de reincidência, mapeamentos, classificações e triagem dos autos para fins de gerenciamento de acervo”.²³² Além disso, condiciona a implantação de soluções que envolvam reconhecimento facial à prévia autorização do Conselho Nacional de Justiça.²³³ Dentro desse quadro de indefinições, está a certeza de que as salvaguardas precisam ser asseguradas aos administrados, pois o dever de motivação existe para os atos administrativos individuais, e a sua (semi)automatização não o exclui, pelo contrário, o qualifica pelo dever de explicabilidade algorítmica.

3.3. Motivação e explicabilidade da decisão administrativa algorítmica pela boa Administração Pública

Cabe, então, desenhar os novos contornos do dever de motivação das decisões administrativas algorítmicas para garantir que sirva como meio de controle efetivo não só contra a falibilidade humana, mas também em face dos novos riscos decorrentes da opacidade algorítmica.

²³² Artigo 23.

²³³ Artigo 22, parágrafo 2º.

Nessa linha, em primeiro lugar, é preciso perceber que assegurar a intervenção humana ao final do processo não é suficiente. Apesar de ser importante a previsão de uma instância revisora humana para as decisões administrativas algorítmicas, essa providência, por si só, revela-se fraca se a explicabilidade não houver sido assegurada nas fases anteriores de desenvolvimento, treinamento, validação e implementação do modelo adotado.

Isso porque, como visto, se o próprio agente humano revisor não for capaz de compreender o resultado fornecido pela máquina, as suas possibilidades de ação ficam demasiadamente limitadas. Por um lado, é verdade que, se o erro for manifesto e envolver fatores mais objetivos (e.g., no caso de um reconhecimento facial equivocado), o revisor humano terá facilidade para identificá-lo e substituir a decisão algorítmica que falhou. No entanto, se a decisão envolver maior complexidade e incorporar fatores de caráter subjetivo, é possível – e até provável – que o agente revisor que não compreende o resultado algorítmico tenda a acreditar na suposta objetividade dele.

Nessa linha, o dever de motivação das decisões administrativas algorítmicas passa a estar qualificado pelo dever de explicabilidade, uma vez que, no regime jurídico administrativo, a opacidade não se coaduna com o necessário controle dos atos administrativos individuais, que requer sejam reveladas as suas respectivas razões fundantes. Por conseguinte, essa relação de qualificação entre dever de motivação e explicabilidade algorítmica deve ser uma via de mão dupla. Numa via, a explicabilidade qualifica a motivação no sentido de que o dever de motivar deverá ser acrescido de explicações sobre o funcionamento algorítmico.

Na outra, o dever de motivação dos atos administrativos qualificará a explicabilidade algorítmica no sentido de que essa deverá ser tal que atenda às exigências do regime jurídico administrativo, o que significa que deverá observar altos padrões de exigência. Essa preocupação, como demonstrado nos capítulos anteriores, abrange não apenas o conjunto de dados de treinamento e teste, mas um olhar cuidadoso sobre todas as escolhas humanas feitas em todas as fases de concepção e implementação do modelo, incluindo, evidentemente, a própria escolha do tipo de algoritmo que será utilizado.

Assim, para os atos administrativos individuais, objeto deste estudo, não basta que o modelo ofereça alto nível de acurácia se não oferecer também compreensibilidade sobre os resultados. O melhor algoritmo para automatização das decisões administrativas que produzem efeitos jurídicos concretos sobre os cidadãos, será aquele que, dentro do conjunto de algoritmos explicáveis, apresentar elevado grau de precisão. Assim, em todo o caminho de concepção à implantação de um sistema de decisão administrativa algorítmica, a Administração Pública deverá estar norteada por escolhas humanas que atendam a *standard*

de explicabilidade mais exigente.²³⁴ Idealmente, e em especial quando o escopo do modelo for subsidiar a tomada de decisões administrativas, a explicabilidade global e local deve ser assegurada desde a concepção do sistema algorítmico, no seu próprio desenho (*explainability by design*).

Com efeito, como demonstrado, o movimento de transformação digital da Administração Pública deve caminhar no sentido de promover uma relação ressignificada de efetiva colaboração e participação do cidadão nas atividades administrativas,²³⁵ tendo como objetivo último melhorar as relações entre as instituições governamentais e os administrados.²³⁶ Nesse contexto, a utilização de sistemas algorítmicos que produzem resultados inexplicáveis está em absoluta dissonância com esse movimento de horizontalidade, transparência e cooperação. A legítima expectativa do cidadão do século XXI é que a Administração Pública se modernize, esteja presente no ambiente digital, se torne mais eficiente e acessível. No entanto, também se espera que ela esteja movida pelo objetivo de promover direitos (em especial, o direito fundamental à boa administração pública²³⁷), ampliar a qualidade dos serviços prestados e aumentar a confiança pela atuação transparente.

Essa direção está alinhada com toda a evolução histórica que levou à chamada “era da motivação administrativa”, que exige, como regra, que a boa Administração Pública observe seu dever de explícita justificação dos atos administrativos, como escudo dos cidadãos contra arbitrariedade e desvios de poder.²³⁸ Assim, tendo em vista o escopo ampliado e a importância do dever de motivação como viabilizador do controle dos atos administrativos, deve-se atentar para não extrair conclusões equivocadas ao importar da LGPD subsídios do direito à explicação das decisões automatizadas.

Primeiramente, cabe dizer que a preocupação com o segredo comercial e industrial, resguardada pela Lei Geral de Proteção de Dados, é própria do setor privado e, no caso das decisões tomadas pela Administração Pública, esse não poderá ser um motivo para negar o fornecimento de informações sobre o funcionamento algorítmico, já que a própria atuação

²³⁴ Washington (n 85) 149.

²³⁵ Brega (n 16) 34-35.

²³⁶ Faleiros Júnior (n 20) 258.

²³⁷ Freitas (n 48) 20. Segundo o autor, o direito fundamental à boa administração pública “trata-se do direito fundamental à administração pública eficiente e eficaz, proporcional cumpridora de seus deveres, com transparência, motivação, imparcialidade e respeito à moralidade, à participação social e à plena responsabilidade por suas condutas omissivas e comissivas. A tal direito corresponde o dever de a administração pública observar, nas relações administrativas, a cogência da totalidade dos princípios constitucionais que a regem.”

²³⁸ *ibid* 48-50.

administrativa é, constitucionalmente, regida pelo princípio da publicidade,²³⁹ entendido como máxima transparência a viabilizar a participação fiscalizatória da cidadania.²⁴⁰ Ademais, consoante já referido, o dever de motivar é exigência da administração democrática, uma vez que o mínimo que se pode exigir enquanto titular último do poder é o conhecimento dos motivos das decisões tomadas por quem deve resguardar o interesse público.²⁴¹

Mesmo que a LGPD tenha previsto que o controlador deva fornecer informações sobre “critérios” e “procedimentos” para decisões “unicamente” automatizadas, a Administração Pública está obrigada a assegurar a explicabilidade para os sistemas algorítmicos que utilizar, o que inclui também as decisões semiautomatizadas. Na mesma linha, em que pese o Artigo 20 da LGPD não garantir o direito de revisão das decisões semiautomatizadas, se estiverem submetidas ao direito público, estarão sujeitas à revisão por força do regime jurídico-administrativo, e exigirão motivação qualificada pela explicabilidade no tocante ao aspecto algorítmico.

Em relação à intervenção humana no processo de revisão dessas decisões administrativas, sejam total ou semiautomatizadas, há uma lacuna legislativa que deve ser suprida. Isso se justifica, em especial, diante da Estratégia Brasileira de Inteligência Artificial, que recomenda a criação de parâmetros sobre a intervenção humana em contextos de inteligência artificial quando implicar alto risco de dano para o indivíduo. Da mesma forma, na já citada recomendação da OCDE sobre o tema, da qual o Brasil é signatário, inclui-se a garantia de salvaguardas para os afetados pelas decisões automatizadas, incluindo a intervenção humana quando apropriado ao contexto em que utilizada a inteligência artificial.²⁴²

No ordenamento jurídico brasileiro, tanto a EBIA quanto a citada Resolução nº 332/2020 do Conselho Nacional de Justiça (CNJ), constituem guias de excelência para os caminhos que a Administração Pública deve trilhar no uso de sistemas de decisões automatizadas. Tendo em vista uma análise global da citada Resolução, com destaque para a preocupação em conferir tratamento isonômico aos cidadãos e em reconhecer a utilidade da inteligência artificial para “aprofundar maior compreensão entre a lei e o agir humano”, o Poder Judiciário parece ter compreendido que a máquina pode contribuir para a produção de decisões humanas mais ágeis, precisas e coerentes.

²³⁹ Artigo 37, *caput*, da Constituição da República Federativa do Brasil de 1988.

²⁴⁰ Freitas (n 47) 56.

²⁴¹ Bandeira de Mello (n 42) 113.

²⁴² OCDE (n 102) item 1.2.b.

Nessa linha, se bem concebidos e desenvolvidos, os sistemas de decisões algorítmicas podem enriquecer o exercício da discricionariedade, oferecendo distintos cenários para nortear o processo decisório humano em relação às consequências práticas das escolhas possíveis.²⁴³ A Lei nº 13.655/2018, que dispõe sobre normas gerais sobre segurança jurídica e eficiência na aplicação do direito público, estabelece que “nas esferas administrativa, controladora e judicial, não se decidirá com base em valores jurídicos abstratos sem que sejam consideradas as consequências práticas da decisão”. Essa norma exige que a motivação demonstre a necessidade e adequação da decisão “inclusive em face das possíveis alternativas”. Então, não há dúvida de que a análise estatística pode auxiliar o tomador de decisão a dar cumprimento a essa análise consequencialista, permitindo, por meio do processamento de grande volume de dados, identificar correlações que não seriam facilmente perceptíveis ao olhar puramente humano.

Além disso, a conjugação do julgamento humano e da máquina, em especial o auxílio da análise estatística, pode contribuir para decisões não só mais eficientes, mas especialmente mais isonômicas em prol dos administrados. Com efeito, seria possível cogitar um escopo algorítmico cuja finalidade fosse a própria redução do ruído nas decisões humanas, ou seja, um modelo destinado a identificar a variabilidade inexplicável do próprio julgamento humano. Por exemplo, em matéria de fixação de indenizações ou de imposição de penalidades administrativas, um sistema nesse sentido seria de indiscutível utilidade para fazer convergir o julgamento humano e o da máquina em direção à isonomia de tratamento. Entretanto, como já ressaltado, no ordenamento jurídico brasileiro, diferentemente do ecossistema estruturado no âmbito do Poder Judiciário, na esfera do Poder Executivo não há o acompanhamento dos projetos que envolvam sistemas de decisões algorítmicas na Administração Pública por um órgão centralizado. Apesar de recente iniciativa do Tribunal de Contas da União nesse sentido, sua abrangência está restrita, por competência, aos órgãos administrativos federais.²⁴⁴

Há, portanto, todo um plexo de órgãos federais, estaduais e municipais neste País de dimensões continentais que estão à margem de governança. Desconhecem-se projetos em desenvolvimento ou até mesmo já em prática nas diferentes esferas da Administração Pública. Diante da ausência de um mecanismo institucionalizado que propicie o

²⁴³ Lírio do Valle (n 238) 187.

²⁴⁴ Tribunal de Contas da União, ‘TCU avalia uso de inteligência artificial pelo governo federal’ (TCU 2022) <<https://portal.tcu.gov.br/imprensa/noticias/tcu-avalia-uso-de-inteligencia-artificial-pelo-governo-federal.htm>> acesso em 20 ago de 2022.

compartilhamento dessas iniciativas, perde-se eficiência; e, diante da ausência de regulação específica, perde-se segurança jurídica.

Assim, no Brasil, remanescem dúvidas a respeito dos tipos de decisões administrativas que podem ser automatizadas e do grau de automatização apropriado em cada caso. Diante de uma LGPD que não veda, como regra, decisões totalmente automatizadas e de uma estratégia nacional que, declaradamente, deseja alavancar a Administração Pública Digital, é de se esperar que, assim como ocorreu no âmbito do Poder Judiciário, surjam as mais diversas iniciativas espalhadas pelos diferentes órgãos administrativos.

Por todas as razões até aqui expostas, mostra-se recomendável que, seguindo o caminho já trilhado pelo Conselho Nacional de Justiça, o uso de sistemas de decisões algorítmicas na Administração Pública seja regulamentado, de forma a preencher as lacunas existentes e assegurar as garantias necessárias para garantir muito além de uma revisão humana ao final do processo. O papel que cabe ao ser humano é muito mais desafiador e precisa estar presente, com consciência e responsabilidade, em todas as fases, desde a concepção até a implantação, permanecendo posteriormente, na supervisão perene e na auditoria periódica desses sistemas.

A grande decisão humana há de ser a definição de como os algoritmos devem ser desenvolvidos e integrados aos processos administrativos, o que demanda a elaboração de análises de impacto para avaliar os riscos e estabelecer as medidas necessárias para mitigá-los.²⁴⁵ Sem dúvidas, como importante forma de controle dessas decisões estará o dever de motivação, que há de ser qualificado pela explicabilidade algorítmica em *standard* elevado.

²⁴⁵ Coglianese e Lehr (n 37) 1177.

Conclusão

Dentre os diversos desafios ao uso de algoritmos para a tomada de decisões administrativas, o presente trabalho buscou enfrentar a preocupação em assegurar que os administrados afetados por essas decisões possam compreendê-las e desafiá-las, ou, dito de outra forma, que possam contra-argumentar o algoritmo.

Ressaltou-se, nesse sentido, a importância do dever de motivação como elemento essencial para viabilizar o efetivo controle dos atos administrativos, destacando-se que o alargamento do âmbito de aplicação da motivação coincide com a própria evolução do Estado de Direito. Demonstrou-se que o atual desafio está em definir os contornos desse dever em relação às decisões administrativas algorítmicas para enfrentar, além da natural falibilidade humana, os novos riscos que surgem em decorrência do uso da tecnologia.

Nesse contexto, o estudo abordou quatro aspectos que podem contribuir com a falibilidade da decisão algorítmica: a falta de compreensão de que há decisões humanas no processo de construção de algoritmos e bases de dados; a transposição de vieses pré-existentes para os algoritmos; a necessidade de interpretação dos resultados combinada com a tendência humana de superestimar a sua objetividade; e a opacidade, que consistiu no tema central do trabalho.

Enfrentou-se o problema da explicação algorítmica com olhar técnico e jurídico. No aspecto técnico, demonstrou-se a dificuldade em precisar o grau de influência de cada variável de entrada no resultado, especialmente diante do crescente aumento da complexidade dos algoritmos usados em aplicações de *machine-learning* e inteligência artificial que “aprendem” de forma autônoma. Além disso, foi exposto o desafio adicional de compreender como o algoritmo se comporta em ação, o que depende não só da interpretação do código, mas, substancialmente, da observação de como ele interage com os dados na prática – ponto crucial da opacidade inerente aos sistemas de aprendizado de máquina.

Diante desse cenário, o presente trabalho evidenciou que crescem os estudos que objetivam produzir modelos preocupados com a capacidade de fornecer razões para as previsões algorítmicas. Com efeito, tecnicamente, há meios de assegurar maior nível de explicabilidade, seja por meio do próprio desenho dos modelos (o que exige boas escolhas humanas nas fases de seleção do algoritmo e do seu treinamento), seja por meio do desenvolvimento de métodos para compreender o funcionamento de sistemas algorítmicos existentes, via engenharia reversa.

No aspecto jurídico, a preocupação com a opacidade de sistemas de decisões algorítmicas resultou em recentes normativas sobre o tema, em especial, na matéria de proteção de dados pessoais. Nessa linha, realizou-se um estudo comparativo entre o RGPD e a LGPD no que toca, em particular, ao direito à explicação das decisões automatizadas. Restou demonstrado que, embora não haja previsão específica de um direito à explicação geral e aplicável a todas as decisões automatizadas na LGPD, em especial diante do escopo restrito do Artigo 20, este pode ser entendido como a consagração da diretriz da explicabilidade e do vinculado princípio da motivação decisória algorítmica no Direito brasileiro. Ademais, sustentou-se que é possível extrair o direito à explicação a partir de uma interpretação holística da LGPD, que está fundada em um conjunto de normas que assegura robusta proteção aos direitos dos indivíduos.

Avançando na análise do ordenamento jurídico brasileiro, verificou-se que não há ainda uma legislação com abrangência nacional que regule a utilização de sistemas de decisões algorítmicas no setor público. Tratando-se de Estado federado com dimensões continentais, causa preocupação a ausência de controle por um órgão centralizado sobre esse ambiente de desenvolvimento e utilização de sistemas de decisões automatizadas pelos diversos órgãos da Administração Pública.

Ao final, desenharam-se os novos contornos do dever de motivação das decisões administrativas algorítmicas. Sustentou-se que há uma relação de qualificação entre o dever de motivação das decisões administrativas e o dever de explicabilidade algorítmica, que assim se manifesta: a explicabilidade qualifica a motivação no sentido de que o dever de motivar deverá ser acrescido de explicações sobre o funcionamento algorítmico; por sua vez, o dever de motivação dos atos administrativos exige que a explicabilidade algorítmica atenda aos altos padrões de exigência do regime jurídico-administrativo.

Assim, para fazer uso de decisões administrativas algorítmicas, a Administração Pública deverá resguardar a explicabilidade como requisito da motivação, de forma a determinar que a escolha do algoritmo observe padrões exigentes de acessibilidade e compreensibilidade. Como demonstrado ao longo do trabalho, esse exigente *standard* de explicabilidade abrange não apenas o conjunto de dados de treinamento e teste, mas um olhar cuidadoso sobre todas as escolhas humanas feitas em todas as fases de concepção e implementação do modelo (*explicability by design*).

No tocante à revisão da decisão, demonstrou-se que não está claro se o ordenamento jurídico brasileiro garante ou não a intervenção humana, sendo possíveis interpretações em ambos os sentidos. Por um lado, a LGPD não assegura a participação humana no processo de revisão da decisão automatizada. Por outro, a EBIA recomenda a criação de parâmetros sobre a intervenção humana em contextos em que o resultado de uma decisão automatizada implica um “alto risco de dano” para o indivíduo, mas nem toda decisão administrativa será enquadrável nesse conceito a depender do seu contexto.

Para além disso, o trabalho propôs-se a alertar para o fato de que assegurar a intervenção humana no final do processo, como instância revisora, revela-se medida insuficiente se a explicabilidade não houver sido assegurada nas fases anteriores de desenvolvimento, treinamento, validação e implementação do modelo adotado. Por meio da análise do caso Loomis, buscou-se demonstrar que mesmo uma decisão semiautomatizada motivada pode, na prática, impedir que o afetado a contra-argumente diante de um cenário de ausência de explicabilidade algorítmica.

Por fim, em especial diante da incerteza em relação aos tipos de decisões administrativas que podem ser automatizadas e do grau de automatização apropriado em cada caso, cabe recomendar a regulação do tema, em linha com o que já fez o Poder Judiciário brasileiro, de modo que sejam supridas as lacunas identificadas ao longo desse trabalho para que a utilização de sistemas de decisões administrativas algorítmicas possa infundir confiança e resguardar segurança jurídica para os administrados. Ademais, sugere-se a realização de futuros estudos com escopo específico de trabalhar o uso de algoritmos para identificar a variabilidade inexplicável do próprio julgamento humano e fazê-lo convergir com a máquina em direção à produção de decisões mais isonômicas em prol dos administrados.

Bibliografia

__ 'I Jornada de Direito Administrativo aprova 40 enunciados' (*Conselho da Justiça Federal*, 10 ago 2022) <<https://www.cjf.jus.br/cjf/noticias/2020/08-agosto/i-jornada-de-direito-administrativo-aprova-40-enunciados>> acesso em 10 ago 2022.

__ 'What is Deep Learning?' (*IBM Cloud Education*, 1 mai 2020) <<https://www.ibm.com/cloud/learn/deep-learning>> acesso em 23 jun 2022.

__ 'Correlation vs Causation' (*The Decision Lab*) <<https://thedecisionlab.com/reference-guide/philosophy/correlation-vs-causation>> acesso em 10 set 2022

Administrative Review Council, 'Report 46 - Automated Assistance in Administrative Decision Making 2004' (Commonwealth of Australia 2004) <<https://www.ag.gov.au/legal-system/publications/report-46-automated-assistance-administrative-decision-making-2004>> acesso em 10 set 2022.

Ananny M e Crawford K, 'Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability' (2018) 20(3) *New Media and Society* 973

Angwin J e outros, 'Machine Bias' (*ProPublica*, 23 de maio de 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> acesso em 07 de setembro de 2022

Araújo F, *Motivação e controle do ato administrativo* (2ª edição, Del Rey 2005)

Article 29 Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' WP251(rev 6 February 2018)

Bannister F e Connolly R, 'Trust and transformational government: A proposed framework for research' (2011) 28(2) *Government Information Quarterly* 137

Bobbitt Z, 'Correlation Does Not Imply Causation: 5 Real-World Examples' (18 ago 2021) <<https://www.statology.org/correlation-does-not-imply-causation-examples/>> acesso em 15 ago 2022

Borgesius F, 'Discrimination, Artificial Intelligence and Algorithm Decision-Making' (2018) Council of Europe

Bounabat B, 'From e-Government to digital Government: Stakes and Evolution Models' (2017) 10 *Electronic Journal of Information Technology*

Brega J, 'Governo Eletrônico e Direito Administrativo' (Tese de doutorado, Universidade de São Paulo 2012)

Burrell J, 'How the machine "thinks": Understanding opacity in machine learning algorithms' (2016) *Big Data & Society*

Cabral T, 'AI and the Right to Explanation: Three Legal Bases under the GDPR' in Dara Hallinan, Ronald Leenes e Paul De Hert (eds), *Data Protection and Privacy: Data Protection and Artificial Intelligence* (Hart Publishing 2021)

Caliskan A, Bryson J e Narayanan A, 'Semantics derived automatically from language corpora contain human-like biases' (2017) 356 (6334) *Science* 183

Carvalho Filho J, *Manual de Direito Administrativo* (32ª edição, Atlas 2018)

Castelluccia C e Métayer D, 'Understanding algorithmic decision-making: Opportunities and challenges' (European Parliamentary Research Service, 2019) 1

Chaves G e Veiga E, 'A inteligência artificial na formação dos precedentes do STJ: Sistema Sócrates 2.0' (*Portal Migalhas*, 31 mai 2021) <<https://www.migalhas.com.br/depeso/346278/a-inteligencia-artificial-na-formacao-dos-precedentes-do-stj>> acesso em 20 jul 2022

Christin A, Rosenblat A e Boyd D, 'Courts and Predictive Algorithms' (2015) *Data & Civil Rights: A New Era of Policing and Justice* <http://www.datacivilrights.org/pubs/2015-1027/Courts_and_Predictive_Algorithms.pdf> acesso em 10 set 2022

Cintra A, *Motivo e motivação do ato administrativo* (RT 1979)

Código de Processo Penal brasileiro, Decreto-Lei nº 3.689, de 3 de outubro de 1941

Coglianese C e Lehr D, 'Regulating by Robot: Administrative Decision Making in the Machine-Learning Era' (2017) 105 *The Georgetown Law Journal* 1147

Coglianese C, 'Robot Regulators could eliminate human error' (*The Regulatory Review*, 16 mai 2016) < <https://www.theregreview.org/2016/05/16/coglianese-robot-regulators-eliminate-error/?msclkid=0b239623a94411ec803bebc5f6522e9a>> acesso em 07 jul 2022

Commonwealth Ombudsman, *Automated Decision-Making Better Practice Guide* (2020). <https://www.ombudsman.gov.au/__data/assets/pdf_file/0030/109596/OMB1188-Automated-Decision-Making-Report_Final-A1898885.pdf> acesso em 14 jul 2022.

Council on Digital Government Strategies' (2014) 1, 2 <<https://www.oecd.org/gov/digital-government/recommendation-on-digital-government-strategies.htm>> acesso em 10 jun 2022

Cretella Júnior J, *Curso de Direito Administrativo* (Forense 1986)

Delua J, 'Supervised vs. Unsupervised Learning: What's the Difference?' (*IBM Cloud*, 12 mar 2021) <<https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning>> acesso em 23 jun 2022

Di Pietro M, *Direito Administrativo* (22ª edição, Atlas 2008)

Di Pietro M, *Discrecionalidade administrativa na Constituição de 1988* (3ª edição, Atlas 2012)

European Commission, 'EU eGovernment Action Plan 2016-2020' (Communication) COM(2016) 179 final

European Commission, *Orientações éticas para uma IA de confiança* (Publications Office 2019) <<https://data.europa.eu/doi/10.2759/2686>> acesso em 10 jul 2022

European Union Agency for Fundamental Rights, 'Big data: discrimination in data-supported decision making' (Publications Office 2018)

Faggella D, 'What is Machine Learning?' (*Emerj*, 26 fev 2020) <<https://emerj.com/ai-glossary-terms/what-is-machine-learning/>> acesso em 23 jun 2022

Faleiros Júnior J, *Administração Pública Digital: proposições para o aperfeiçoamento do Regime Jurídico Administrativo na sociedade da informação* (Foco 2020)

Freitas J e Freitas T, *Direito e Inteligência Artificial: em defesa do humano* (Forum 2021)

Freitas J, *Discrecionalidade administrativa e o direito fundamental à boa administração pública* (Malheiros 2007)

Freitas J, *O controle dos atos administrativos e os princípios fundamentais* (Malheiros 2004)

Government of South Australia, 'Digital by Default Declaration' (2014) <https://www.dit.sa.gov.au/__data/assets/pdf_file/0005/264695/Digital-by-Default-Declaration.pdf> acesso em 07 set 2022.

Grissinger M, 'Understanding Human Over-Reliance On Technology' (2019) 44(6) PT 375.

Guidotti R e outros, 'A Survey of methods for explaining black box models' (2019) 51(5) ACM Computing Surveys

Jeff Larson e outros, 'How We Analyzed the COMPAS Recidivism Algorithm' (*ProPublica*, 23 de maio de 2016) <<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>> acesso em 07 de setembro de 2022

Kahneman D, Sibony O e Sunstein C, *Ruído: uma falha no julgamento humano* (Cássio de Arantes Leite tr, Objetiva 2021)

Kaminski M, 'The Right to Explanation, Explained' (2019) 34 Berkeley Technology Law Journal 189; Selbst A e Powles J, 'Meaningful information and the right to explanation' (2017) 7(4) International Data Privacy Law 233

Kavlakoglu E, 'AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference?' (*IBM Cloud*, 27 mai 2020) <<https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>> acesso em 23 jun 2022

Khan G, "Social Media-based Government Explained: Utilization Model, Implementation Scenarios, and Relationships" in Boughzala I, Janssen M e Assar S, *Case Studies in e-Government 2.0: Changing Citizen Relationships* (Springer 2015)

Knight W, 'The Dark Secret at the Heart of AI' (2017) 120 MIT Technology Review <<https://www.technologyreview.com/2017/04/11/51113/the-dark-secret-at-the-heart-of-ai/>> acesso em 07 jul 2022

Lage F, *Manual de Inteligência Artificial no Direito Brasileiro* (JusPodivm 2021)

Lehr D e Ohm P, 'Playing with the Data: What Legal Scholars Should Learn About Machine Learning' (2017) 51 Davis University of California 653

Lei brasileira nº 13.709 de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais) (2018)

Lírio do Valle V, 'Inteligência artificial incorporada à Administração Pública: mitos e desafios teóricos' (2020) 81 Revista de Direito Administrativo e Constitucional 179

Lowry S e Macpherson G, 'A blot on the profession' (1988) 296(6623) British medical journal 657

Maeji V, 'Justiça 4.0: Inteligência Artificial está presente na maioria dos tribunais brasileiros' (*Conselho Nacional de Justiça* 2022) <<https://www.cnj.jus.br/justica-4-0-inteligencia-artificial-esta-presente-na-maioria-dos-tribunais-brasileiros/>> acesso em 10 set 2022

Magrani E, Perrone C e Souza C, 'O direito à explicação entre a experiência europeia e a sua positivação na LGPD' in Doneda D e outros (coord), *Tratado de proteção de dados pessoais* (Forense 2021)

Malgieri G e Comandé G, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (2017) 7(4) International Data Privacy Law 243.

Malgieri G, 'Automated Decision-Making in the EU Member States: The Right to Explanation and Other "Suitable Safeguards" for Algorithmic Decisions in the EU National Legislations' (2019) 35(5) Computer Law & Security Review

Mello C, *Curso de Direito Administrativo* (26ª edição, Malheiros 2009)

Mello C, *Discrecionabilidade e controle jurisdicional* (2ª edição, Malheiros 2003)

Ministério da Ciência, Tecnologia e Inovações Secretaria de Empreendedorismo e Inovação, 'Estratégia Brasileira de Inteligência Artificial' (2021) <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivos/inteligenciaartificial/ebia-diagramacao_4-979_2021.pdf> acesso em 10 set 2022

Mittelstadt B e outros, 'The ethics of algorithms: Mapping the debate' (2016) Big Data & Society

Organização para a Cooperação e Desenvolvimento Econômico, 'Recommendation of the Council on Digital Government Strategies' (2014) 1, 2 <<https://www.oecd.org/gov/digital-government/recommendation-on-digital-government-strategies.htm>> acesso em 10 jun 2022

Organização para a Cooperação e Desenvolvimento Econômico, ‘Recommendation of the Council on Artificial Intelligence’ (2019) C/MIN(2019)3/FINAL <<https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>> acesso em 05 jul 2022

Organização para a Cooperação e Desenvolvimento Econômico, ‘Trust in Government: Evidence, Policies and Decision-making’ (2013) C/MIN(2013)4/FINAL<[http://www2.oecd.org/oecdinfo/info.aspx?app=OLIScoteEN&Ref=C/MIN\(2013\)4/FINAL](http://www2.oecd.org/oecdinfo/info.aspx?app=OLIScoteEN&Ref=C/MIN(2013)4/FINAL)> acesso em 05 jul 2022

Pasquale F e Citron D, ‘The scored society: due process for automated predictions’ (2014) 89 Washington Law Review

Pyle D e San José C, ‘An executive’s guide to machine learning’ (*McKinsey Quarterly*, 1 jun 2015) <<https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/an-executives-guide-to-machine-learning>> acesso em 23 jun 2022

Regulamento (EU) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016 relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados) (2016) OJ L119/1

Ribeiro M, Singh S e Guestrin C, ‘Why Should I Trust You?: Explaining the Predictions of Any Classifier’ (2016) arXiv <<https://arxiv.org/abs/1602.04938>> acesso em 10 set 2022.

Richardson R, ‘Defining and Demystifying Automated Decision Systems’ 81(3) Maryland Law Review 785

Richmond K, ‘AI, Machine Learning, and International Criminal Investigations: The lessons from forensic science’ (iCourts Working Paper Series 222/2020) <<https://ssrn.com/abstract=3727899>> acesso em 10 ago 2022

Santa Rosa G, ‘Lei da inteligência artificial no Brasil: entenda o projeto aprovado na Câmara’ (*Tecnoblog* 2022) <<https://tecnoblog.net/especiais/lei-da-inteligencia-artificial-no-brasil-entenda-o-projeto-aprovado-na-camara/>> acesso em 13 jul 2022

Stachewski A, ‘Como a Estônia se tornou uma nação digital e um berço de negócios inovadores’ (Revista Pequenas Empresas & Grandes Negócios, 16 ago 2022) <<https://revistapegn.globo.com/Startups/noticia/2022/08/como-estonia-se-tornou-uma-nacao-digital-e-um-berco-de-negocios-inovadores.html>> acesso em 5 set 2022

State of Wisconsin v Eric L Loomis (2016) 881 NW2d (Supreme Court of Wisconsin)

Surden H, ‘Machine Learning and Law’ 89 Washington Law Review (2014) 87

UK Cabinet Office, ‘Government Digital Strategy: action 6 - digital by default service standard’ (2013) <<https://www.gov.uk/government/publications/government-digital-strategy-action-6>> acesso em 12 jul 2022

Valente J, 'Órgãos públicos usam inteligência artificial para combater corrupção' (Agência Brasil, 3 ago 2018) <<https://agenciabrasil.ebc.com.br/geral/noticia/2018-08/orgaos-publicos-usam-inteligencia-artificial-para-combater-corrupcao> > acesso em 23 jul 2022

Wachter S, Mittelstadt B e Floridi L, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7(2) International Data Privacy Law 76

Washington A, 'How to Argue with an Algorithm: Lessons from the COMPAS ProPublica Debate' (2019) 17(1) Colorado Technology Law Journal 131

Završnik A, 'Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings' (2019) European Journal of Criminology

Índice

INTRODUÇÃO	1
1. OS ALGORITMOS NA ADMINISTRAÇÃO PÚBLICA DIGITAL	4
1.1. A AUTOMATIZAÇÃO DOS SERVIÇOS PÚBLICOS	4
1.2. MOTIVAÇÃO E CONTROLE DA FALIBILIDADE HUMANA	9
1.3. MOTIVAÇÃO E CONTROLE DA FALIBILIDADE ALGORÍTMICA	14
2. A DECISÃO ALGORÍTMICA: ENTRE O DEVER DE MOTIVAÇÃO E A EXPLICAÇÃO POSSÍVEL	20
2.1. OS DESAFIOS IMPOSTOS PELOS SISTEMAS DE DECISÕES ALGORÍTMICAS	20
2.2. O PROBLEMA DA OPACIDADE ALGORÍTMICA	26
2.3. O DIREITO À EXPLICAÇÃO NO RGPD E NA LGPD	31
3. OS NOVOS CONTORNOS DO DIREITO À MOTIVAÇÃO NA ADMINISTRAÇÃO PÚBLICA DIGITAL	37
3.1. O CIDADÃO, O ALGORITMO E O CONTRA-ARGUMENTO	37
3.2. AS DECISÕES ALGORÍTMICAS E A EXIGÊNCIA DE EXPLICABILIDADE NO CONTEXTO NORMATIVO BRASILEIRO	42
3.3. MOTIVAÇÃO E EXPLICABILIDADE DA DECISÃO ADMINISTRATIVA ALGORÍTMICA PELA BOA ADMINISTRAÇÃO PÚBLICA	48
CONCLUSÃO	54
BIBLIOGRAFIA	57