

INTELIGÊNCIA ARTIFICIAL E MODELOS DE DECISÃO AUTOMÁTICOS: O ENVIESAMENTO DOS ALGORITMOS

ARTIFICIAL INTELLIGENCE AND AUTOMATIC DECISION MODELS: THE BIASING OF ALGORITHMS

ROBSON PASCOAL PEREIRA¹
ANDREI MENESES LORENZETTO²

Resumo

A ascensão do uso de sistemas com inteligência artificial, desde situações do cotidiano até as mais complexas, faz com que surja a necessidade de se verificar sobre quais bases, critérios e modelos estes sistemas têm sido construídos. Diante do potencial de tomada de decisão autônoma e de modelos não supervisionados por humanos, este estudo tem a intenção de entender o alcance destas decisões e seus impactos no dia a dia dos indivíduos. Utilizando o método analítico-dedutivo, procura entender como os dados, tidos como matéria prima destes sistemas de inteligência artificial, podem prover modelos preconceituosos ou discriminadores, diante da carga valorativa pelas quais se dá a coleta e análise. Averigua ainda, quais as diretrizes éticas, direitos e garantias devem ser resguardados dos cidadãos nestes tratamentos, e qual a capacidade destes de se oporem ao resultado automatizado. Por fim, o artigo objetiva analisar os novos institutos da regulamentação do uso consciente da inteligência artificial, mormente as diretrizes éticas para inteligência artificial confiável europeia, além da amplamente citada atualmente, Lei Geral de Proteção de Dados e

¹ Bacharel em Ciência da Computação pela Universidade Federal do Paraná (2005). Pós-graduado pela Pontifícia Universidade Católica em Engenharia de Software (2008). Pós-graduado em Gestão de Tecnologia da Informação pelas Faculdades Opet (2013). Bacharel em Direito pelo Centro Universitário Opet (2020). Advogado inscrito na Ordem dos Advogados do Brasil, seccional Paraná. Mestrando em Direito pelo Centro Universitário Autônomo do Brasil (UniBrasil). E-mail: pascoal.robson@gmail.com

² Graduado em Direito pela Pontifícia Universidade Católica de Minas Gerais (2003). Mestre (2021) e Doutorando do Programa de Pós-graduação Stricto Sensu em Direitos Fundamentais e Democracia pelo Centro Universitário Autônomo do Brasil - UniBrasil/PR. Membro do núcleo de pesquisa em Direito Constitucional (Nupeconst) do Programa de Pós-graduação (PPGD) do Centro Universitário Autônomo do Brasil - UniBrasil/PR. Advogado.

o Projeto de Lei 21/2020, que busca estabelecer princípios para o uso da inteligência artificial no Brasil.

Palavras-chave: Inteligência artificial, diretrizes éticas, enviesamento.

Abstract

The rise in the use of systems with artificial intelligence, from everyday situations to the most complex, makes it necessary to verify on what bases, criteria and models these systems have been built. Given the potential for autonomous decision-making and models not supervised by humans, this study intends to understand the scope of these decisions and their impacts on the daily lives of individuals. Using the analytical-deductive method, it seeks to understand how data, taken as the raw material of these artificial intelligence systems, can provide biased or discriminating models, given the value load by which the collection and analysis takes place. It also investigates which ethical guidelines, rights and guarantees must be safeguarded from citizens in these treatments, and what their capacity to oppose the automated result is. Finally, the article aims to analyze the new institutes of regulation of the conscious use of artificial intelligence, especially the ethical guidelines for reliable European artificial intelligence, in addition to the currently widely cited, General Data Protection Law and the Bill 21/2020, which seeks to establish principles for the use of artificial intelligence in Brazil.

Keywords: Artificial intelligence, ethical guidelines, bias.

1 - Introdução

Dizer que a sociedade está migrando da sociedade da informação para a sociedade do conhecimento já não é mais novidade. Muito menos que, neste cenário, os algoritmos deixaram de representar apenas uma sequência de passos pré-definidos com a finalidade de automatizar tarefas. Inteligência Artificial, *Machine Learning*, *Deep Learning*, *Big Data*, análise preditiva, nanotecnologia, não mais representam conceitos abstratos pertencentes somente ao dicionário dos profissionais de tecnologia.

Nesta conjuntura o presente trabalho procurar abordar, utilizando o método analítico-dedutivo, por meio de extensa pesquisa bibliográfica, primeiramente uma conceituação, ainda que de forma preliminar, de inteligência artificial e suas subdivisões. Ademais, através de técnicas de aprendizagem, permitem análise de massa de dados e respostas personalizadas, dotadas de subjetividade e em constante aprimoramento. O trabalho objetiva ainda, salientar a importância dos dados, que nestes modelos são vistos como matéria prima para a inteligência artificial, seja em modelos de respostas simples, seja em algoritmos complexos.

Entendendo que os modelos matemáticos e estatísticos baseados em aprendizagem de máquinas permitem uma análise preditiva, cujas respostas podem partir de modelos não supervisionados por humanos, é preciso compreender quais as premissas de tratamento de dados, e mais, de que forma estes dados são capturados e utilizados. Ainda, interpretar como a transparência e o direito a explicação do tratamento automatizado podem ser lidos neste cenário. Entra-se no âmago desta pesquisa então para questionar, uma vez que se assume (por conta no volume de dados) a impossibilidade humana na compreensão completa destes critérios, quais seriam os possíveis enviesamentos e valores que destes dados se extraem. Algoritmos podem não ser neutros e tem potencial de exprimir modelos discriminatórios, misóginos e racistas, dependendo dos dados que os alimentam.

Tentativas de controles e estabelecimento de um mecanismo confiável nos sistemas de inteligência artificial tem sido objeto de regulamentações estrangeiras e nacionais, como as Diretrizes Éticas estabelecidas pela Comissão Europeia, além da já em vigor, e amplamente citada atualmente, Lei Geral de Proteção de Dados (LGPD), que, juntamente com o Projeto de Lei 21/2020, buscam estabelecer limites de tratamento de dados e uso de sistemas de inteligência artificial.

Entender os impactos destes modelos de inteligência artificial e se estes institutos respondem as necessidades de controle é o intento desta pesquisa, que procura dar luz a este problema, que não é invisível, ainda que não possa ser enxergado por olhos desatentos aos riscos e impactos às garantias e liberdades individuais.

2 – Algoritmos e inteligência artificial

A tecnologia e a capacidade computacional alcançaram tamanha medida que, não cabe mais entender algoritmos como métodos utilizados para respostas objetivas e limitadas. Ainda que a premissa de transformação de uma entrada (*input*) em uma saída (*output*) se mantenha, análises complexas, toneladas de dados³, modelos neurais, modificaram e, ainda parece que tendem a modificar mais, a capacidade de utilização e a adaptação das respostas destes programas.

Esta adaptação dos algoritmos passa pela capacidade, cada vez mais latente, de interação sem a necessidade de interferência humana, sendo esta uma das características da inteligência artificial, ou seja, a habilidade de, através do conjunto de informações a que é submetida, responder de formas distintas e gradativamente mais complexas. Ainda que a definição de inteligência artificial encontre múltiplos caminhos, e que o alcance deste trabalho não permita este aprofundamento, procurar-se-á atribuir um conceito que sirva de parâmetro para as demais reflexões do texto. Neste sentido a inteligência artificial pode ser compreendida “a partir de características como autonomia, habilidade social, cooperação, proatividade e reatividade. Tais atributos põem-se a indicar a aptidão de algoritmos à atuação com ou sem intervenção humana”.⁴

Outro recorte fundamental para a definição da fronteira conceitual da inteligência artificial é desprender da ideia de uma inteligência artificial ficcional, muito abordada em filmes, nos quais a inteligência artificial toma vontade própria e passa a dominar, manipular e até exterminar a humanidade. Aqui, nos interessa a inteligência artificial real, capaz de incorporar funcionalidades práticas que auxiliam no dia a dia, que são utilizadas por milhares de empresas, que automatizam tarefas, que propõe soluções que assistem as pessoas no desenvolvimento de suas vidas, muito embora seja inevitável que haja uma espécie de inquietação no ser humano, que se expande cada vez mais, frente a uma suposta eliminação do limite entre homens e máquinas.⁵

³ Aqui o termo toneladas de dados, refere-se tão somente a quantidade imensurável de dados que os computadores são capazes de processar em tempos cada vez menores.

⁴ TEPEDINO, Gustavo; GUIA DA SILVA, Rodrigo. **Desafios da inteligência artificial em matéria de responsabilidade civil**. Revista Brasileira de Direito Civil – RBDCivil, Belo Horizonte, v. 21 Revista Brasileira de Direito Civil, v. 21, n. 02, 2019. p. 111

⁵ Barboza, distingue inteligência artificial real de inteligência artificial ficcional. “A IA real salva vidas, ao incorporar mecanismos de segurança, como freios automáticos, e otimizar processos de previsão de falhas. Centenas de empresas existem e milhares de pesquisadores e engenheiros estudam nesta área, porque objetivam produzir soluções e ajudar as pessoas a desenvolverem suas vidas. O outro tipo de IA, que compreende IA superinteligentes que matarão todos, “é ficção” e os cientistas tendem

Ainda que sempre relacionada a inteligência artificial, o subcampo que possibilita o aprendizado pode ser denominado como *Machine Learning* (Aprendizado de Máquina). Conforme Maini e Sabri, através deste subcampo é que o computador é capaz de identificar padrões em dados observados e construir modelos que preveem comportamentos sem tê-los sido previamente definidos.⁶

Descendo mais um nível de abstração, encontra-se o *Deep Learning*, que pode ser entendido como uma especialização do *Machine Learning*, porém com maior capacidade preditiva e personalizada, através de camadas de comunicação em redes neurais artificiais, que se assemelhariam a transmissão neural humana. Saindo um pouco do “*tecniquês*”, o *Deep Learning* não é algo para o futuro, está presente em diversas aplicações, desde a indicação que o Netflix faz com base nos filmes que assistiu, até aquela propaganda insistente que são oferecidas após uma busca do Google, que por sinal, só são possíveis por conta dos dados que são obtidos com ou sem consentimento e ciência do usuário.

Isso significa que, a partir de um conjunto de dados que podem ser supervisionados ou não, ou seja, sem terem sido previamente classificados ou filtrados por seres humanos, a máquina é capaz de estabelecer conexões semânticas, identificar padrões subjetivos e inferir respostas que variam de acordo com outra característica da *machine learning*, qual seja a retroalimentação. A partir da análise da assertividade das respostas – grau de precisão – a máquina aperfeiçoa seu modelo e passa a entregar novos resultados, em um processo cíclico de aprendizagem.⁷

Um exemplo de utilização de técnicas não supervisionadas de inteligência artificial, segundo Maini e Sabri é a segmentação de clientes do provedor de dados de marketing *Personicx*. Este serviço segmenta as famílias dos EUA em grupos distintos dentro de estágios de vida para possibilitar anúncios direcionados do Facebook em campanhas publicitárias.⁸ Outro exemplo, ainda mais conhecido, é o

a trabalhar na IA real.” BARBOZA, Heloisa Helena; ALMEIDA, Vitor. **Pessoas com (d)eficiência e inteligência artificial: primeiras reflexões**. In: TEPEDINO, Gustavo; SILVA, Rodrigo da Guia (coord.). O Direito Civil na era da inteligência artificial. São Paulo: Revista dos Tribunais, 2020. p. 83-102.

⁶ MAINI, Vishal; SABRI, Samer. **Machine Learning for Humans**. [s.l.: s.n.], 2017. Disponível em <https://everythingcomputerscience.com/books/Machine Learning for Humans.pdf>. Acesso em 10 jun. 2021. p. 16

⁷ *Ibid.*

⁸ No original: “An interesting example of clustering in the real world is marketing data provider Acxiom’s life stage clustering system, Personicx. This service segments U.S. households into 70 distinct clusters within 21 life stage groups that are used by advertisers when targeting Facebook ads, display ads, direct

Airbnb, que agrupa dados demográficos e hábitos de compra semelhantes para definir anúncios relevantes.

Neste ponto surge o primeiro questionamento sobre a utilização da inteligência artificial, (que daqui em diante utilizar-se-á no seu conceito mais amplo, abrangendo todos os seus subcampos), quais sejam, as fontes de dados. Fica nítido que só é possível imaginar uma aprendizagem complexa tendo como entrada um grande volume de dados.

Ocorre que a produção de dados se verifica a todo instante. Ao dormir, o sono é monitorado por relógios inteligentes. Ao despertar conecta-se ao celular, utiliza-se serviços de transporte compartilhado para se deslocar ao trabalho e o GPS está ligado a todo momento. Por mais que se queira desconectar, é praticamente nulo o serviço que não gere, colete ou armazene dados. Neste cenário, os dados representam a matéria prima para os algoritmos. “Esta nova matéria-prima com a qual tudo é “possível” conhecer, e as grandes empresas, deste novo tempo, são as detentoras deste novo “ouro” e nós somos os seus fornecedores”.⁹

Embora a legitimidade e o correto tratamento destes dados sejam de fundamental importância na discussão jurídica, importa neste momento, entender, mais do que a forma como foram obtidos ou a quem pertencem, quais características são possíveis de serem extraídas deles e além disso, sob qual critério de valores serão utilizados.

Tal questionamento toma relevância na medida que, partindo que já foi exposto, em relação a análises não supervisionadas de inteligências artificiais, o risco de generalizações pode formatar um modelo discriminatório, enviesado e, sob a perspectiva de tomada de decisões automáticas, este modelo afeta diretamente os indivíduos que se veem, por força da redução de custos operacionais, da automação, ou de qualquer outra razão que permita a análise subjetiva, reféns destes resultados.

3 – Modelos de tomada de decisão e análise preditiva

mail campaigns, etc”. MAINI, Vishal; SABRI, Samer. **Machine Learning for Humans**. [s.l.: s.n.], 2017. Disponível em <https://everythingcomputerscience.com/books/Machine Learning for Humans.pdf>. Acesso em 10 jun. 2021. p. 56-58

⁹ NOVAIS, Paulo; FREITAS, Pedro Miguel. **Inteligência Artificial e Regulação de Algoritmos**. Diálogos União Europeia-Brasil. In: BRASIL: Ministério da Ciência, Tecnologias, Inovação e Comunicação, 2018. Disponível em: <https://bitly.com/rsuJh>. Acesso em 30 maio. 2021. p. 10

Só faz sentido pensar em automatização com utilização de inteligência artificial se, por questões econômicas, o tempo despendido ou o custo da operação puder ser reduzido. Por óbvio a capacidade de analisar um volume grande de dados é incomparável entre homem e máquina. Em vista disso, empresas constroem modelos matemáticos que passo a passo vêm apoiando nas decisões de negócios, tendo por vezes, deliberações independentes de avaliação humana, como no caso da já citada *machine learning* não supervisionada.

São vastos os cenários que atualmente já contam com modelos de decisão automatizadas, tais como: avaliação de concessões de crédito, processos seletivos de recrutamento, contratos de seguro, etc... Tais modelos são construídos com base em dados informados pelo interessado, mas que, na maioria dos casos, obtém dados em fontes públicas e privadas, redes sociais, ou quaisquer outras fontes de dados que puder interessar as empresas.

À primeira vista, ainda que tal decisão recaia sobre critérios objetivos, com o aprendizado e a retroalimentação da inteligência artificial, identificar qual informação ou requisito foi fundamental para uma resposta positiva ou negativa, torna-se uma tarefa hercúlea. Há de se convir que para quem está participando de um processo seletivo de contratação, saber quais dados seus são utilizados, ou quais critérios valorados são de fundamental importância.

Além disso, utilizando o mesmo exemplo do candidato ao processo seletivo de contratação, as respostas do modelo de tomada de decisão podem ter sido baseadas em dados incompletos, incorretos, imprecisos e, nestes casos, prejudicando uma resposta assertiva, ainda que os critérios e modelos tenham sido suficientemente corretos. Ademais, ainda que a inteligência artificial seja capaz de agrupar determinados padrões de dados para inferir resultados, invariavelmente seu comportamento está associado a umas das duas questões: a resposta que se espera por parte de quem programou o algoritmo, ou a massa de dados fornecida. Neste ponto situam-se as principais críticas a este modelo de negócios: é possível que haja convicções pessoais incorporadas na inteligência artificial ou, sobre quais valores éticos e morais a massa de dados foi aferida.

Outro ponto relevante neste estudo diz respeito a previsões com bases nestes dados coletados. Aqui, não se trata apenas de respostas valorativas, mas de modelos sugestivos, que condicionam o comportamento, apresentam alternativas, ou seja, que

indicam o que e quando deve ser consumido. Entender o passado e o presente para prever o futuro, medir comportamentos para influenciar decisões, pertencem a chamada análise preditiva, cuja intenção é, determinar padrões de dados para antever resultados e tendências. Para tal, é necessário que a inteligência artificial realize tomadas de decisões baseada sobre incertezas, que são críveis somente com a maturidade dos dados e um autoaperfeiçoamento recursivo. Mesclando matemática e tecnologia, a análise dos dados permite construir máquinas que aprendem com suas próprias experiências.¹⁰

Essa imprevisibilidade sobre os resultados faz parte da análise preditiva e Ana Frazão afirma que aos poucos os algoritmos vão adquirindo a capacidade de decodificar as “pegadas digitais das pessoas”. Assim podem prever até mesmo o que não se tem consciência. “Por meio do aprendizado de máquina, os algoritmos ainda podem “aprender” e modificar sua própria estrutura e conduta, sem que haja propriamente controle ou mesmo previsibilidade sobre tais alterações e os resultados que daí decorrerão”.¹¹

Entretanto, a imprevisibilidade não pode ser lida na contramão da transparência. O fato de os algoritmos possuírem alto grau de abstração os fazem serem entendidos, se é que são entendidos, como uma caixa-preta, cuja chave não está disponível para os usuários. Não se sabe ao certo como funcionam e normalmente somente os desenvolvedores são capazes de interpretá-los. Ademais, do ponto de vista jurídico, delegar essas decisões as máquinas não deveria ser uma alternativa isenta de responsabilidades. “A questão a ser enfrentada aqui é saber em que medida podemos terceirizar nossas responsabilidades para máquinas e quais as repercussões jurídicas disso”.¹²

Perigo maior do que a indicação de um filme incorreto ou de uma propaganda inconveniente está nas decisões automatizadas não supervisionadas em campos mais delicados como a medicina ou trânsito. Quando tratamentos incorretos forem

¹⁰ NIVEL, E. et al. **Bounded Recursive Self-Improvement**. Reykjavik University – School of Computer Science. p. 1-56. 2013.

¹¹ FRAZÃO, Ana. **Algoritmos e inteligência artificial: repercussões da sua utilização sobre a responsabilidade civil e punitiva das empresas**. Revista Jota, 15 maio 2018. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/algoritmos-e-inteligencia-artificial-15052018>. Acesso em: 30 maio 2021. p. 2-6

¹² *Ibid.*

prescritos, acidentes de trânsito ocorrerem, essas caixas-pretas dos algoritmos servirão de obstáculos? Quem, e em que medida, será responsável para reparação do dano? Não cabe aqui, no escopo da pesquisa abordar a responsabilização civil, por erros de decisão das máquinas. Certamente este tema vale um maior aprofundamento, dada a importância prática deste tema. De todo modo, cumpre papel a cobrança por um direito a explicação, a um modelo transparente, nos quais os usuários possam ao menos ter um mecanismo de rastreabilidade que permitia ter a ciência de que estão interagindo com uma inteligência artificial e ainda serem informados de suas capacidades e limitações.

A esse direito de explicação, Gabriela Buarque e Marcos Ehrhardt apontam que essa exigibilidade é diretamente proporcional ao impacto da tomada de decisão da vida do indivíduo. A explicação deve ser oportuna e verificável pelo envolvido.¹³

Pelo menos dois problemas surgem na disponibilização dos meios que os algoritmos processam os dados para retornar um resultado. O primeiro, de natureza legal, se ampara na proteção ao segredo comercial, que se funda na ideia de uma vantagem competitiva como fator determinante para o negócio. O segundo, de natureza lógica, implica que, se as pessoas soubessem quais os critérios são usados e de que forma são, poderiam manipular as entradas, com vista a obtenção de um resultado direcionado. Costa Trigo traz outro exemplo que ilustra o potencial direcionamento e problema da exposição do código neste sentido. Declarações anuais de imposto de renda são analisadas por algoritmos. “Em tais situações, a revelação dos parâmetros utilizados acabaria por possibilitar que o contribuinte fraudasse o sistema, orientando suas declarações de acordo com o que fosse – ou não – dele esperado”.¹⁴

Mesmo na eventualidade de ter sido exposto o código fonte ao usuário, o problema da transparência encontra uma nova e talvez a mais complexa barreira. Como citado no capítulo anterior, o volume de dados processado pela inteligência artificial, para a decisão proposta, pode ser da ordem de grandeza inviável para

¹³ SILVA, Gabriela Buarque Pereira; JÚNIOR, Marcos Ehrardt. **DIRETRIZES ÉTICAS PARA A INTELIGÊNCIA ARTIFICIAL CONFIÁVEL NA UNIÃO EUROPEIA E A REGULAÇÃO JURÍDICA NO BRASIL**. Revista IBERC v. 3, n. 3, set./dez. 2020. p. 1-28

¹⁴ TRIGO, Alberto Lucas Albuquerque da Costa. **Breves notas sobre o controle das decisões informadas por algorítmicos**. In: TEPEDINO, Gustavo; SILVA, Rodrigo da Guia (coord.). **O Direito Civil na era da inteligência artificial**. São Paulo: Revista dos Tribunais, 2020. p. 65-80.

análise do indivíduo que manifestou seu direito a explicação, sem o auxílio de um perito ou especialista. Como exemplo, o sistema de Inteligência Artificial chinês Wu Dao 2.0 que conta com 1,75 trilhão de parâmetros.¹⁵

É impossível para um ser humano analisar mais de 1 trilhão de parâmetros, ainda que este número seja relativo ao maior sistema de inteligência artificial até então criado. “Algoritmos apenas podem ser considerados compreensíveis quando o ser humano é capaz de articular a lógica de uma decisão específica, explicando, por exemplo, a influência de determinados inputs ou propriedades para a decisão”.¹⁶

Ainda que fosse diminuído para a casa do bilhão, do milhão, dos milhares ou até mesmo das centenas de dados, precisaríamos de um software para analisar os resultados produzidos por outro software. Não é preciso se estender muito para compreender o ciclo que se forma. “O código-fonte é, portanto, apenas uma parte desse quebra-cabeça, e sua divulgação não é suficiente para demonstrar a assertividade do processo decisório”.¹⁷

Neste sentido, Frazão aponta que o segredo comercial e o caráter técnico não podem servir de barreiras para o direito individual do cidadão em ter informações claras e adequadas dos meios de tratamento de seus dados. Não se faz necessário demonstrar as minúcias das técnicas empregadas para a decisão automatizada, as fórmulas protegidas por segredos comerciais, ou as vantagens competitivas oriundas deste desenvolvimento automatizado. De todo modo, aponta que a explicação deve ser apresentada em linguagem natural. Assim como se espera que um juiz fundamente sua sentença, a decisão automatizada não se deve limitar a mera publicidade dos parâmetros em linguagem técnica.¹⁸

¹⁵ Pesquisadores da Academia de Inteligência Artificial de Pequim (BAAI) na China recentemente introduziram o Wu Dao 2.0, um sistema de IA multimodal com 1,75 trilhão de parâmetros. Isso ocorreu pouco mais de um ano após a introdução do GPT-3 e é uma ordem de magnitude maior. Como o GPT-3, o multimodal Wu Dao - que significa “iluminação” - pode executar tarefas de processamento de linguagem natural, geração de texto, reconhecimento de imagem e geração de imagem. Mas pode fazer isso mais rápido, sem dúvida melhor, e pode até cantar. (Tradução livre) COPELAND, Michael. **What’s the difference between artificial intelligence, machine learning and deep learning?** Disponível em: <<https://bit.ly/3qCkxvd>>. Acesso em: 02 jul. 2021.

¹⁶ BECKER, Daniel; FERRARI, Isabela; WOLKART, Erik Navarro. **Arbitrium Ex Machina : Panorama, Riscos E a Necessidade De Regulação das Decisões Informadas por Algoritmos.** Revista dos Tribunais, v. 995, n. Set, p. 1–16, 2018.

¹⁷ *Ibid.*

¹⁸ FRAZÃO, Ana; GOETTENAUER, Carlos. **O jogo da imitação jurídica: o direito à revisão de decisões algorítmicas como um mecanismo para a necessária conciliação entre linguagem**

Se já não fosse suficientemente grande o problema do volume de dados na análise de critérios para tomada de decisão, o que dizer dos dados enviesados, que possuem uma carga valorativa preconceituosa, discriminatória e por conseguinte geram resultados discriminadores. Não são poucos os relatos na internet de algoritmos que condicionam grau de periculosidade maior a pessoas negras, que negam crédito aos mais carentes, softwares de simulações reais que apresentam comportamentos abusivos, misóginos e homofóbicos.

Richard Lee, um estudante de Taiwan que mora na Austrália teve seu passaporte negado pois o algoritmo entendeu que o estudante, de origem asiática, não abriu os olhos para a foto.¹⁹ Em 2015, o Google Imagens identificou um usuário negro como gorila.²⁰ Esse viés fez a empresa pedir desculpa e “corrigir” sua inteligência artificial. A correção, no entanto, foi remover o indexador léxico gorila ou chipanzé de seu mecanismo de retorno das análises de imagens, ou seja, a forma de resolver o problema, foi apagando-o. A Amazon desativou seu sistema de recrutamento por inteligência artificial quando percebeu que o algoritmo estava sendo sexista ao preterir homens no processo seletivo. Isso porque os dados utilizados foram compostos pela base de currículo dos últimos 10 anos que, em sua maioria, eram de homens.²¹

Estes são exemplos conhecidos, pois tornaram-se notícias em sites de tecnologia, direito e curiosidades. Expandindo este modelo para outras situações, ainda que hipotéticas, seria possível imaginar facilmente a seguinte situação: Um software de recrutamento de cargos de direção de grandes empresas certamente buscaria casos de sucesso, ou características de grandes CEO’s para identificar o

natural e infraestrutura Matemática. In: TEPEDINO, Gustavo; SILVA, Rodrigo da Guia (coord.). O Direito Civil na era da inteligência artificial. São Paulo: Revista dos Tribunais, 2020. p. 45-63

¹⁹ Rapaz tirou uma foto para o novo passaporte, mas o clique foi recusado por um estranho motivo. Disponível em: <https://vejasp.abril.com.br/blog/pop/este-rapaz-tirou-uma-foto-para-o-novo-passaporte-mas-o-clique-foi-recusado-por-um-estranho-motivo/>. Acesso em 02 jul. 2021

²⁰ Google conserta seu algoritmo “racista” apagando os gorilas. Disponível em: https://brasil.elpais.com/brasil/2018/01/14/tecnologia/1515955554_803955.html. Acesso em 02 jul. 2021

²¹ “By 2015, it was clear that the system was not rating candidates in a gender-neutral way because it was built on data accumulated from CVs submitted to the firm mostly from males, Reuters claimed. The system started to penalise CVs which included the word women. The program was edited to make it neutral to the term but it became clear that the system could not be relied upon, Reuters was told.” Disponível em: <https://www.bbc.com/news/technology-45809919>. Acesso em 02 jul. 2021

padrão de contratação. Ocorre que por um preconceito estrutural e ainda vigente, estes cargos são ocupados em sua maioria por homens brancos. Assim, o software entenderá que este é o perfil adequado para a seleção e descartará mulheres, negros ou pardos, fazendo com que novos homens brancos ocupem estas posições, agravando o problema. Nas palavras de Isabela Ferrari, Erik Navarro Wolkart e Daniel Becker “Algoritmos aprendem pelo exemplo. Se os dados aos quais expostos refletirem o preconceito (consciente ou inconscientemente) presente na sociedade, as decisões daí derivadas irão refleti-lo e reforçá-lo.”²²

Resta claro que a qualidade da decisão tem ligação direta com a qualidade dos dados “*inputados*”. Essa possível perpetuação do enviesamento da inteligência artificial só pode ser rompida com a quebra do modelo de análise meramente estatístico.²³ A sociedade muda, ainda que por vezes tenha-se dúvida se para melhor, e estas mudanças resultam em decisões distintas para situações semelhantes. Evoluir significa a passagem gradual por um processo de transformação.

Algoritmos precisam ser treinados. E se eles aprendem pelo exemplo, estes treinamentos devem ser inclusive para que o algoritmo desaprenda algo que não é mais válido. As variáveis que compõe o resultado da análise, a interpretação dada a uma informação, a resposta apresentada a uma situação nova deve ser avaliada para minimizar os desvios e aumentar a taxa de assertividade do algoritmo.

4 – Diretrizes éticas para inteligência artificial

São inúmeras as vantagens que a inteligência artificial traz a sociedade, em todos os campos de desenvolvimento. Afinal, nos acostumamos com incontáveis serviços que levam além de conforto, comodidade, saúde e demais facilidades a coletividade. Todavia, apontou-se até então que as decisões decorrentes do uso da

²² BECKER, Daniel; FERRARI, Isabela; WOLKART, Erik Navarro. **Arbitrium Ex Machina : Panorama , Riscos E a Necessidade De Regulação das Decisões Informadas por Algoritmos**. Revista dos Tribunais, v. 995, n. Set, p. 1–16, 2018.

²³ De acordo com BRITZ: “Nesse contexto, é possível a ocorrência da discriminação por erro estatístico, o que decorreria tanto de dados incorretamente capturados como também de modelo estatístico de bases científicas frágeis”. SOUZA, Carlos Affonso; PADRÃO, Vinicius. **Responsabilidade civil e os "novos turcos": o papel da conduta humana nos danos relacionados à inteligência artificial**. In: TEPEDINO, Gustavo; SILVA, Rodrigo da Guia (coord.). O Direito Civil na era da inteligência artificial. São Paulo: Revista dos Tribunais, 2020.

inteligência artificial podem gerar riscos a parâmetros regulatórios já consolidados no ordenamento jurídico.

Neste cenário, normas de proteção de tratamento de dados têm sido apontadas como uma solução a estas ameaças às garantias constitucionais, apesar de forma muito incipiente. A própria Lei 13.709/2018, Lei Geral de Proteção de Dados (LGPD) não aborda diretamente os riscos quanto ao uso da inteligência artificial. Seu escopo está direcionado ao correto tratamento dos dados de pessoas naturais. De todo modo, sancionada em 14 de agosto de 2018, é o principal instrumento de privacidade de dados dos indivíduos, quanto ao uso não consentido. Tem como objetivo o estabelecimento de regras claras que visam garantir a proteção de dados pessoais, por meio de práticas transparentes aos titulares dos dados, qual sejam pessoas naturais.²⁴

Ao mesmo ponto que se vislumbra a problema na coleta de dados, Frazão aponta que a limitação do tratamento pode tornar os algoritmos menos eficientes e agravar o problema das decisões automatizadas.²⁵ Nesta perspectiva aponta-se que os sistemas algoritmos devem ser lidos em um processo conciliador e não por um modelo de restrição ou eliminação do tratamento.

Desta forma, não se trata de eliminar a evolução tecnológica ou impedir seu avanço, mas sim, como qualquer serviço que oferece riscos a princípios e direitos, regulamentá-la. Especialmente nestes casos que tendem a gradualmente fazer parte do cotidiano das pessoas naturais. Neste sentido, não obstante Barboza qualificar a inteligência artificial como definitiva, traz a preocupação da sua regulação, quando

²⁴ BRASIL, Lei nº 13.709 de 14 de ago. 2018. Disponível em http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em 30 mai. 2021. Em relação ao direito a explicação, abordado como direito do titular do dado, a LGPD, traz em seu 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.

²⁵ “De fato, a limitação de propósito de tratamento e a minimização de coletas, ideias consolidadas na legislação brasileira de proteção de dados, podem servir para reduzir a influência dos algoritmos na vida dos cidadãos. [...] A redução da entrada de dados diminui a capacidade de monitoramento do sistema e, portanto, pode torná-lo menos eficiente e influente. Se de alguma forma esses dois mecanismos diminuem a capacidade dos sistemas algoritmos controlarem a conduta dos cidadãos, essa redução não se dá por conciliação, mas por sufocamento.” FRAZÃO, Ana; GOETTENAUER, Carlos. **O jogo da imitação jurídica: o direito à revisão de decisões algorítmicas como um mecanismo para a necessária conciliação entre linguagem natural e infraestrutura Matemática.** In: TEPEDINO, Gustavo; SILVA, Rodrigo da Guia (coord.). *O Direito Civil na era da inteligência artificial.* São Paulo: Revista dos Tribunais, 2020. p. 45-63.

aponta que, apesar de não restarem dúvidas quanto aos benefícios proporcionados pela inteligência artificial a toda humanidade, “não se deve afastar os questionamentos ... de pertinência pela IA, como os que dizem respeito ao cabimento e aos limites de regulamentação da matéria, sob diversas vertentes, inclusive a ética.”²⁶

Normatizar, neste sentido, significa estabelecer diretrizes morais e éticas que tragam transparência a seu uso e compatibilize a inteligência artificial com os demais instrumentos do ordenamento jurídico. Em abril de 2019, A Comissão Europeia, criou um grupo de especialistas em inteligência artificial que apresentou as diretrizes éticas para uma inteligência artificial confiável. Este grupo objetiva o desenvolvimento de um mercado digital inteligente, porém sustentável e inclusivo.²⁷

Dentre os pilares do modelo proposto para as inteligências artificiais, o grupo destacou o respeito a legalidade, a ética e a robustez, que se desdobra tanto em perspectiva técnica, tanto em consideração ao ambiente social. Essas diretrizes foram compiladas em um guia que apresenta um conjunto exemplificativo de sete requisitos (supervisão humana; robustez técnica e segurança; privacidade e governança; transparência; diversidade, não discriminação e justiça; bem-estar social e responsabilidade) que todos os sistemas de inteligência artificial devem atender para ser considerado confiável.²⁸

Destaca-se o primeiro requisito, supervisão humana, o qual demonstra que a preocupação dos sistemas de inteligência artificial está na centralidade do ser humano, e no compromisso com o bem-estar e liberdade dos cidadãos. Essa necessidade de supervisão é um reconhecimento que, embora tragam grandes facilidades e oportunidades, a inteligência artificial também apresenta riscos que devem ser minimizados de forma adequada, eliminando possíveis distorções, como já apontado no capítulo anterior. “É imprescindível que haja uma avaliação dos impactos

²⁶ BARBOZA, Heloisa Helena; ALMEIDA, Vitor. **Pessoas com (d)eficiência e inteligência artificial: primeiras reflexões**. In: TEPEDINO, Gustavo; SILVA, Rodrigo da Guia (coord.). O Direito Civil na era da inteligência artificial. São Paulo: Revista dos Tribunais, 2020. p. 83-102

²⁷ Disponível em https://ec.europa.eu/info/index_en. Acesso em: 28 jul. 2021

²⁸ A confiança no desenvolvimento, na implantação e na utilização dos sistemas de IA diz respeito não só às propriedades inerentes à tecnologia, mas também às qualidades dos sistemas sociotécnicos que envolvem aplicações de IA. À semelhança das questões de (perda de) confiança na segurança da aviação, da energia nuclear ou dos alimentos, não são apenas as componentes do sistema de IA, mas o próprio sistema no seu contexto global, que podem, ou não, gerar confiança. Disponível em <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>. Acesso em: 09 jul. 2021

e uma tentativa de mitigação dos riscos necessários em uma sociedade democrática, com vistas a não infringir as esferas jurídicas dos indivíduos”. Assim como mecanismos de feedback externos aos próprios sistemas de inteligência artificial, a ponto de permitir um diálogo com a sociedade.²⁹

Estes mecanismos de supervisão e feedback, pelo menos em tese, se apresentam como um meio de defesa a pressupostos legais e podem ser lidos no sentido que, sistemas que utilizam inteligência artificial podem harmonizar com a visão conciliadora entre inovação e respeito a proteção de direitos fundamentais.

Além destes requisitos a Comissão Europeia ainda enumerou quatro princípios (respeito da autonomia humana; prevenção de danos; equidade e explicabilidade) que, conforme o próprio guia aponta, devem ser lidos em conjuntos com as demais normas internas e externas de respeito aos direitos fundamentais, não impondo qualquer hierarquia entre estas.³⁰

Embora alguns princípios se confundam com os requisitos, cumprem destaque para o presente artigo os princípios da equidade e da explicabilidade. O primeiro é entendido, em sua dimensão substantiva, como a eliminação de enviesamentos, discriminação e estigmatização dos sistemas de IA. Em sua dimensão processual significa a possibilidade de o indivíduo contestar as decisões tomadas pelos sistemas de IA. Essa contestação só é possível através do princípio da explicabilidade. Este diz respeito justamente a transparência da decisão cuja finalidade de tratamento deve ser clara aos por ela afetados.

Muito embora os problemas práticos levantados, a Comissão Europeia indica, para os casos que não seja possível a explicação dos modelos que geraram o resultado (algoritmos caixa-preta), outras medidas devem ser tomadas como rastreabilidade ou auditabilidade. Este último trecho demonstra que ainda não se tem uma compreensão completa de como esses direitos serão exercidos, razão pela qual se demonstra que o tema ainda apresenta um problema latente.

No Brasil, o Projeto de Lei 21/2020, apresentado à mesa diretora da Câmara do Deputados, pelo Deputado Eduardo Bismarck (PDT/CE), em fevereiro de 2020, "Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil,

²⁹ SILVA, Gabriela Buarque Pereira; JÚNIOR, Marcos Ehrardt. **DIRETRIZES ÉTICAS PARA A INTELIGÊNCIA ARTIFICIAL CONFIÁVEL NA UNIÃO EUROPEIA E A REGULAÇÃO JURÍDICA NO BRASIL**. Revista IBERC v. 3, n. 3, p. 1-28, set./dez. 2020. p. 15-16

³⁰ Disponível em https://ec.europa.eu/info/index_en. Acesso em: 28 jul. 2021

e dá outras providências"³¹. O projeto foi submetido à Comissão de Ciência e Tecnologia, Comunicação e Informática (CCTCI), que destacou a importância do tema, depreendendo que o objetivo do projeto não é frear o avanço da inteligência artificial, mas promover a confiança pública no desenvolvimento de novos mercados que utilizam inteligência artificial.

Destacam-se no projeto os seus fundamentos (art. 4º) e princípios (art. 6º), os quais trazem equivalências com elencados na Diretriz Europeia, principalmente os fundamentos da não discriminação e privacidade³², assim como os princípios da finalidade e explicabilidade.

Se estas normas vão conseguir produzir a proteção esperada, é algo que se verá em breve, pois as mudanças na economia são rápidas e não aguardam as regulamentações para produzir efeitos na vida dos cidadãos. Basta ver os inúmeros exemplos citados na mídia em que a inteligência artificial, e principalmente os sistemas de decisões automatizadas não supervisionados, tem gerado resultados distorcidos ou obscuros.

5 - Considerações Finais

Não se trata de criminalizar ou diminuir a importância da inteligência artificial. Ela não é ficcional, nem muito menos algo do futuro. Na medida que a inteligência artificial ganha protagonismo em uma sociedade de tecnologias digitais, entender suas possibilidades e principalmente suas limitações é fundamental para o uso correto deste instrumento. Mapear as formas de aplicação de algoritmos e principalmente quais decisões podem ser delegadas a máquinas é papel do direito, tendo em vista a capacidade de interconexão com garantias fundamentais consolidadas do Estado.

A natureza matemática das decisões algorítmicas não diminui a vulnerabilidade destas decisões, pois, como já apontado, não se trata de um erro de programação que possa ser corrigido, mas sim, de um desvio dos modelos estatísticos ou pior (e

³¹ Disponível em https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=1853928 Acesso em 09 jul. 2021.

³² FRAZÃO, Ana. **Algoritmos e inteligência artificial: repercussões da sua utilização sobre a responsabilidade civil e punitiva das empresas**. Revista Jota, 15 maio 2018. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/algoritmos-e-inteligencia-artificial-15052018>. Acesso em: 30 maio 2021.

mais provável) dos dados que compõe estes modelos, enviesando seus resultados, trazendo consigo cargas valorativas preconceituosas.

Tanto a Diretriz Europeia como o Projeto de Lei 21/2020, aliado à LGPD, buscam em seu texto estabelecer princípios, direitos e deveres que permitam o uso confiável dos sistemas de inteligência artificial. Conquistar a confiança do cidadão preocupado com a exposição de dados e o tratamento inadequado requer a percepção de um sistema legal que efetivamente controle o uso indevido e possa agir em casos de discriminação, enviesamentos e obscuridades.

É necessário enfrentar adequadamente o problema, entendendo que a regulamentação não significa nenhuma barreira para a evolução tecnológica, e tão somente, a delimitação de diretrizes éticas e principiológicas na utilização da inteligência artificial, não podendo se esconder diante da incapacidade ou da diferença de velocidade que caminham tecnologia e direito.

6- Referências bibliográficas

BRASIL. Lei nº 13.709, de 14 de ago. 2018. Lei Geral de Proteção de Dados Pessoais (LGPD), Brasília, DF. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 30 maio. 2021.

BARBOZA, Heloisa Helena; ALMEIDA, Vitor. **Pessoas com (d)eficiência e inteligência artificial: primeiras reflexões**. In: TEPEDINO, Gustavo; SILVA, Rodrigo da Guia (coord.). O Direito Civil na era da inteligência artificial. São Paulo: Revista dos Tribunais, 2020. p. 83-102.

BECKER, DANIEL; FERRARI, ISABELA; WOLKART, Erik Navarro. **Arbitrium Ex Machina: Panorama, Riscos E a Necessidade De Regulação das Decisões Informadas por Algoritmos**. Revista dos Tribunais, v. 995, n. Set, p. 1–16, 2018.

COPELAND, Michael. **What's the difference between artificial intelligence, machine learning and deep learning?** Nvidia Blog, 29 jul. 2016. Disponível em: <https://bit.ly/3qCkxvd>. Acesso em: 02 jul. 2021

FRAZÃO, Ana. **Algoritmos e inteligência artificial: repercussões da sua utilização**

sobre a responsabilidade civil e punitiva das empresas. Revista Jota, 15 maio 2018. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/algoritmos-e-inteligencia-artificial-15052018>. Acesso em: 30 maio 2021.

FRAZÃO, Ana; GOETTENAUER, Carlos. **O jogo da imitação jurídica: o direito à revisão de decisões algorítmicas como um mecanismo para a necessária conciliação entre linguagem natural e infraestrutura Matemática.** In: TEPEDINO, Gustavo; SILVA, Rodrigo da Guia (coord.). O Direito Civil na era da inteligência artificial. São Paulo: Revista dos Tribunais, 2020. p. 45-63

MAINI, Vishal; SABRI, Samer. **Machine Learning for Humans.** [s.l.: s.n.], 2017. Disponível em <https://everythingcomputerscience.com/books/Machine Learning for Humans.pdf>. Acesso em 10 jun. 2021;

NIVEL, E. et al. **Bounded Recursive Self-Improvement.** Reykjavik University – School of Computer Science. p. 1-56. 2013.

NOVAIS, Paulo; FREITAS, Pedro Miguel. **Inteligência Artificial e Regulação de Algoritmos.** Diálogos União Europeia-Brasil. In: BRASIL: Ministério da Ciência, Tecnologias, Inovação e Comunicação, 2018. Disponível em: <https://bityli.com/rsuJh>. Acesso em 30 maio. 2021

SILVA, Gabriela Buarque Pereira; JÚNIOR, Marcos Ehrardt. **DIRETRIZES ÉTICAS PARA A INTELIGÊNCIA ARTIFICIAL CONFIÁVEL NA UNIÃO EUROPEIA E A REGULAÇÃO JURÍDICA NO BRASIL.** Revista IBERC v. 3, n. 3, p. 1-28, set./dez. 2020.

SOUZA, Carlos Affonso; PADRÃO, Vinicius. **Responsabilidade civil e os "novos turcos": o papel da conduta humana nos danos relacionados à inteligência artificial.** In: TEPEDINO, Gustavo; SILVA, Rodrigo da Guia (coord.). O Direito Civil na era da inteligência artificial. São Paulo: Revista dos Tribunais, 2020.

TEPEDINO, Gustavo; GUIA DA SILVA, Rodrigo. **Desafios da inteligência artificial em matéria de responsabilidade civil.** Revista Brasileira de Direito Civil – RBDCivil, Belo Horizonte, v. 21 Revista Brasileira de Direito Civil, v. 21, n. 02, 2019.

TRIGO, Alberto Lucas Albuquerque da Costa. **Breves notas sobre o controle das decisões informadas por algorítmicos.** In: TEPEDINO, Gustavo; SILVA, Rodrigo da Guia (coord.). O Direito Civil na era da inteligência artificial. São Paulo: Revista dos Tribunais, 2020. p. 65-80.