Algoritmos e direitos fundamentais: riscos, transparência e *accountability* no uso de técnicas de automação decisória

Algorithms and civil rights: risks, transparency and accountability in automated decision-making

Assista aos comentários do autor para este artigo

João Paulo Lordelo

Pós-doutor em Direitos Humanos pela Universidade de Coimbra. Doutor em Direito pela UFBA. *Academic Visitor* na Universidade de Oxford. Professor do programa de pós-graduação em Direito do Centro Universitário Paraíso – UniFAP e do Instituto Brasiliense de Direito Público – IDP. Coordenador pedagógico e professor da Escola Superior do Ministério Público da União (ESMPU). Membro da Comissão de Juristas da Câmara dos Deputados designada para elaboração de anteprojeto de reforma da Lei de Lavagem. Ex-Defensor Público Federal (2010-2014). Membro do Ministério Público Federal (Procurador da República).
ORCID: [https://orcid.org/0000-0002-2528-7699].
Lattes: http://lattes.cnpq.br/3696439036747412
joaolordelo@gmail.com

Recebido em: 12.04.2021
Aprovado em: 22.09.2021
Última versão do autor: 25.09.2021

Áreas do Direito: Digital; Direitos Humanos

Resumo: O artigo pretende responder ao seguinte problema de pesquisa: como garantir o dever de transparência e promover adequadamente o *accountability* no uso de ferramentas algorítmicas de automação decisória por parte do Poder Público, especialmente no campo da persecução penal? A hipótese é a de que o uso de recursos tecnológicos de inteligência artificial pelo Poder Público é algo irrefreável, sendo capaz de proporcionar diversos benefícios, notadamente no âmbito decisório. A sua utilização, especialmente no campo da persecução penal, implica o reconhecimento de deveres acentuados de transparência e *accountability*, havendo meios tecnológicos de promovê-los de forma satisfatória e sem

Abstract: The article intends to solve the following research problem: how to guarantee the duty of transparency and adequately promote accountability in the use of algorithmic decision-making automation tools by the Government, especially in the field of criminal prosecution? The hypothesis is that the use of technological resources of artificial intelligence by the Public Power is something unstoppable, being able to provide several benefits, notable in the field of decision making. Its use, especially in the field of criminal prosecution, implies the recognition of enhanced duties of transparency and accountability, with technological means to promote them in a satisfactory manner and without

Lordelo, João Paulo. Algoritmos e direitos fundamentais: riscos, transparência e *accountability* no uso de técnicas de automação decisória. *Revista Brasileira de Ciências Criminais*. vol. 186. ano 29. p. 205-236. São Paulo: Ed. RT, dezembro 2021.

prejuízo à propriedade industrial. Os suportes fáticos e teóricos são fornecidos por uma análise comparatista, com destaque apara os relatórios e a disciplina normativa produzia no âmbito da Comunidade Europeia. O método de abordagem empregado é o hipotético-dedutivo.

Palavras-chave: Algoritmos – Direitos fundamentais – Automação decisória – Transparência – Devido processo legal.

prejudice to industrial property. The factual and theoretical supports are provided by a comparative analysis, with emphasis on the reports and the normative discipline produced within the European Community. The approach method used is the hypothetical-deductive one.

Keywords: Algorithms – Civil rights – Automated decision-making – Transparency. Due process of law.

1. Introdução

Cada vez mais, as relações humanas são marcadas pela intermediação de ferramentas tecnológicas complexas, cujo uso tem promovido profundas transformações nos variados campos de interação social.

O desenvolvimento científico exponencial tem apresentado, a cada ano, novas técnicas de processamento de dados, capazes não apenas de possibilitar às pessoas a busca de informações na internet, mas também de automatizar processos decisórios anteriormente atribuídos apenas a seres humanos.

No âmbito das relações privadas, os exemplos variam desde as ferramentas de individualização publicitária em redes sociais até a elaboração de "perfis de risco" por seguradoras e instituições financeiras, ou mesmo a automação decisória na contratação de trabalhadores.

No campo do Poder Público, são variados os usos de algoritmos de inteligência artificial no mundo, a exemplo do emprego de sistemas preditivos pelas autoridades policiais, das ferramentas de avaliação de empregados públicos e das técnicas de auxílio decisório no Poder Judiciário.

Essas rápidas e constantes inovações tecnológicas têm o potencial de promover uma ampla gama de benefícios. Mas os atributos que as fazem tão especiais – a exemplo da capacidade de aprendizado a partir de dados e, portanto, de evoluir no tempo sem um *input* humano explícito – também levantam preocupações.

Isso ocorre em especial no que diz respeito aos desafios necessários à defesa dos direitos fundamentais, como o direito à vida, o direito ao devido processo legal, o direito à presunção de inocência, o direito à privacidade, os direitos trabalhistas, o direito à liberdade de expressão e o direito às eleições livres.

Um documento de referência consiste no estudo denominado "Algoritmos e Direitos Humanos", desenvolvido pelo comitê de especialistas aprovado pelo Conselho da Europa em 2017[1]. Conforme apontado no referido trabalho, a elevada incerteza quanto aos possíveis alcances dos algoritmos de inteligência artificial sequer permitem uma projeção segura quanto à sua suscetibilidade a eventuais marcos regulatórios.

Firmadas tais premissas, o presente artigo pretende responder ao seguinte problema de pesquisa: como garantir o dever de transparência e promover adequadamente o *accountability* no uso de ferramentas algorítmicas de automação decisória por parte do Poder Público, especialmente no campo da persecução penal?

A hipótese é a de que o uso de recursos tecnológicos de inteligência artificial pelo Poder Público é algo irrefreável, sendo capaz de proporcionar diversos benefícios, notadamente no âmbito decisório. A sua utilização, especialmente no campo da persecução penal, implica o reconhecimento de deveres acentuados de transparência e *accountability*, havendo meios tecnológicos de promovê-los de forma satisfatória e sem prejuízo à propriedade industrial.

Para tanto, serão inicialmente desenvolvidos conceitos fundamentais para a adequada compreensão do tema, a exemplo das noções de algoritmos, inteligência artificial e *machine learning*.

Em seguida, serão expostos alguns riscos no uso de algoritmos nos âmbitos privado e público.

Ao final, serão desenvolvidos alguns princípios e regras mínimas que podem orientar o desenvolvimento de marcos regulatórios sobre o tema.

Os suportes fáticos e teóricos são fornecidos por uma análise comparatista, com destaque apara os relatórios e a disciplina normativa produzia no âmbito da Comunidade Europeia.

O método de abordagem empregado é o hipotético-dedutivo.

2. Conceitos fundamentais

Três são os conceitos fundamentais para a compreensão do atual estado da arte das inovações tecnológicas no campo da informática: algoritmo, inteligência artificial e *machine learning*. Sem o seu conhecimento prévio, é impossível avançar na análise das funcionalidades e dos limites à regulação das ferramentas de automação decisória.

1. CONSELHO DA EUROPA, 2018.

2.1. Algoritmos

A cada dia, decisões humanas são delegadas a ferramentas de inteligência artificial desenvolvidas a partir do uso de algoritmos, assim compreendidos os conjuntos finitos de instruções que, seguidas, realizam uma tarefa específica.

Embora possam ser empregados em complexas aplicações desenvolvidas no campo da ciência da computação, os algoritmos possuem um conceito bastante simples. Trata-se de um rito, um conjunto finito de instruções. Na *Encyclopaedia Britannica*, colhe-se o seguinte conceito: "procedimento sistemático que produz, em um finito número de passos, a resposta apara uma questão ou a solução de um problema"[2].

Cuida-se de conceito cuja origem histórica é imprecisa. A sua aplicação remonta à antiguidade, a exemplo do algoritmo de Euclides, utilizado com o objetivo de identificar o máximo divisor comum entre dois números inteiros diferentes de zero[3].

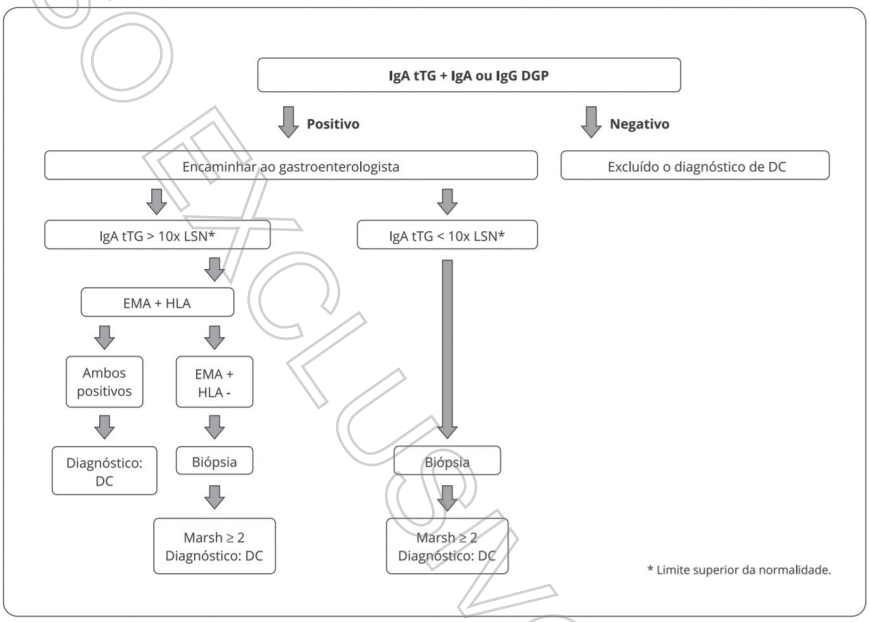
Os algoritmos costumam ser construídos, na sua forma mais básica, por meio das expressões "se" e "então", que expressam funções da linguagem de programação utilizada (Java, Phyton, C++ etc.). Eles podem ser empregados em variadas atividades, não apenas em ferramentas desenvolvidas por profissionais da ciência da computação. Mesmo na informática, os algoritmos podem integrar aplicações mais simples, distantes da complexidade presente em ferramentas de inteligência artificial.

Graficamente, costumam ser representados por meio de "diagramas de fluxo". Um exemplo comum, fora da ciência da computação, são os algoritmos de diagnóstico, fluxogramas utilizados por profissionais médicos que orientam para um diagnóstico correto. É o caso do algoritmo para diagnóstico da doença celíaca, exposto na figura a seguir:

2. ALPAYDIN, 2016, p. 16.

3. EUCLIDES, 2009.

Figura 1 – Algoritmo para diagnóstico de doença celíaca (DC) em crianças ou adolescentes com sintomas ou sinais sugestivos de doença celíaca, incluindo sintomas atípicos



Fonte: SIQUEIRA et al[4].

Qualquer que seja o seu ambiente de aplicação, os algoritmos funcionam a partir de dados (*inputs*) que lhes são fornecidos, para, por meio de uma sequência finita de regras, produzir um resultado (*output*), a exemplo de uma decisão administrativa ou judicial – ou, na figura supra, um diagnóstico médico[5]. No estágio atual da inovação tecnológica do Direito, essas regras são especialmente voltadas à classificação de casos, documentos, eventos ou pessoas, havendo um contínuo avanço para tarefas bastante sofisticadas.

Problemas podem surgir em todas as etapas da aplicação do algoritmo, sobretudo na definição das suas regras e no emprego de dados indevidos. Diante disso, são crescentes os trabalhos vocacionados a levantar preocupações quanto a três grandes problemas que podem decorrer do seu uso: a) o emprego de *data sets* (*inputs*) viciados; b) a opacidade dos algoritmos não programados; c) a discriminação que pode ser gerada por algoritmos de *machine learning*[6].

4. SIQUEIRA; FONSECA; PAULA; NOVAIS, 2014.
5. PEIXOTO; SILVA, 2019, p. 71.
6. FERRARI; BECKER; WOLKART, 2018, p. 635-655.

Nessa linha, Dierle Nunes e Ana Luiza Pinto Coelho Marques exemplificam diversos casos de modelos enviesados, identificados em ferramentas de uso comum:

> "Também é possível verificar vieses algorítmicos no sistema de concessão de crédito europeu e norte-americano, na medida em que diversas companhias utilizam modelos de IA para análise do risco do empréstimo. Muitos desses modelos utilizam até mesmo dados das redes sociais do solicitante para o cálculo do *credit score*, baseando-se, assim, nas conexões sociais do indivíduo. Dessa forma, o resultado vincula-se diretamente ao grupo social no qual o solicitante está inserido. Corroborando tal fato, um relatório de 2007, apresentado pela *Federal Reserve* ao Congresso dos Estados Unidos, apontou que negros e hispânicos têm um *credit score* significativamente inferior ao de brancos e asiáticos.
>
> Outros exemplos de modelos enviesados podem ser citados: um sistema de reconhecimento facial criado pela Google identificou pessoas negras como gorilas; o sistema de busca de contatos do aplicativo Linkedin demonstrou uma preferência por nomes de pessoas do sexo masculino; Tay, mecanismo de IA lançado pela Microsoft para interagir com usuários do Twitter, passou a reproduzir mensagens xenofóbicas, racistas e antissemitas; o aplicativo de *chat* SimSimi, que utiliza inteligência artificial para conversar com os usuários, foi suspenso no Brasil por reproduzir ameaças, palavrões e conversas de teor sexual."[7]

É possível, portanto, que uma ferramenta processual destinada a sugerir ao julgador uma determinada sanção aplicável a um condenado em processo criminal incorpore os mesmos vieses da pessoa a quem busca auxiliar, absorvendo, seja nas regras ou nos dados, os vícios já existentes. A percepção desse tipo de problema tem suscitado preocupações crescentes, expostos ao longo do presente artigo.

2.2. Inteligência artificial

Algoritmos e inteligência artificial são expressões distintas. Existem algoritmos fora da inteligência artificial – e até mesmo fora da ciência da computação –, embora não seja possível cogitar uma ferramenta de inteligência artificial sem o uso de algoritmos[8].

Inexiste um conceito preciso para a expressão "inteligência artificial" (IA). Isso se deve, em especial, às objeções metafísicas que costumam atribuir o atributo da inteligência à "consciência" ou até mesmo à "alma humana", expressões igualmente imprecisas.

De forma ampla, costuma-se compreendê-la como a possibilidade de as máquinas, de alguma forma, desenvolverem a capacidade de *pensar* – ou, ao menos, *imitar* a capacidade de pensamento humana[9].

A PwC, rede de firmas especializadas em auditoria e asseguração, em informe de 2017, conceituou a IA como o "termo coletivo para sistemas de computadores que

7. NUNES; MARQUES, 2018, p. 421-447.

8. VILLASENOR; FOGGO, 2020, p. 301.

9. FENOLL, 2018, p. 20; PEIXOTO, SILVA, 2019, p. 20.

podem perceber o seu ambiente, pensar, aprender e tomar ações em resposta aos dados que apreendidos e seus objetivos"[10].

As discussões iniciais sobre o tema remontam ao artigo *Computering machinery and intelligence*, do célebre matemático inglês Alan Turing, publicada na revista *Mind*, da Universidade de Oxford, no ano de 1950. Em suas primeiras linhas, o autor indica a pergunta que pretende responder em seu estudo: "as máquinas podem pensar?"[11].

Nesse artigo, a apresentação do problema é ilustrada por meio do chamado "jogo da imitação", em que participam três pessoas: um homem (A), uma mulher (B) e um interrogador (C), que pode ser de qualquer gênero. O objetivo do jogo, para o interrogador, consiste em descobrir qual dos outros dois participantes é o homem e qual é a mulher, baseando-se exclusivamente em respostas escritas a questionamentos por ele formulados. O objetivo de A é fazer com que C se equivoque na identificação. Já o objetivo de B é ajudar o interrogador.

Estabelecido o modelo, Turing questiona: o que acontecerá se uma máquina tomar o lugar de A no jogo? Os resultados serão os mesmos?

A partir desses questionamentos, tomando-se ainda o potencial dos computadores no desempenho de tarefas inicialmente a cargo exclusivo de seres humanos, a pergunta original ("as máquinas podem pensar?") é readequada para a seguinte: "existem computadores capazes de ter um bom desempenho no jogo da imitação?"[12].

A forma de colocação das perguntas por Alan Turing é capaz de afastar objeções de natureza teológica, que atribuem a capacidade de pensamento à "alma humana". O autor entende que os pontos de vista das variadas religiões a respeito da alma humana – ou mesmo dos demais animais – não apresentam nenhum obstáculo concreto, mas meras especulações[13].

Uma outra objeção, lançada à época pelo professor Jefferson Lister em 1949, sustentava que "até o dia em que uma máquina puder escrever um soneto um compor um concerto em razão dos seus pensamentos e emoções, não podemos concordar com a ideia de que uma máquina equivaleria ao cérebro humano"[14]. Em resposta, Turing afirma que, por essa objeção, a única forma de saber se uma máquina pensa consiste em *ser* essa máquina e se sentir como um ser pensante. Igualmente, a única maneira de saber se um determinado ser humano pensa consiste em *ser* esse determinado ser humano. O autor conclui que, embora existam, os mistérios sobre a consciência humana não precisam ser resolvidos para que se possa avançar no tema da inteligência das máquinas.

Mesmo que distante de um exemplo concreto – consideradas as limitações tecnológicas da época –, valendo-se de hipóteses, Turing encerra: observadas as adequadas

10. ALPAYDIN, 2016.
11. TURING, 1947; VILLASENOR; FOGGO, 2020.
12. TURING, 1950, p. 442.
13. TURING, 1950, p. 443.
14. TURING, 1950, p. 445.

capacidades de armazenamento e de programação, é possível *imaginar* uma máquina com capacidade de aprendizado, sujeita a uma forma própria de desenvolvimento – como o cérebro de uma criança. Embora o autor não apresente uma resposta conclusiva sobre o tema, as hipóteses levantadas acompanharam o desenvolvimento tecnológico experimentado nas décadas seguintes.

Poucos anos após a publicação do artigo de Alan Turin, John McCarthy, renomado cientista da computação, teria cunhado o termo "inteligência artificial", em conferência realizada no ano de 1956, no Darmouth Colege.

Naquele ano, um pequeno grupo de cientistas desenvolveu o *Dartmouth Summer Research Project on Artificial Intelligence*, que deu início ao campo de pesquisa. Segundo o então professor de matemática John McCarthy, organizador do evento, a conferência tinha "como base a conjectura de que, em princípio, qualquer aspecto do aprendizado ou outra capacidade da inteligência pode ser descrita de forma tão precisa que uma máquina poderia ser criada para simular"[15].

Nos dias atuais, passadas sete décadas, muito do que anteriormente habitava o plano das projeções e hipóteses se transformou em realidade. Os algoritmos de IA acompanham a grande maioria da população mundial, seja durante a navegação na internet, seja em ferramentas tecnológicas domésticas – como assistentes pessoais ou *smartphones*.

2.3. Machine learning e deep learning

Em razão da possibilidade de aprender pela experiência, os algoritmos de IA são capazes de adaptação e evolução. Consequentemente, os sistemas podem criar as próprias regras ou perguntas, sem a necessidade de que um programador preveja soluções específicas para as situações possíveis.

Isso se deve, sobretudo, às técnicas de *machine learning* (aprendizado automático ou aprendizado de máquina), assim compreendido o "campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados". A definição foi cunhada em 1959 por Arthur Samuel, cientista da computação vinculado à IBM[16]. Cuida-se de pressuposto para a inteligência artificial[17].

A capacidade de aprendizado derivada dos processos de *machine learning* depende da análise de um conjunto mínimo de dados. Justamente por isso, o chamado *Big Data* – conjuntos de megadados não estruturados gerados produzidos na era da internet – tende a ser o alimento ideal para o desenvolvimento de múltiplas funcionalidades.

Um exemplo de algoritmo de *machine learning* (simples) consiste no *Naive Bayes*, capaz de separar mensagens de e-mail legítimas de spam[18].

15. Dartmouth, 1956.
16. STANFORD UNIVERSITY INFOLAB, 1964.
17. ALPAYDIN, 2016, p. 17.
18. PEIXOTO; SILVA, 2019, p. 89.

As técnicas de *machine learning* são divididas em algumas categorias, destacando-se as seguintes: *deep learning,* aprendizado por reforço, aprendizado de máquina supervisionado e aprendizado de máquina não supervisionado[19].

Deep learning consiste em um método específico e complexo de aprendizado que utiliza redes neurais em camadas sucessivas no processo de aprendizagem. Redes neurais artificiais são projetadas para emular o funcionamento do cérebro humano, permitindo que as máquinas lidem com abstrações e problemas mal definidos.

Os algoritmos de aprendizagem supervisionada têm o objetivo de classificar dados em categorias previamente definidas. Destinam-se a encontrar padrões em dados que possam ser aplicados em um processo analítico. Por exemplo, um aplicativo de *machine learning* desse tipo é capaz de analisar diversas imagens de animais, classificando-as em espécies previamente rotuladas.

A aprendizagem não supervisionada diz respeito às ferramentas de *machine learning* que demandam uma grande quantia de dados não rotulados para a solução de problemas. Elas buscam, a partir de uma ampla massa de *dataset,* identificar propriedades úteis, sem a intervenção humana. É o que ocorre nas redes sociais, em que circulam grandes quantias de dados não rotulados, cuja classificação pode ser útil, por exemplo, para fins de monetização e publicidade. Da mesma forma, a tecnologia de detecção de *spam* em mensagens de e-mail usa métodos de aprendizagem não supervisionada[20].

Finalmente, a aprendizagem por reforço, de uso menos comum, consiste num modelo comportamental, em que o algoritmo interage como *datasets* que não são fixos e recebe o *feedback* da sua análise de dados, orientando o usuário para o melhor resultado. Aqui, o sistema não é treinado com um conjunto de dados específicos, aprendendo por meio de tentativa e erro. Os acertos, ao resolverem o problema, são reforçados pelo usuário.

3. Algoritmos, direito à privacidade e proteção de dados

A primeiras preocupações levantadas a partir do uso de algoritmos no processamento de dados recaíram sobre o direito à privacidade. Há décadas, em pioneiro artigo publicado em 1970, Arthur J. Sills já havia registrado que, já naquele momento, sob o aspecto técnico, "não haveria limites para a precisão e a extensão pela qual informações podem ser reunidas, guardadas, coletadas e disseminadas por sistemas de processamento de dados em massa"[21].

De fato, no que concerne aos dados pessoais, algoritmos são capazes de ampliar ao menos cinco capacidades, especialmente no âmbito virtual: *identificação*, assim compreendida a atividade de reconhecimento de determinadas informações como dados

19. IBM, 2020; PEIXOTO, SILVA, 2019, p. 20-21; ALPAYDIN, 2016.

20. IBM, 2020.

21. SILLS, 1970.

relacionados a pessoas naturais ou jurídicas; *individualização*, consistente no reconhecimento de determinada informação como dados relacionados a pessoas específicas; *coleta*, assim compreendida a possibilidade de, por meio de técnicas variadas, reunir, de forma automatizada, os dados identificados; *armazenamento,* consistente na guarda desses dados em locais específicos, permitindo o seu acesso posterior; e *disseminação* desses dados a pessoas interessadas.

Daí podem resultar possíveis violações ao direito à privacidade, cuja proteção encontra previsão em diversas constituições, bem como no art. 8º da Convenção Europeia de Direitos Humanos e no art. 11 da Convenção Interamericana de Direitos Humanos.

Algoritmos costumam ser utilizados com o objetivo de rastrear o tráfego *online* e fixar um perfil de cada usuário, sobretudo por meio dos metadados registrados em *cookies*. Até mesmo impressões digitais e registros de buscas podem ser secretamente monitorados por algoritmos, especialmente aqueles presentes em aplicações de *smartphones*. Eventualmente, todos esses dados coletados podem vir a ser utilizados como evidências digitais em processos criminais.

De acordo com pesquisa publicada em 2015 por Wu Youyou, Michal Kosinski, e David Stillwell, das universidades de Cambridge e Stanford, as inferências que os algoritmos de inteligência artificial podem fazer a respeito de uma determinada pessoa conseguem ser até mesmo mais precisas que aquelas realizadas por amigos e familiares[22].

Além disso, pesquisa realizada em 2012, na Universidade de Berkeley, concluiu que o uso de algoritmos invasivos de monitoramento oculto aumentou sensivelmente nos últimos anos, em resposta à prática crescente que alguns usuários têm adotado de apagar ou desabilitar *cookies* em navegadores[23].

Conforme compreendido em estudo do Conselho da Europa, o uso de dados reunidos em perfis criados por algoritmos afeta diretamente o direito informacional e de autodeterminação[24]. Além disso, as informações obtidas podem ser utilizadas de forma prejudicial aos indivíduos, servindo de substrato em processos de contratação, persecução penal e até mesmo de substrato para a manipulação de eleições.

4. Os riscos decorrentes do uso de algoritmos de inteligência artificial no campo da persecução penal

4.1. Prevenção criminal e devido processo legal

Em 2018, Marielle Francisco da Silva, vereadora no município do Rio de Janeiro conhecida como Marielle Franco, foi assassinada a tiros na sua cidade. Junto com ela estava seu motorista, Anderson Pedro Mathias Gomes, também vítima do crime. As

22. Youyou; Kosinski; Stillwell, p. 1036-1040, 2015.

23. HOOFNAGLE; SOLTANI; GOOD; WAMBACH; AYENSON, 2012.

24. CONSELHO DA EUROPA, 2018.

circunstâncias do delito o destacaram imediatamente nos meios de comunicação em massa.

Após uma complexa investigação cibernética conduzida pelo Ministério Público e pela Polícia Civil do Estado do Rio de Janeiro, os executores do crime foram rapidamente identificados. Uma das técnicas investigativas utilizadas foi a requisição de dados ao Google, para que informasse os usuários que teriam buscado informações sobre a vereadora na véspera do crime. Mediante prévia autorização judicial, o Google foi obrigado a entregar a lista dos IPs e Device IDs de usuários que pesquisaram as combinações de palavras "Marielle Franco", "Vereadora Marielle", "Agenda vereadora Marielle", "Casa das Pretas", "Agenda vereadora Marielle" e "Rua dos Inválidos", entre os dias 7 de 14 de março de 2018 – ocasião em que a vereadora e o seu motorista foram assassinados[25].

Essa decisão levou o Google e recorrer ao Superior Tribunal de Justiça (STJ). A sua principal irresignação dizia respeito à abrangência da medida, na medida em que as autoridades de investigação não individualizaram os investigados, requisitando informações sobre pessoas em geral que tivessem buscado tais palavras.

A Terceira Seção do STJ, em agosto de 2020, manteve a decisão que determinou à empresa o fornecimento de informações de usuários de seus serviços no âmbito das investigações. Segundo compreendeu o ministro Rogerio Schietti Cruz, relator, a quebra do sigilo de dados armazenados, de forma autônoma ou associada a outras informações pessoais, não necessita da prévia indicação das pessoas que estão sendo investigadas, na medida em que o objetivo da medida é justamente proporcionar a identificação de usuários do serviço ou de terminais utilizados[26].

Casos assim revelam os desafios jurídicos no campo das investigações cibernéticas, em que, cada vez mais, complexas ferramentas tecnológicas são desenvolvidas.

Para além da quebra de sigilo de dados, procedimento normalmente utilizado mediante prévia autorização judicial e com o objetivo específico de desvendar um ilícito já ocorrido (dimensão *retrospectiva*), novas técnicas têm sido desenvolvidas para uso também no campo *preditivo,* ou seja, com o objetivo de prevenir ilícitos (dimensão *prospectiva*) ou conhecer delitos ainda desconhecidos.

Dois conceitos são fundamentais para a compreensão dos algoritmos de prevenção criminal: correspondência de dados (*data matching*) e mineração de dados (*data mining*).

Entende-se por *data matching* o conjunto de técnicas de comparação automatizada de dois ou mais bancos de dados pessoais. Também conhecidas como *subject-based inquiries*, essas técnicas iniciam com um sujeito previamente identificado, na tentativa de construção de um perfil completo a partir das suas atividades, relação com outros indivíduos, locais ou eventos. No setor público, elas auxiliam na identificação de fraudes e abusos na obtenção de benefícios assistenciais[27]. Os resultados do *data matching* podem ser utili-

25. GLOBO, 2020.

26. SUPERIOR TRIBUNAL DE JUSTIÇA, 2020.

27. STEINBOCK, 2005, p. 10-11.

zados de forma *reativa*, de modo a revelar eventos passados – a exemplo da comparação da análise dos dados profissionais, fiscais e patrimoniais de um determinado indivíduo, que podem indicar enriquecimentos ilícitos.

Diversamente, a mineração de dados (*data mining*) é um conjunto de técnicas que têm como foco padrões qualitativos e comportamentais utilizados em julgamentos *preditivos*. Essas técnicas utilizam análise estatística e modelagem, de forma a identificar padrões em dados que permitam inferir regras capazes de prever resultados futuros. O *data mining* é também conhecido como *knowlege discovery, pattern-matching* ou *dataveillance*[28]. A sua origem remonta às práticas comerciais de individualização de marketing, bem como às ferramentas de definição de riscos na indústria dos seguros.

Em relatório publicado no ano de 2004, o *U.S. Government Accountability Office* (GAO) identificou o uso de 199 (cento e noventa e nove) ferramentas de *data mining* nas agências e departamento federais, indicando os seguintes objetivos principais: melhoria do serviço, detecção de fraudes ou abusos, análise de informações científicas e de pesquisa, gestão de recursos humanos, detecção de atividades ou padrões criminosos, análise de inteligência e detecção de atividades terroristas:

Figura 2 – Principais seis propósitos do uso de mineração de dados em departamentos e agências americanas

Table 1: Top Six Purposes of Data Mining Efforts in Departments and Agencies and Number of Efforts Reported

Department or agency	Improving service or performance	Detecting fraud, waste, and abuse	Analyzing scientific and research information	Managing human resources	Detecting criminal activities or patterns	Analyzing intelligence and detecting terrorist activities
Department of Agriculture	8	1				
Department of Commerce						
Department of Defense	19	1	1	14	1	5
Department of Education	6	9			3	1
Department of Energy					3	
Department of Health and Human Services	4		1			1
Department of Homeland Security	5			2	2	4
Department of the Interior	1					
Department of Justice	1			1	3	3
Department of Labor	3	1				
Department of State		2				
Department of Transportation		1				
Department of the Treasury	4	1			2	
Department of Veterans Affairs	5	5			1	
Environmental Protection Agency		1				
Export-Import Bank of the United States	1					
Federal Deposit Insurance Corporation	1					
Federal Reserve System		1				
National Aeronautics and Space Administration	1	1	21			
Nuclear Regulatory Commission	1					
Office of Personnel Management	1					
Pension Benefit Guaranty Corporation	2					
Railroad Retirement Board	1					
Small Business Administration	1					
Total	**65**	**24**	**23**	**17**	**15**	**14**

Source: GAO analysis of agency-provided data.

Fonte: U.S. GOVERNMENT ACCOUNTABILITY OFFICE[29].

28. STEINBOCK, 2005, p. 13.

29. U.S. Government Accountability Office, 2004.

Nos Estados Unidos, especialmente após o decreto do *Patriot Act,* assinado logo após o atentado de 11 de setembro de 2001, o uso de técnicas de *data matching* e *data mining* para fins *preventivos* – em contraste com as práticas reativas – cresceu sensivelmente. A título de exemplo, no ano de 2003, foi criado o *Terrorist Screening Center* (TSC), agência vinculada ao *Federal Bureau of Investigation* (FBI), responsável por consolidar uma *watchlist*, banco de dados contendo nome de suspeitos de participarem de atividades terroristas[30].

Além disso, novos algoritmos têm sido utilizados não apenas para auxiliar investigações, mas também para fins *decisórios* – o que pode variar desde o impedimento de empréstimo de livros em bibliotecas públicas até a proibição de viagens aéreas a passageiros suspeitos (*passenger profiling* e *no-fly list*[31]). Há, portanto, desde 2001, um sensível movimento na análise de dados: da reação à prevenção e da investigação à decisão.

Embora seja inegável o salto de eficiência proporcionado por algoritmos no campo da prevenção criminal, não são poucas as pesquisas acadêmicas que têm levantado preocupações, especialmente no que concerne à discriminação em desfavor de minorias.

Em artigo publicado em 2017, Danielle Kehl, Priscilla Guo e Samuel Kessler registram que, por um lado, os algoritmos preditivos podem trazer maior objetividade nos processos criminais, reduzindo o risco de erros humanos. Por outro, existe o risco de que, mal-utilizados, possam reforçar vieses e prejudicar a aplicação da justiça, sobretudo se a raça, o gênero e a posição social de indivíduos forem considerados como variáveis preditivas[32].

Em trabalho publicado em 2019, Sandra Mayson defende que, no que diz respeito aos vieses raciais, o problema reside na própria natureza da *previsão.* Todos os sistemas preditivos observam o passado para tentar adivinhar futuros eventos. Assim, em um mundo racialmente estratificado, qualquer método de predição projetará essas desigualdades do passado para o futuro[33].

Uma possível forma de harmonização dessas técnicas preditivas com os direitos fundamentais consiste na sua permeabilidade à cláusula do devido processo legal, como sugerido por Daniel Steinbock, sempre que utilizadas para fins de privação da liberdade ou da propriedade[34].

Daí derivariam quatro providências essenciais.

A primeira consiste na realização de audiências sumárias antes de decisões que afetem a propriedade ou a liberdade individual, a exemplo da privação do direito de acesso a transportes aéreos. A segunda consiste na abertura de oportunidades de correção após uma primeira consequência derivada da aplicação de técnicas de *data matching* ou *data*

30. FBI, 2020.

31. O'CONNOR; RUMANN, 2003.

32. Kehl; Guo; Kessler, 2017.

33. MAYSON, 2019.

34. STEINBOCK, 2005, p. 83.

mining. A terceira, na possibilidade de responsabilização por resultados do tipo "falso-positivo", indenizando-se as vítimas de constrangimentos indevidos. Finalmente, em razão da necessidade de preservação do sigilo dos dados existentes em bancos de dados de órgãos de persecução penal, o seu conteúdo poderia ser avaliado por instituições independentes (*independent oversight bodies*), permitindo-se assim o controle social.

4.2. Vieses cognitivos e decisões judiciais

Há muito, o dogma da ação racional, elemento central da economia clássica, tem sido confrontado por experimentos no campo da psicologia cognitiva[35] e da economia comportamental (*behavioral economics*)[36], especialmente a partir da pesquisa seminal realizada na década de 1970, por Amos Tversky e Daniel Kahneman[37].

O escopo do trabalho dos aludidos autores consistiu no processo decisório humano em ambientes de incertezas. Isso porque "muitas decisões são baseadas em crenças relativas à probabilidade de eventos incertos, a exemplo do resultado de uma eleição, a culpabilidade de um réu ou o valor futuro do dólar"[38]. Os experimentos conduzidos por Tversky e Kahneman foram capazes de revelar que, durante tais processos decisórios, muitas pessoas se valem de um limitado número de princípios heurísticos automatizados.

Esses princípios heurísticos consistem em atalhos mentais[39] ou estratégias inconscientes – heranças do processo evolutivo humano –, meios utilizados na redução da complexidade das tarefas de conhecimento de situações e tomadas de decisões. Eles funcionam como algoritmos cognitivos que permitem que as decisões humanas sejam mais eficientes. Para tanto, são realizados processos inconscientes de categorização de objetos, pessoas e suas ocorrências em grupos ou tipos.

Além de adaptativas, as heurísticas são inegavelmente úteis, contribuindo para uma maior eficiência nas decisões diárias, notadamente aquelas que integram o chamado "Sistema 1" de pensamento, em que os julgamentos e escolhas são mais intuitivos, experimentais e automatizados[40].

Apesar disso, elas podem resultar em erros severos e sistemáticos[41], conduzindo a falácias, vieses e ilusões. Cuida-se do "preço que pagamos para tamanha eficiência"[42].

35. NUNES; LUD; PEDRON, 2018, p. 10; FENOLL, 2019, p. 24.
36. DYSON, 2011.
37. TVERSY; KAHNEMAN, 1973, p. 207-232.
38. TVERSY; KAHNEMAN, 1974, p. 1.124-1.131.
39. PEER; GAMLIEL, 2013, p. 114-119; JONES, 2013, p. 49-122; JONES; RANKIN, 2014.
40. KAHNEMAN, 2012.
41. SUNSTEIN, 2000; NUNES; LUD, 2018; FENOLL, 2019; COSTA, 2018.
42. NEGOWETTI, 2014.

Existem fortes experimentos que demonstram, empiricamente, a existência de vieses implícitos capazes ensejar a tomada de decisões irracionais, muitas delas em desfavor de minorias[43].

Conforme dados apresentados pelo Georgia Department of Corrections[44], não apenas os negros recebem sentenças 4,25% mais altas do que os brancos, mas os negros de pele média e escura também recebem sentenças 4,8% mais altas do que as dos brancos. Por outro lado, os negros de pele clara recebem sentenças quase da mesma severidade que os brancos. Pesquisadores também concluíram que, em casos que envolvem uma vítima branca, quanto mais fenotipicamente negro o réu era considerado, maior era a probabilidade de ele ser condenado à morte[45].

De igual modo, outros estudos apontam que juízes tendem a fixar fianças consideravelmente mais elevadas, em acréscimo de 25% (vinte e cinco por cento), bem como punições cerca de 12% (doze por cento) mais longas em desfavor de réus negros em situações similares a réus brancos[46].

Recentes estudos demonstram até mesmo fatores aparentemente irrelevantes são considerados em processos heurísticos decisórios. A título de exemplo, juízes levados a refletir sobre sua própria morte apresentaram maior probabilidade de proferir decisões mais conservadoras. Do mesmo modo, juízes em tribunais que se encontravam temporalmente mais distantes da sua última refeição apresentaram maior probabilidade de confirmar a decisão recorrida. Até mesmo o momento das alegações finais e as vestimentas dos advogados foram identificados como fatores relevantes no processo decisório[47].

Daí surge o seguinte questionamento: o uso de algoritmos decisórios é capaz de reforçar os vieses implícitos dos julgadores?

Em alguns casos, o seu uso não se resume à mera classificação processual, auxiliando diretamente o processo decisório. A título de exemplo, o *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS) consiste em um mecanismo utilizado nos Estados Unidos para avaliar o risco de reincidência dos acusados no país, auxiliando os juízes a decidirem a respeito das prisões cautelares[48].

O COMPAS já foi apontado como um sistema que produz resultados discriminatórios. Segundo artigo publicado por autores vinculados à ProPublica – instituição sem fins lucrativos sediada em Nova York que tem por missão a realização de jornalismo investigativo de interesse público –, o algoritmo utilizado tende a classificar erroneamente

43. PEREIRA; ÁLVARO; OLIVEIRA; DANTAS, 2011, p. 144-153; ROOTH, 2007; GOFF; JACKSON; DI LEONE; CULOTTA, DITOMASSO, 2014, p. 526-545.

44. BURCH, 2015, p. 395-420.

45. PARKS; DAVIS, 2016; EBERHARDT; DAVIES; PURDIE-VAUGHNS; JOHNSON, 2006, p. 383-386.

46. WALDFOGEL; AYERS, 1994, p. 987-1048.

47. JONES, 2013.

48. NUNES; MARQUES, 2018, p. 421-447.

acusados negros como prováveis reincidentes, além de enquadrar, também de forma equivocada, acusados brancos como indivíduos com baixo risco de reincidência.

Demais disso, a empresa desenvolvedora do *software* – Northpointe – não disponibiliza ao público o algoritmo no qual se baseia o índice de reincidência do acusado, limitando-se a divulgar as perguntas feitas aos acusados, cujas respostas são utilizadas no cálculo. Consequentemente, o acusado não é capaz de conhecer as razões pelas quais ostenta um alto ou baixo indicador, tampouco a forma com que suas respostas afetam o resultado final[49].

Situações assim revelam a preocupação relevante no campo dos processos judiciais, em que ferramentas de automação decisória, ao "aprender" com os dados dispostos, seriam capazes incorporar os mesmos vieses da pessoa a quem buscam auxiliar, absorvendo, seja nas regras ou nos dados, os vícios já existentes.

5. Regulação algorítmica

5.1. O estado da arte na União Europeia

Em maio de 2018, entrou em vigor, na União Europeia, o Regulamento 2016/679, relativo à proteção das pessoas físicas no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados. O diploma tem por objeto o processamento automatizado e semiautomatizado de dados pessoais, não se aplicando, todavia, entre outros campos, ao tratamento efetuado pelas autoridades competentes para efeitos de prevenção, investigação, detecção e repressão de infrações penais ou da execução de sanções penais, "incluindo a salvaguarda e a prevenção de ameaças à segurança pública"[50].

O diploma, além de ter um objeto bastante específico – proteção de dados pessoais –, apresenta limitações no âmbito de abrangência, em especial no caso de tratamento de dados com finalidade não econômica, o que inclui atividades estatais. Ainda assim, tendo em vista a preocupação comum no tratamento de dados pessoais, existem aspectos comuns que devem ser analisados, em relação à regulação do uso de algoritmos.

Nesse sentido, são estabelecidos princípios relativos ao tratamento de dados pessoais, destacando-se a transparência e a lealdade. Estabelece o art. 5º do Regulamento (UE) 2016/679:

> "Artigo 5º
>
> Princípios relativos ao tratamento de dados pessoais
>
> 1. Os dados pessoais são:

49. LARSON; MATTU; LAUREN; ANGWINHOW, 2016.

50. PARLAMENTO EUROPEU E CONSELHO DA UNIÃO EUROPEIA, 2016.

a) Objeto de um tratamento lícito, leal e transparente em relação ao titular dos dados («licitude, lealdade e transparência»);

b) Recolhidos para finalidades determinadas, explícitas e legítimas e não podendo ser tratados posteriormente de uma forma incompatível com essas finalidades; o tratamento posterior para fins de arquivo de interesse público, ou para fins de investigação científica ou histórica ou para fins estatísticos, não é considerado incompatível com as finalidades iniciais, em conformidade com o artigo 89º, n. 1 («limitação das finalidades»);

c) Adequados, pertinentes e limitados ao que é necessário relativamente às finalidades para as quais são tratados («minimização dos dados»);

d) Exatos e atualizados sempre que necessário; devem ser adotadas todas as medidas adequadas para que os dados inexatos, tendo em conta as finalidades para que são tratados, sejam apagados ou retificados sem demora («exatidão»);

e) Conservados de uma forma que permita a identificação dos titulares dos dados apenas durante o período necessário para as finalidades para as quais são tratados; os dados pessoais podem ser conservados durante períodos mais longos, desde que sejam tratados exclusivamente para fins de arquivo de interesse público, ou para fins de investigação científica ou histórica ou para fins estatísticos, em conformidade com o artigo 89º, n. 1, sujeitos à aplicação das medidas técnicas e organizativas adequadas exigidas pelo presente regulamento, a fim de salvaguardar os direitos e liberdades do titular dos dados («limitação da conservação»);

f) Tratados de uma forma que garanta a sua segurança, incluindo a proteção contra o seu tratamento não autorizado ou ilícito e contra a sua perda, destruição ou danificação acidental, adotando as medidas técnicas ou organizativas adequadas («integridade e confidencialidade»);

2. O responsável pelo tratamento é responsável pelo cumprimento do disposto no n. 1 e tem de poder comprová-lo («responsabilidade»)".

Tais regramentos, que seguem uma linha comum em outros países, são certamente aplicáveis às relações privadas marcadas pela presença de algoritmos – ex.: avaliação de crédito, redes sociais, contratação de trabalhadores, individualização de marketing *on-line* etc.) –, na medida em que utilizados dados pessoais.

Especificamente quanto ao uso de ferramentas de inteligência artificial e automação decisória, a Comissão Europeia publicou, em outubro de 2017, as orientações gerais relativas à aplicação do Regulamento (UE) 2016/679 – *Guidelines on Automated individual decision-making and Profiling*[51]. De acordo com as orientações, ao titular dos dados devem ser adequadamente assegurados, entre outros, os direitos à informação, acesso, retificação e objeção. Esse documento foi seguido das comunicações COM(2018)237 e

51. COMISSÃO EUROPEIA, 2016.

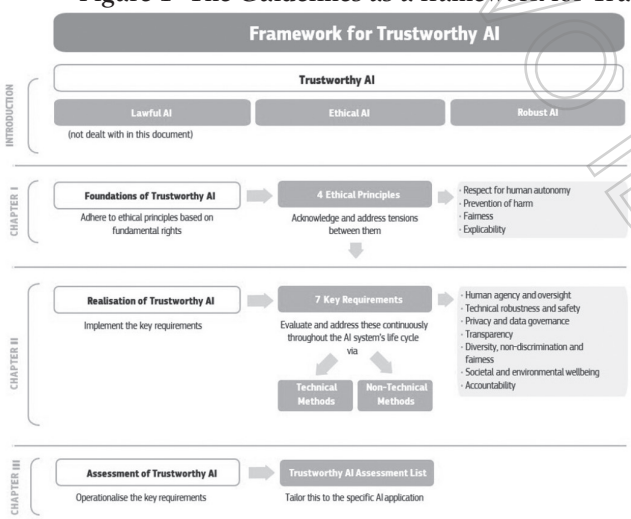
COM(2018)795, que apresentaram um plano coordenado do continente para o desenvolvimento tecnológico tendo por objeto a inteligência artificial.

Em abril de 2019, dando seguimento ao seu plano coordenado, a Comissão publicou um documento intitulado *Ethics Guidelines for Trustworthy AI,* elaborado por um grupo de experts independentes reunidos por ela[52]. O documento tem por objetivo estabelecer princípios éticos fundamentais para que a maximização dos benefícios decorrentes do desenvolvimento de ferramentas de inteligência artificial seja acompanhada da prevenção e da diminuição dos riscos inerentes. Assim, para que se possa falar em uma inteligência artificial "confiável" e "fidedigna", são estabelecidos três componentes: a) ela deve ser *legal,* respeitando a legislação aplicável; b) ela deve ser ética, garantindo a adesão a princípios e valores éticos; c) ela deve ser *robusta*, tanto do ponto de vista técnico quanto social, na medida em que, mesmo com boas intenções, podem ser causados danos não intencionais.

Para que os três componentes das orientações éticas sejam observados, são desenvolvidos quatro princípios: a) respeito pela autonomia humana; b) prevenção de danos; c) justiça; d) explicação.

Por fim, as *Ethics Guidelines* consolidam requerimentos mínimos (*key requirements*) para a efetivação dos princípios éticos: a) agência humana e supervisão; b) robustez técnica e segurança; c) privacidade e governança de dados; d) transparência; e) diversidade, não discriminação e justiça; f) bem-estar social e ambiental; g) *accountability*.

Figure 1- The Guidelines as a framework for Trustworthy AI



Acesse para visualizar a imagem colorida

Fonte: COMISSÃO EUROPEIA[53].

52. COMISSÃO EUROPEIA, 2019.

53. COMISSÃO EUROPEIA, 2018.

Posteriormente, em fevereiro de 2020, a Comissão Europeia publicou seu *White Paper* sobre o tema: *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*[54]. Nele é demonstrada a necessidade de regulação de ferramentas de inteligência artificial não apenas no setor privado, mas também no campo público. O documento expõe, em sua introdução, a relevância do emprego de ferramentas de automação decisória de interesse público, capazes de reduzir os custos dos serviços prestados, promover sustentabilidade e equipar autoridades de persecução penal, entre outros aspectos. São também expostas as preocupações concernentes à necessária proteção dos direitos fundamentais.

Além de prever a aplicação imediata da legislação europeia relacionada à segurança e responsabilização, incluindo regras específicas de cada setor e as complementações nacionais, o *White Paper* estabelece ainda o escopo para um futuro marco regulatório comunitário sobre inteligência artificial. A Comissão apresenta sua opinião no sentido de que a ferramenta de inteligência artificial deve ser considerada de alto risco – e, portanto, sujeita a uma disciplina diferenciada – quando presentes os seguintes pressupostos cumulativos: a) aplicação em um setor em que, dadas as características, significantes riscos são esperados (transporte, energia, partes do setor público etc.); b) a aplicação da ferramenta no setor é realizada de uma forma tal que potencialize a ocorrência de riscos.

5.2. O estado da arte no Brasil

De forma similar ao ocorrido no âmbito europeu em 2016, a Lei 13.709/2018 disciplina, no Brasil, o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural.

De acordo com o seu art. 4º, essa Lei não se aplica ao tratamento de dados pessoais realizado para fins particulares e não econômico, para fins jornalísticos, artísticos, acadêmicos, bem como para fins de segurança pública, defesa nacional, segurança do Estado ou atividades de investigação e repressão de infrações penais. Há, portanto, uma relevante lacuna do tema no que diz respeito à persecução penal.

Por sua vez, o art. 6º da Lei 13.709/2018 estabelece um rol de princípios, entre eles a transparência, a segurança e a não discriminação:

> "Art. 6º As atividades de tratamento de dados pessoais deverão observar a boa-fé e os seguintes princípios:
>
> I – finalidade: realização do tratamento para propósitos legítimos, específicos, explícitos e informados ao titular, sem possibilidade de tratamento posterior de forma incompatível com essas finalidades;

54. COMISSÃO EUROPEIA, 2020.

> II – adequação: compatibilidade do tratamento com as finalidades informadas ao titular, de acordo com o contexto do tratamento;
>
> III – necessidade: limitação do tratamento ao mínimo necessário para a realização de suas finalidades, com abrangência dos dados pertinentes, proporcionais e não excessivos em relação às finalidades do tratamento de dados;
>
> IV – livre acesso: garantia, aos titulares, de consulta facilitada e gratuita sobre a forma e a duração do tratamento, bem como sobre a integralidade de seus dados pessoais;
>
> V – qualidade dos dados: garantia, aos titulares, de exatidão, clareza, relevância e atualização dos dados, de acordo com a necessidade e para o cumprimento da finalidade de seu tratamento;
>
> VI – transparência: garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial;
>
> VII – segurança: utilização de medidas técnicas e administrativas aptas a proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou difusão;
>
> VIII – prevenção: adoção de medidas para prevenir a ocorrência de danos em virtude do tratamento de dados pessoais;
>
> IX – não discriminação: impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos;
>
> X – responsabilização e prestação de contas: demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento das normas de proteção de dados pessoais e, inclusive, da eficácia dessas medidas."

Passados dois anos, o Conselho Nacional de Justiça, órgão responsável pelo controle da atuação administrativa e financeira do Poder Judiciário e do cumprimento dos deveres funcionais dos juízes, editou a Resolução 332, de 21 de agosto de 2020. O ato tem por objeto "a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário".

Trata-se, portanto, de diploma voltado, de forma específica, à disciplina do emprego de ferramentas de automação alimentadas por inteligência artificial no âmbito do Poder Judiciário.

Entre as suas considerações iniciais, encontra-se a de que

> "as decisões judiciais apoiadas pela Inteligência Artificial devem preservar a igualdade, a não discriminação, a pluralidade, a solidariedade e o julgamento justo, com a viabilização de meios destinados a eliminar ou minimizar a opressão, a marginalização do ser humano e os erros de julgamento decorrentes de preconceitos; [...]"

Os seus dispositivos introdutórios ("disposições gerais") apresentam os conceitos fundamentais para a disciplina do tema, entre eles o de algoritmo:

> "Art. 3º Para o disposto nesta Resolução, considera-se:
>
> I – Algoritmo: sequência finita de instruções executadas por um programa de computador, com o objetivo de processar informações para um fim específico;
>
> II – Modelo de Inteligência Artificial: conjunto de dados e algoritmos computacionais, concebidos a partir de modelos matemáticos, cujo objetivo é oferecer resultados inteligentes, associados ou comparáveis a determinados aspectos do pensamento, do saber ou da atividade humana;
>
> III – Sinapses: solução computacional, mantida pelo Conselho Nacional de Justiça, com o objetivo de armazenar, testar, treinar, distribuir e auditar modelos de Inteligência Artificial;
>
> IV – Usuário: 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;
>
> V – Usuário interno: membro, servidor ou colaborador do Poder Judiciário que desenvolva ou utilize o sistema inteligente;
>
> VI – Usuário externo: pessoa que, mesmo sem ser membro, servidor ou colaborador do Poder Judiciário, utiliza ou mantém qualquer espécie de contato com o sistema inteligente, notadamente jurisdicionados, advogados, defensores públicos, procuradores, membros do Ministério Público, peritos, assistentes técnicos, entre outros."

Em seguida o ato estabelece a necessidade de respeito aos direitos fundamentais (Capítulo II), passando a dispor, no seu Capítulo III, sobre o princípio da não discriminação:

> "CAPÍTULO III
>
> DA NÃO DISCRIMINAÇÃO
>
> Art. 7º As decisões judiciais apoiadas em ferramentas de Inteligência Artificial devem preservar a igualdade, a não discriminação, a pluralidade e a solidariedade, auxiliando no julgamento justo, com criação de condições que visem eliminar ou minimizar a opressão, a marginalização do ser humano e os erros de julgamento decorrentes de preconceitos. [...]"

Uma importante medida preventiva é disciplinada no art. 7º, § 1º, da Resolução: antes de ser colocado em produção, é necessário que o modelo de inteligência artificial seja homologado, de forma a identificar se preconceitos ou generalizações influenciaram seu desenvolvimento, acarretando tendências discriminatórias no seu funcionamento. Identificado o viés discriminatório, a utilização será impedida, com o registro do projeto das razões que levaram à decisão.

Essa preocupação alcança até mesmo a composição de equipes para pesquisa, desenvolvimento e implantação das soluções computacionais que se utilizem de inteligência artificial, que deverá buscar a "diversidade em seu mais amplo espectro, incluindo gênero, raça, etnia, cor, orientação sexual, pessoas com deficiência, geração e demais características individuais" (art. 20, *caput*).

Além disso, a participação representativa deverá existir "em todas as etapas do processo, tais como planejamento, coleta e processamento de dados, construção, verificação, validação e implementação dos modelos, tanto nas áreas técnicas como negociais" (art. 20, § 1º).

No que diz respeito à publicidade e transparência, são estabelecidos, no art. 8º da Resolução, alguns deveres, entre eles: a) a divulgação responsável; b) a indicação dos objetivos e resultados pretendidos pelo modelo; c) a documentação dos riscos e indicação de instrumentos de informação e controle; d) apresentação dos mecanismos de auditoria e certificação de boas práticas, além do; e) fornecimento de explicação satisfatória e passível de auditoria por autoridade humana quanto a qualquer proposta de decisão apresentada pelo modelo, especialmente quando essa for de natureza judicial.

O direito à explicação satisfatória há de ser concretizado em linguagem clara e precisa, quanto à utilização de sistema inteligente nos serviços que forem prestados (art. 18).

Além disso, duas importantes limitações foram introduzidas.

A primeira diz respeito ao controle do usuário. De acordo com o art. 17 da Resolução CNJ 332/2020, o sistema inteligente deverá assegurar a autonomia dos usuários internos, com o uso de modelos que proporcionem o incremento, e não a restrição. No presente momento, não é possível o desenvolvimento de ferramentas de automação que eliminem a autonomia decisória humana no âmbito do Poder Judiciário brasileiro.

Desse modo, na hipótese de o modelo apresentar uma proposta de decisão, é necessário que o sistema permita a revisão da proposta e dos dados utilizados para sua elaboração, sem que haja vinculação à solução apresentada pela inteligência artificial.

Uma segunda – e polêmica – limitação reside no art. 23, ao estabelecer que a utilização de modelos de inteligência artificial em matéria penal não deve ser estimulada, sobretudo com relação à sugestão de modelos de decisões preditivas.

Essa limitação é flexibilizada pelo § 1º, que a afasta quando se tratar de utilização de soluções computacionais destinadas à automação e ao oferecimento de subsídios destinados ao cálculo de penas, prescrição, verificação de reincidência, mapeamentos, classificações e triagem dos autos para fins de gerenciamento de acervo.

De forma semelhante, o § 2º do art. 23 dispõe que os modelos de inteligência artificial destinados à verificação de reincidência penal não devem indicar conclusão mais prejudicial ao réu do que aquela a que o magistrado chegaria sem sua utilização.

5.3. Uma proposta de tripé normativo fundamental

Excluídos os princípios aplicáveis, de forma geral, às relações públicas (boa-fé, legalidade, transparência, não discriminação etc.), uma análise conjunta dos riscos do uso

de algoritmos decisórios e das normas que já disciplinam o processamento de dados pessoais permite à definição de um esboço normativo próprio ao uso de algoritmos no ambiente estatal, o que inclui a persecução penal.

Como exposto, as preocupações relativas à violação de direitos fundamentais pelo uso de algoritmos decorrem, em especial, de três achados: a opacidade dos algoritmos, o emprego de *inputs* viciados e a discriminação que pode ser gerada com o seu uso.

Uma proposta de regime jurídico de mínima regulação algorítmica há de, ao menos, considerar esses três problemas, possibilitando a sua correção.

5.3.1. Princípio da auditabilidade

Entende-se por auditabilidade a capacidade de preservação de todo o conjunto de informações utilizadas na cadeia de funcionamento de uma determinada ferramenta. Sem auditabilidade, qualquer ferramenta se transformará em uma caixa-preta, impossibilitando tanto o conhecimento da sua funcionalidade quanto a transparência[55].

Em se tratando de algoritmos de automação decisória, essas informações compreendem ao menos três categorias[56].

A primeira categoria consiste no registro de todos os dados abstratos que são levados em consideração pelo algoritmo (*inputs* e dados estatísticos abstratamente utilizados). A título exemplificativo, em uma ferramenta voltada à classificação de usuários de serviços públicos em ordem prioritária, é necessário saber quais dados são levados em consideração (estatísticas relativas à idade, enfermidade, renda etc.). De igual modo, em se tratando de ferramenta voltada à análise da periculosidade de investigados ou réus em processos criminais, para fins de prisão preventiva, é importante saber que tipo de dado é utilizado (estatísticas relativas à reincidência, circunstâncias do crime etc.).

A segunda categoria consiste nas informações específicas utilizadas numa decisão concreta (*inputs* concretamente utilizados). No segundo exemplo utilizado supra, seria o caso de ser utilizado o histórico prisional de um determinado réu.

A terceira categoria consiste na regra do algoritmo. É necessário ter, em registro, toda a cadeia lógica e matemática utilizada para, mediante processamento dos dados de entrada (*inputs*), ser produzido o resultado (*output*). Essa terceira categoria informacional é particularmente potencializada no âmbito público, por decorrência do princípio do devido processo legal. Uma forma de "armazenar" um algoritmo consiste no registro da cadeia completa de códigos utilizada para a sua implementação. Embora normalmente seja bastante difícil para alguém compreender as linhas de programação, esse registro possibilita o controle por agências independentes. Além, disso, é possível o emprego de

55. BATHAEE, 2018 p. 889-938; YANISKY-RAVID; HALLISEY, 2019, p. 428-486.

56. VILLASENOR; FOGGO, 2020, p. 339-340.

pseudocode, comumente utilizado no campo da ciência da computação[57], de modo a registrar os algoritmos de uma forma mais compreensível, mediante técnicas representativas[58].

Uma particular observação diz respeito aos estágios evolutivos dos algoritmos de *machine learnig*. Levando-se em consideração a sua capacidade de aprendizado e "mutação", uma forma segura de registro das informações consiste numa espécie de "captura de imagem" (*snapshot*) cada vez que o algoritmo é acionado, permitindo a sua testagem futura.

Feitas essas considerações, resta o seguinte questionamento: quem deve guardar essas informações?

Uma primeira opção consiste em atribuir essa obrigação às empresas, órgãos ou instituições responsáveis pelo desenvolvimento da ferramenta. A vantagem está no reconhecimento de que aquele que desenvolveu o algoritmo é, presumidamente, quem detém maior conhecimento sobre o seu funcionamento. Consequentemente, também detém uma capacidade mais elevada de monitoramento e correção. Por outro lado, é possível que os dados a serem armazenados sejam sensíveis, envolvendo a privacidade de muitas pessoas. Além disso, o desenvolvedor do produto é, presumidamente, o menos interessado em revelar eventuais falhas encontradas.

Uma segunda opção consiste em atribuir a responsabilidade pelo armazenamento de informações às pessoas ou órgãos responsáveis pela sua aplicação. Essa parece ser uma forma mais segura de preservar a segurança e a privacidade dos dados, permitindo-se ao desenvolvedor o seu acesso parcial apenas em situações específicas, mediante tráfego criptografado, para fins de auditabilidade.

5.3.2. Princípio da transparência

Enquanto a auditabilidade diz respeito ao registro de informações, o princípio da transparência impõe que essas informações possam ser adequadamente acessadas e explicadas.

Da transparência, extraem-se duas obrigações: acesso (publicidade) e explicação.

Quanto ao acesso, uma objeção comum consiste no argumento voltado à proteção do segredo industrial. De fato, o acesso amplo e ilimitado a algoritmos desenvolvidos por empresas pode colocá-las em situação de desvantagem econômica em relação aos seus competidores no mercado comum. Por outro lado, a negativa de acesso em absoluto aos interessados pode representar uma grave violação ao princípio do devido processo legal. Basta imaginar o emprego de ferramentas de avaliação de risco utilizadas em processos criminais[59].

Esse impasse pode ser suficientemente resolvido de algumas formas.

57. BBC BITESIZE, 2020.

58. VILLASENOR; FOGGO, 2020, p. 341.

59. VILLASENOR; FOGGO, 2020, p. 343.

Uma delas consiste no emprego de *protective orders*, bastante comuns em processos civis em que discutidas supostas violações de patentes. Em casos assim, assistentes técnicos dos autores podem ter acesso ao código-fonte de aplicações, bem como a outros dados secretos, mediante o emprego de técnicas protetivas e obrigações específicas. Um exemplo consiste na disponibilização do código em um único terminal informático sem acesso à internet, em uma sala dedicada e na presença de representantes do desenvolvedor[60].

Uma outra forma consiste na disponibilização de dados apenas a instituições de checagem certificadas e a órgãos e instituições públicas, impondo-se o dever de manutenção da confidencialidade.

Para além do acesso (publicidade), a segunda dimensão da transparência consiste no direito à explicação.

Cuida-se de um desdobramento do *right to explanation*, registrado no Considerando nº 71 do Regulamento (UE) 2016/679 (*General Data Protection Regulation*)[61]. Embora o art. 22 do aludido diploma, ao disciplinar as decisões individuais automatizadas, não tenha mencionado o direito à explicação, o seu art. 15 assegura ao titular de dados, em caso de decisões automatizadas, o direito de acesso a "informações úteis relativas à lógica subjacente, bem como à importância e às consequências previstas de tal tratamento para o titular dos dados". Embora possa haver certa controvérsia quanto ao alcance do direito à explicação no âmbito das relações privadas, em se tratando de aplicações desenvolvidas ou aplicadas por instituições públicas, parece não haver dúvidas quanto à exigência, que decorre do devido processo legal.

O cumprimento do dever de explicação não se confunde com a mera disponibilização do código-fonte. Como registram Kroll et al., uma solução ingênua para a verificação da regularidade procedimental consiste em demandar transparência do código fonte, bem como dos *inputs* e *outputs* para relevantes decisões[62]. A publicidade, por si só, não é suficiente para promover *accountability*, sobretudo porque não explica o porquê de uma determinada decisão ter sido tomada. Para que se atinja esse benefício, a instituição ou órgão que utiliza determinado algoritmo deve fornecer uma explicação mínima sobre o seu funcionamento numa determinada situação concreta, em extensão suficiente a permitir o exercício do direito de defesa.

5.3.3. Princípio da consistência ou regularidade procedimental

Pelo princípio da consistência ou regularidade procedimental, é necessário assegurar que a cada destinatário ou usuário de uma determinada ferramenta construída a partir de

60. VILLASENOR; FOGGO, 2020, p. 344.

61. PARLAMENTO EUROPEU E CONSELHO DA UNIÃO EUROPEIA, 2016.

62. KROLL; HUEY; BAROCAS; FELTEN; REIDENBERG; ROBINSON, 2017, p. 657.

algoritmos seja aplicado o mesmo *procedimento*, bem como que esse procedimento não tenha sido desenvolvido de forma que lhe cause desvantagens a alguém em particular[63].

Cuida-se de uma decorrência não apenas do devido processo legal, mas também do princípio da isonomia. Além disso, para que seja realizado, o princípio da consistência depende da prévia auditabilidade do sistema. O mais importante aqui é o exame dos *inputs* e *outputs*, o que afasta as preocupações relativas à proteção de segredos industriais, próprias do princípio da transparência. Embora exista uma relação próxima com a "justiça" do algoritmo – algo inerente a qualquer atividade estatal –, o escopo da consistência é mais restrito.

A regularidade procedimental deve conduzir à consistência dos *outputs,* evitando que *inputs* similares produzam resultados discrepantes. Um exemplo é fornecido por Villasenor e Foggo: considerem-se dois diferentes réus com perfis idênticos quanto aos *inputs* específicos submetidos a um mesmo algoritmo de avaliação de risco (*risk assessments*). O primeiro é avaliado em março, o segundo em outubro. Embora seja possível a apresentação de resultados diversos em razão de melhoras na acurácia do algoritmo de inteligência artificial, aos réus que, numa visão retrospectiva, foram avaliados de forma mais severa, deve ser dada a possibilidade de conhecimento e de busca de correção[64].

Em uma ferramenta desprovida de algoritmos de inteligência artificial, *inputs* idênticos ou similares implicariam sempre o mesmo resultado. Com inteligência artificial, por outro lado, as técnicas de *machine learning* podem ensejar a "evolução" do algoritmo. Isso não significa que, em qualquer mudança algorítmica, todos os casos anteriores tenham que ser individualmente verificados, na medida em que essa forma de proceder poderia resultar na impraticabilidade do emprego da ferramenta. Em verdade, essa checagem pode ser feita também de forma automatizada – igualmente por meio de algoritmos auditáveis – e em casos de mudanças sensíveis proporcionadas pela inteligência artificial.

A consistência algorítmica é algo que pode ser avaliado por algoritmos de checagens. Ao invés de um ser humano realizar a checagem procedimental de cada caso, é possível que sistemas sejam desenvolvida para isso, reforçando a segurança de forma automatizada.

Kroll et al. registram que o atual estágio tecnológico da ciência da computação contempla ferramentas capazes de avaliar a regularidade procedimental, incorporando-se no desenho dos sistemas. De forma específica, essas ferramentas podem assegurar que: a) a mesma disciplina foi utilizada na construção de cada decisão; b) a disciplina foi integralmente especificada antes de os sujeitos envolvidos serem conhecidos, reduzindo-se a possibilidade de desenvolvimento de um procedimento em desfavor de alguém em particular; c) cada decisão é o produto das regras e dos *inputs* aplicados; d) se a decisão requer *inputs* randômicos, a sua escolha ocorreu para além do interesse de qualquer pessoa[65].

63. KROLL; HUEY; BAROCAS; FELTEN; REIDENBERG; ROBINSON; YU, 2017, p. 656.

64. VILLASENOR; FOGGO, 2020, p. 343.

65. KROLL; HUEY; BAROCAS; FELTEN; REIDENBERG; ROBINSON; YU, 2017, p. 657.

6. Considerações finais

Ao final do exposto, apresentam-se as seguintes conclusões, sem prejuízo de outras ilações realizadas ao longo do texto:

Em razão da possibilidade de aprender pela experiência, os algoritmos de IA são capazes de adaptação e evolução. Consequentemente, os sistemas podem criar as próprias regras ou perguntas, sem a necessidade de que um programador preveja soluções específicas para as situações possíveis. Isso se deve, sobretudo, às técnicas de *machine learning*.

A primeiras preocupações levantadas a partir do uso de algoritmos no processamento de dados recaíram sobre o direito à privacidade. Algoritmos são capazes de ampliar ao menos cinco capacidades, especialmente no âmbito virtual: identificação, individualização, coleta, armazenamento e disseminação desses dados a pessoas interessadas.

Além disso, algoritmos em buscadores são capazes de ocultar ou induzir informações de forma deliberada, favorecendo interesses políticos e econômicos específicos, afetando diretamente o exercício da liberdade de expressão.

Também o campo laboral pode ser negativamente afetado pelo uso de algoritmos. É o caso dos processos de contratação, avaliação e demissão de trabalhadores, que podem se valer de ferramentas automatizadas desprovidas de transparência e de dados estatísticos capazes de reforçar círculos viciosos de preconceito de raça, classe ou gênero.

No âmbito público, os riscos do emprego de algoritmos de inteligência artificial envolvem atividades como a prevenção criminal. Novas técnicas têm sido desenvolvidas para uso também no campo *preditivo,* ou seja, com o objetivo de prevenir ilícitos (dimensão *prospectiva*) ou conhecer delitos ainda desconhecidos. É o caso do *data matching* e da mineração de dados (*data mining*). Não são poucas as pesquisas acadêmicas que têm levantado preocupações, especialmente no que concerne à discriminação em desfavor de minorias.

Demais disso, o reconhecimento dos chamados vieses implícitos revela a preocupação relevante, no campo dos processos judiciais, em que ferramentas de automação decisória, ao "aprender" com os dados dispostos, seriam capazes incorporar os mesmos vieses da pessoa a quem busca auxiliar, absorvendo, seja nas regras ou nos dados, os vícios já existentes.

Em síntese, as preocupações relativas à violação de direitos fundamentais pelo uso de algoritmos decisórios, em especial, de três achados: a opacidade dos algoritmos, o emprego de *inputs* viciados e a discriminação que pode ser gerada com o seu uso.

Excluídos os princípios aplicáveis, de forma geral, às relações públicas (boa-fé, legalidade, transparência, não discriminação etc.), uma análise conjunta dos riscos do uso de algoritmos decisórios e das normas que já disciplinam o processamento de dados pessoais permite a definição de um esboço normativo próprio ao uso de algoritmos no ambiente estatal.

A proposta apresentada no presente trabalho contempla três princípios: auditabilidade, transparência e consistência ou regularidade procedimental.

O primeiro é a auditabilidade, consistente na capacidade de preservação de todo o conjunto de informações utilizados na cadeia de funcionamento de uma determinada ferramenta. Em se tratando de algoritmos de automação decisória, essas informações compreendem ao menos três categorias: a) registro de todos os dados abstratos que são levados em consideração pelo algoritmo (*inputs* e dados estatísticos abstratamente utilizados); b) informações específicas utilizadas numa decisão concreta (*inputs* concretamente utilizados); c) regra do algoritmo.

Enquanto a auditabilidade diz respeito ao registro de informações, o princípio da transparência impõe que essas informações possam ser adequadamente acessadas e explicadas. Da transparência, extraem-se duas obrigações: acesso (publicidade) e explicação.

Uma forma de conciliar a transparência com a proteção do segredo industrial consiste no emprego de *protective orders*. Em casos assim, assistentes técnicos dos autores podem ter acesso ao código-fonte de aplicações, bem como a outros dados secretos, mediante o emprego de técnicas protetivas e obrigações específicas. Outra forma consiste na disponibilização de dados apenas a instituições de checagem certificadas e a órgãos e instituições públicas, impondo-se o dever de manutenção da confidencialidade.

Pelo princípio da consistência ou regularidade procedimental, é necessário assegurar que a cada destinatário ou usuário de uma determinada ferramenta construída a partir de algoritmos seja aplicado o mesmo *procedimento*, bem como que esse procedimento não tenha sido desenvolvido de forma que lhe cause desvantagens a alguém em particular. A regularidade procedimental deve conduzir à consistência dos *outputs*, evitando que *inputs* similares produzam resultados discrepantes.

O atual estágio tecnológico da ciência da computação contempla ferramentas capazes de avaliar a regularidade procedimental, incorporando-se no desenho dos sistemas. Essas ferramentas podem assegurar que: a) a mesma disciplina foi utilizada na construção de cada decisão; b) a disciplina foi integralmente especificada antes de os sujeitos envolvidos serem conhecidos, reduzindo-se a possibilidade de desenvolvimento de um procedimento em desfavor de alguém em particular; c) cada decisão é o produto das regras e dos *inputs* aplicados; d) se a decisão requer *inputs* randômicos, a sua escolha ocorreu para além do interesse de qualquer pessoa.

7. Referências

ALPAYDIN, Ethem. *Machine Learning*. Cambridge, MA: The MIT Press, 2016.

ATHAEE, Yavar. Artificial intelligence black box and the failure of intent and causation. *Harvard Law Review,* v. 31, n. 2, p. 889-938, 2018.

BBC BITESIZE. *Representing an algorithm:* pseudocode. Disponível em: [www.bbc.co.uk/bitesize/guides/zpp49j6/revision/2#:~:text=Writing%20in%20pseudocode%20is%20similar,pseudocode%2C%20INPUT%20asks%20a%20question]. Acesso em: 24.10.2020.

BURCH, Traci. Skin color and the criminal justice system: beyond black-white disparities in sentencing. *Journal of Empirical Legal Studies,* v. 12, p. 395-420, 2015.

COMISSÃO EUROPEIA. *Ethics guidelines for trustworthy AI*. Disponível em: [https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top]. Acesso em: 25 out. 2020.

COMISSÃO EUROPEIA. *Guidelines on automated individual decision-making and Profiling*. Disponível em: [https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053]. Acesso em: 25.10.2020.

COMISSÃO EUROPEIA. *White paper on artificial intelligence:* a european approach to excellence and trust. Disponível em: [https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf]. Acesso: 25.10.2020.

CONSELHO DA EUROPA. *Algorithms and human rights*. Disponível em: [https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5]. Acesso em: 09.10.2020.

CONSELHO DA UNIÃO EUROPEIA. Recomendação CM/Rec(2012)3 do Comitê de Ministros. Disponível em: [https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805caa87]. Acesso em: 11.10.2020.

COSTA, Eduardo José da Fonseca. *Levando a imparcialidade a sério:* proposta de um modelo interseccional entre direito processual, economia e psicologia. Salvador: JusPodivm, 2018.

DITOMASSO, Natalie Ann. The Essence of Innocence: Consequences of Dehumanizing Black Children. *Journal of Personality and Social Psychology*, v. 106, n. 4, p. 526-545, 2014. Disponível em: [www.apa.org/pubs/journals/releases/psp-a0035663.pdf]. Acesso em: 30.08.2020.

DYSON, Freeman. How to dispel your illusions. *New York Review of Books*, dez. 2011. Disponível em: [www.nybooks.com/articles/archives/2011/dec/22/how-dispel-your-illusions/]. Acesso em: 23.08.2020.

EBERHARDT, Jennifer L; DAVIES, Paul G.; PURDIE-VAUGHNS, Valerie J.; JOHNSON, Sheri Lynn. Looking deathworthy: perceived stereotypicality of black defendants predicts capital-sentencing outcomes. *Psychological Science,* v. 17, p. 383-386, 2006.

EUCLIDES. *Os elementos*. Trad. Irineu Bicudo. São Paulo: UNESP, 2009.

FENOLL, Jordi Nieva. *Inteligencia artificial y proceso judicial*. Madrid, Barcelona, Buenos Aires, São Paulo: Marcial Pons, 2018.

FENOLL, Jordi Nieva. Transfondo psicológico de la independencia judicial. In: FENOLL, Jordi Nieva; OTEIZA, Eduardo (Dir.). *La independência judicial:* un constante asedio. Madri, Barcelona, Buenos Aires, São Paulo: Marcial Pons, 2019.

FERRARI, Isabela; BECKER, Daniel; WOLKART, Erik Navarro. Arbitrium ex machina: panorama, riscos e a necessidade de regulação das decisões informadas por algoritmos. *Revista dos Tribunais*, v. 995, p. 635-655, set. 2018.

GOFF, Phillip Atiba; JACKSON, Matthew Christian; DI LEONE, Brooke Allison Lewis; CULOTTA, Carmen Marie. The essence of innocence: Consequences of dehumanizing Black children. *Journal of Personality and Social Psychology*, v. 106, n. 4, p. 526–545, 2014.

HOOFNAGLE, Chris; SOLTANI, Ashkan; GOOD, Nathan; WAMBACH, Dietrich James; AYENSON, Mika D. Behavioral advertising: the offer you cannot refuse. *Harvard Law & Policy Review,* v. 6, 2012.

IBM. Machine Learning e Ciência de dados com IBM Watson. Disponível em: [www.ibm.com/br-pt/analytics/machine-learning].

JONES, Craig. The troubling new science of legal persuasion: heuristics and biases in judicial decision-making. *Advocates Quaterly,* v. 41, p. 49-122, 2013.

JONES, Craig; RANKIN, Micah B. E. Justice as a rounding error? Evidence of subconscious bias in second-degree murder sentences in Canada. *Osgoode Digital Commons,* v. 10, n. 81, 2014.

KAHNEMAN, Daniel. *Rápido e devagar.* Trad. Cássio de Arantes Leite. Rio de Janeiro: Objetiva, 2012.

Kehl, Danielle; Guo, Priscilla; Kessler, Samuel. Algorithms in the criminal justice system: assessing the use of risk assessments in sentencing. *Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School,* 2017. Disponível em: [https://dash.harvard.edu/bitstream/handle/1/33746041/2017-07_responsivecommunities_2.pdf]. Acesso em: 12.10.2020.

KING, Allan G.; MRKONICH, Marko J. "Big data" and the risk of employment discrimination. *Oklahoma Law Review,* v. 68, p. 555-584, 2016.

KROLL, Joshua A.; HUEY, Joanna; BAROCAS, Solon; FELTEN, Edward W.; REIDENBERG, Joel R.; ROBINSON, David G.; YU, Harlan. Accountable algorithms. *University of Pennsylvania Law Review*, v. 165, n. 3, p. 633-706, 2017.

LARSON, Jeff; MATTU, Surya; LAUREN, Kirchner; ANGWINHOW, Julia. *We analyzed the COMPAS recidivism algorithm*. Disponível em: [www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm]. Acesso em: 10.08.2019.

MAINKA, Spencer M. Algorithm-based recruiting technology in the workspace. *Texas A&M Journal of Property Law,* v. 5, p. 801-922, 2019.

MAYSON, Sandra G. Bias in, bias out. *Yale Law Journal,* v. 128, 2019.

NEGOWETTI, Nicole. E. Judicial Decisionmaking, Empathy and the Limits of Perception. *Akron Law Review,* v. 47, 2014.

NUNES, Dierle; LUD, Nathanael; PEDRON, Flávio. *Desconfiando da imparcialidade dos sujeitos processuais*. Salvador: JusPodivm, 2018.

NUNES, Dierle; MARQUES, Ana Luiza Pinto Coelho. Inteligência artificial e direito processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas. *Revista de Processo*, v. 285, p. 421-447, nov. 2018.

O'CONNOR, Michael P; RUMANN, Celia. Into the fire: how to avoid getting burned by the same mistakes made fighting terrorism in Northern Ireland. *Cardozo Law Review*, v. 24, 2003.

PARKS, Gregory S.; DAVIS, Andre M. Confronting implicit bias: an imperative for judges in capital prosecutions. *Human Rights Magazine,* v. 42, n. 2, 2016. Disponível em: [www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/2016-17-vol-42/vol--42--no--2---the-death-penalty--how-far-have-we-come-/confronting-implicit-bias--an-imperative-for-judges-in-capital-p/]. Acesso em: 30.08.2020.

PARLAMENTO EUROPEU E CONSELHO DA UNIÃO EUROPEIA. *Regulamento (UE) 2016/679*. Disponível em: [https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=celex%3A32016R0679]. Acesso em: 18.10.2020.

PEER, Eyal; GAMLIEL, Eyal. Heuristics and biases in judicial decisions. *Court Review,* v. 49, n. 2, p. 114-119, 2013.

PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. *Inteligência artificial e direito*. Curitiba: Alteridade, 2019.

PEREIRA, Marcos Emanoel; ÁLVARO, José Luis; OLIVEIRA, Andréia C. Oliveira; DANTAS, Gilcimar. Estereótipos e essencialização de brancos e negros: um estudo comparativo. *Psicologia & Sociedade*, v. 1, p. 144-153, 2011.

ROOTH, Dan-Olof. Implicit discrimination in hiring: real world evidence. *Institute for the Study of Labor (IZA)*, 2007. Disponível em: [http://ftp.iza.org/dp2764.pdf]. Acesso em: 30.08.2020.

SILLS, Arthur J. Automated data processing and the issue of privacy. *Seton Hall Law Review*, v. 1, 1970.

SIQUEIRA, Adliz da Rocha; FONSECA, Claudia Salvini Barbosa Martins da; PAULA, Isabela Maria Barbosa de; NOVAIS Marina Magalhães. Doença celíaca: um diagnóstico diferencial a ser lembrado. *Brazilian Journal of Allergy and Immunology*, v. 2, n. 6, 2014.

STANFORD UNIVERSITY INFOLAB. *Arthur Samuel: Pioneer in Machine Learning*. Disponível em: [http://infolab.stanford.edu/pub/voy/museum/samuel.html]. Acesso em: 04.10.2020.

STEINBOCK, Daniel J. Data matching, data mining and due process. *Georgia Law Review,* v. 40, n. 1, 2005.

SUNSTEIN, Cass R. *Behavioral law and economics*. Cambridge: Cambridge University Press, 2000.

TURING, A. M. *Lecture to the London Mathematical Society,* feb. 20, 1947.

TURING, A. M. Computing machinery and intelligence. *Mind,* v. LIX, n. 236, 1950.

TVERSY, Amos; KAHNEMAN, Daniel. Availability: a heuristic for judging frequency and probability. *Cognitive Psychology,* v. 2, n. 5, p. 207-232, 1973.

TVERSY, Amos; KAHNEMAN, Daniel. Judgment under uncertainty: heuristics and biases. *Science,* v. 185, p. 1124-1131, 1974.

U.S. Government Accountability Office. *Data mining:* federal efforts cover a wide range of uses, 2004. Disponível em: [www.gao.gov/products/GAO-04-548]. Acesso em: 12.10.2020.

VAN HAASTERT, Hugo. 2016. Government as a platform: public values in the age of big data. *Oxford Internet Institute,* 2016. Disponível em: [http://blogs.oii.ox.ac.uk/ipp-conference/sites/ipp/files/documents/Government%2520as%2520a%2520Platform%2520-%2520Big%2520data%2520paper.pdf]. Acesso em: 12.10.2020.

VILLASENOR, John; FOGGO, Virginia. Artificial intelligence, due process and criminal sentencing. *Michigan State Law Review,* v. 295, p. 295-354, 2020.

WALDFOGEL, Joel; AYERS, Ian. A market test for race discrimination in bail setting. *Stanford Law Review,* v. 46, p. 987-1048, 1994.

YANISKY-RAVID, Shlomit; HALLISEY, Sean. Equality and privacy by design: a new model of artificial intelligence data transparency via auditing, certification, and safe harbor regimes. *Fordham Urban Law Journal,* v. 46, n. 2, p. 428-486, 2019.

Youyou, Wu; Kosinski, Michal; Stillwell, David. Computer-based personality judgments are more accurate than those made by humans. *PNAS,* v. 112, n. 4, p. 1036-1040, 2015.

Zittrain, Jonathan. Engineering an election. *Harvard Law Review,* v. 127, p. 335-341, 2013-2014.

Pesquisas do Editorial

Veja também Doutrinas relacionadas ao tema

- A ditadura do algoritmo e a proteção da pessoa humana: uma análise do controle do si eletrônico, de Juliana Evangelista de Almeida e Daniel Evangelista Vasconcelos Almeida – *RDPriv* 69/29-43;
- Compreendendo algoritmos aplicados ao sistema de justiça criminal – ilegibilidade, acesso, compreensão, verdade e computabilidade no 'eu' identificado por algoritmos, de Rafael de Deus Garcia e Evandro Piza Duarte – *RBCCrim* 183/199-226; e
- Inteligência artificial, *big data* e algoritmos: policiamento e as novas roupagens de um agir discriminatório, de Rodrigo Ghiringhelli de Azevedo e Luíza Correa de Magalhães Dutra – *RBCCrim* 183/247-268.