**UNIVERSIDADE FEDERAL DA BAHIA**
**FACULDADE DE DIREITO**
**PROGRAMA DE PÓS-GRADUAÇÃO EM DIREITO**

**MARIA CLARA DE SOUZA SEIXAS**

# **"TRUSTWORTHY AI":** CONTESTANDO DECISÕES TOMADAS POR INTELIGÊNCIA ARTIFICIAL À LUZ DA PROTEÇÃO DE DADOS PESSOAIS

Salvador /BA
2024

**MARIA CLARA DE SOUZA SEIXAS**

# **“TRUSTWORTHY AI”:** CONTESTANDO DECISÕES TOMADAS POR INTELIGÊNCIA ARTIFICIAL À LUZ DA PROTEÇÃO DE DADOS PESSOAIS

Dissertação apresentada como requisito parcial à obtenção do título de Mestre em Direito pela Universidade Federal da Bahia.
Linha de Pesquisa: Autonomia e Direito Civil Contemporâneo.

Orientador: Prof. Dr. Maurício Requião

Salvador /BA
2024

Dados Internacionais de Catalogação na Publicação

S462 Seixas, Maria Clara de Souza
"Trustworthy ai" : contestando decisões tomadas por inteligência artificial à luz da proteção de dados pessoais / por Maria Clara de Souza Seixas. – 2024.
161 f.

Orientador: Prof. Dr. Maurício Requião de Sant'ana.
Dissertação (Mestrado) – Universidade Federal da Bahia, Faculdade de Direito, Salvador, 2024.

1. Inteligência artificial. 2. Direito à proteção dos dados informáticos. 3. Proteção de dados. 4. Direito digital. 5. Inteligência artificial - Confiabilidade. I. Sant'ana, Maurício Requião de. II. Universidade Federal da Bahia - Faculdade de Direito. III. Título.

CDD – 342.0858

Biblioteca Teixeira de Freitas, Faculdade de Direito, Universidade Federal da Bahia

**MARIA CLARA DE SOUZA SEIXAS**

# **“TRUSTWORTHY AI”:** CONTESTANDO DECISÕES TOMADAS POR INTELIGÊNCIA ARTIFICIAL À LUZ DA PROTEÇÃO DE DADOS PESSOAIS

Dissertação apresentada como requisito à obtenção do grau de Mestre em Direito pela Universidade Federal da Bahia, na forma do Regimento Interno do Programa de Pós-Graduação *stricto sensu* em Direito da Universidade Federal da Bahia, e aprovada pela Banca Examinadora composta pelos professores abaixo firmados, em sessão pública de defesa no dia 27 de setembro de 2024.

Aprovada em 27 de setembro de 2024, com nota: 10,00

**Banca Examinadora:**

Maurício Requião de Sant’ana – Orientador _____
*Doutor em Direito pela Universidade Federal da Bahia*
Universidade Federal da Bahia.

Saulo José Casali Bahia _____
*Doutor em Direito pela Pontifícia Universidade Católica de São Paulo*
Universidade Federal da Bahia.

Marcos Ehrhardt Jr _____
*Doutor em Direito pela Universidade Federal de Pernambuco*

## AGRADECIMENTOS

Gostaria de registrar meu profundo agradecimento a todas as pessoas que foram fundamentais para a realização deste mestrado, cuja conquista é fruto do apoio de todos que me acompanharam ao longo desta jornada que incluiu desafios de pandemia e maternidades.

Em primeiro lugar, agradeço ao meu marido, Fernando, por seu companheirismo incansável, sua compreensão e apoio em todos os momentos. Cada desafio e cada vitória foram compartilhados com você, que esteve ao meu lado, me incentivando e encorajando a seguir em frente. Sua confiança inabalável na minha capacidade foi fundamental para que eu pudesse concluir esta etapa. Meu amor e minha gratidão por você são imensos, e sei que muitas outras conquistas virão e sempre estaremos um ao lado do outro, como força e suporte mútuo.

Aos meus filhos, Felipe e Maria Fernanda, meu agradecimento especial. Vocês foram o combustível que me manteve motivada todos os dias. Dedico a vocês esta vitória, pois sem o sorriso e o carinho de cada dia, esta caminhada teria sido muito mais árdua.

Aos meus pais, Madalena e Luiz Fernando, agradeço por sempre acreditarem em mim e por me apoiarem. Vocês estiveram presentes, torcendo e vibrando por cada conquista, por menor que fosse. Sou eternamente grata por todo amor e apoio que me fortaleceram nos momentos de incerteza.

À toda a minha família querida, meus irmãos, Henrique e Luiza, minha cunhada Natalia, minha prima Catarina e aos meus tios João, à Andreia, que estiveram ao meu lado nos finais de semana, cuidando dos meus filhos para que eu pudesse me dedicar aos estudos, deixo meu mais sincero agradecimento. A compreensão e o apoio de vocês foram fundamentais para que eu pudesse me concentrar e alcançar este objetivo.

Agradeço ainda à minha equipe da 4S Advocacia. Em especial, minha gratidão à minha amiga e sócia Ana Paula, cujo incentivo constante e parceria foram cruciais ao longo deste percurso. Agradeço por compartilhar comigo tanto o cotidiano desafiador da advocacia quanto a vida acadêmica, fazendo de cada desafio uma oportunidade de crescimento e aprendizado.

Agradeço ao meu orientador, Professor Maurício Requião, cujo apoio e orientação foram fundamentais na construção deste trabalho. Sua experiência, dedicação e conselhos preciosos foram essenciais para que eu pudesse trilhar esse caminho com confiança e alcançar este importante marco em minha vida acadêmica. Sou profundamente grata por ter tido a oportunidade de aprender com alguém que tanto admiro.

Agradeço também aos professores que gentilmente aceitaram o convite para compor a banca examinadora deste trabalho. É uma honra imensa ter o meu trabalho avaliado por

profissionais tão renomados e que são referências para mim e para muitos outros estudiosos do Direito. Agradeço pela oportunidade de aprender ainda mais com cada um de vocês.

Por fim, mas não menos importante, agradeço aos amigos e colegas que, direta ou indiretamente, contribuíram para que eu chegasse até aqui. Tê-los ao meu lado foi e sempre será uma fonte inesgotável de alegria e motivação.

“O aspecto mais triste da vida atual é que a ciência ganha em conhecimento mais rapidamente que a sociedade em sabedoria.”

— Isaac Asimov

## RESUMO

Na atualidade, os sistemas de Inteligência Artificial (IA) e seus algoritmos complexos estão remodelando profundamente o tecido social e econômico global. Estes sistemas, capazes de processar volumes massivos de dados e tomar decisões automatizadas, estão se infiltrando em praticamente todos os aspectos da vida cotidiana, desde a escolha de rotas de tráfego até decisões que impactam profundamente direitos fundamentais como os direitos à liberdade, à vida e à privacidade. A evolução do *machine learning* e do *deep learning* ampliou exponencialmente o alcance e a sofisticação desses sistemas, levantando questões cruciais sobre transparência, equidade e *accountability*.

Diante desse panorama de rápida transformação tecnológica, emergem desafios inéditos para o campo jurídico e ético. A capacidade desses sistemas de IA de influenciar e até mesmo determinar aspectos críticos da vida humana suscita debates urgentes sobre a necessidade de novos marcos regulatórios e princípios éticos. O conceito de "Trustworthy AI" ganha destaque, propondo uma abordagem centrada no ser humano, que prioriza o respeito aos direitos fundamentais e a promoção do bem-estar social. Paralelamente, cresce a demanda por mecanismos efetivos de contestação das decisões tomadas por esses sistemas automatizados, essenciais para salvaguardar a autonomia e os direitos individuais.

Esta monografia se propõe a explorar criticamente o poder crescente dos sistemas de IA como "tomadores de decisão" na sociedade contemporânea. O trabalho analisa os impactos multifacetados dessa tecnologia, com ênfase na questão crucial do direito de contestação das decisões automatizadas. Investigam-se as complexidades inerentes à implementação de uma IA confiável e ética, bem como os desafios para estabelecer estruturas legais e regulatórias que garantam a proteção dos direitos fundamentais frente a potenciais vieses e “injustiças” algorítmicas. Por fim, o estudo se debruça sobre a concepção de um *framework* robusto para o direito de contestação, articulando princípios como a autodeterminação informativa e o devido processo legal no contexto digital, visando equilibrar a inovação tecnológica com a preservação dos valores democráticos e dos direitos humanos.



**Palavras-chave**: Inteligência Artificial. Trustworthy AI. Dados pessoais. Contestação algorítmica.

## ABSTRACT

Currently, Artificial Intelligence (AI) systems and their complex algorithms are profoundly reshaping the global social and economic fabric. These systems, capable of processing massive volumes of data and making automated decisions, are infiltrating virtually every aspect of daily life, from choosing traffic routes to decisions that deeply impact fundamental rights such as the rights to freedom, life, and privacy. The evolution of machine learning and deep learning has exponentially expanded the reach and sophistication of these systems, raising crucial questions about transparency, fairness, and accountability.

Faced with this rapid technological transformation, unprecedented challenges emerge in the legal and ethical fields. The ability of these AI systems to influence and even determine critical aspects of human life raises urgent debates about the need for new regulatory frameworks and ethical principles. The concept of "Trustworthy AI" gains prominence, proposing a human-centered approach that prioritizes respect for fundamental rights and the promotion of social well-being. In parallel, there is a growing demand for effective mechanisms to contest decisions made by these automated systems, essential to safeguard individual autonomy and rights.

This monograph aims to critically explore the growing power of AI systems as "decision-makers" in contemporary society. The work analyzes the multifaceted impacts of this technology, emphasizing the crucial issue of the right to contest automated decisions. It investigates the complexities inherent in implementing reliable and ethical AI, as well as the challenges in establishing legal and regulatory structures that ensure the protection of fundamental rights against potential algorithmic biases and "injustices". Finally, the study focuses on the conception of a robust framework for the right of contestation, articulating principles such as informational self-determination and due process in the digital context, aiming to balance technological innovation with the preservation of democratic values and human rights.



**Keywords**: Artificial Intelligence. Trustworthy AI. Personal data. Algorithm challenge.

**SUMÁRIO**

# 1 INTRODUÇÃO

A Quarta Revolução Industrial[1], na qual a sociedade global se encontra, se caracteriza por uma transição social movida por novos sistemas construídos a partir dos avanços tecnológicos possuidores de conectividade. Estes foram capazes de gerar uma verdadeira revolução digital. Ergueu-se, assim a necessidade de se repensar as relações interpessoais bem como o desenvolvimento, a criação e a interação das pessoas com a tecnologia.

A combinação de novas tecnologias, que se somam com um imenso volume de dados pessoais coletados de forma cotidiana - muitas vezes sem o conhecimento ou aceite dos titulares dos dados - especialmente as informações e dados derivados de transações e relações mediadas por computadores - trazem uma nova lógica social e econômica à qual o Direito não pode se omitir.

Para entender este fenômeno, é importante compreender o funcionamento de algoritmos e como a evolução do seu uso trouxe desafios complexos.

Em termos gerais, um algoritmo é um conjunto de instruções, como uma receita ou regras para um jogo. É uma sequência de operações ou regras que, quando aplicadas a um conjunto de dados, permitem resolver problemas semelhantes. Na área de informática, refere-se a um conjunto de regras e procedimentos lógicos bem definidos que levam à solução de um problema em várias etapas. Podem ser entendidos como as diretrizes seguidas por uma máquina, maneiras de representar matematicamente um processo estruturado para realizar uma tarefa[2]. E são estes algoritmos que constituem a base sobre a qual os sistemas de Inteligência Artificial (IA) são desenvolvidos.

A Inteligência Artificial pode ser entendida como a capacidade de máquinas realizarem tarefas típicas da inteligência humana, como planejamento, compreensão de linguagem natural utilizada pelo humano, o reconhecimento de objetos e sons, o aprendizado, o raciocínio e a resolução de problemas diversos. Em essência, Inteligência Artificial é o campo que se dedica à criação e ao desenvolvimento de sistemas computacionais capazes de executar funções que normalmente requerem inteligência humana.[3]

---

[1] Conceito trabalhado por Klaus Schwab, CEO do Fórum Econômico Mundial. Disponível em: https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/. Acesso em: 06 abr. 2020.

[2] ELIAS, Paulo Sá. Algoritmos, inteligência artificial e o direito. Consultor Jurídico, 2017. Disponível em: https://www.conjur.com.br/dl/algoritmos-inteligencia-artificial.pdf. Acesso em: 04 fev. 2021.

[3] ELIAS, Paulo Sá. Algoritmos, inteligência artificial e o direito. Consultor Jurídico, 2017. Disponível em: https://www.conjur.com.br/dl/algoritmos-inteligencia-artificial.pdf. Acesso em: 04 fev. 2021.

O estágio atual da Inteligência Artificial já permite avançada compreensão visual, de objetos e rostos, como nas tecnologias de reconhecimento fácil, o reconhecimento de voz e compreensão da linguagem humana, o que se verifica nos assistentes virtuais como a Siri[4] ou Alexa[5], e também a tomada de decisões complexas.

A despeito de Inteligência Artificial ser tema recorrente na imaginação e nas projeções humanas há décadas, ela veio a crescer e se tornar uma realidade em razão do desenvolvimento de *graphic processing units* (tecnologia de processamento e sistematização de mineração de dados[6]) e do *big data*[7] (fenômeno decisivo para a criação de uma nova lógica econômica).

O conceito de *big data* não deve ser entendido apenas como o grande volume de dados que é gerado diariamente por empresas, dispositivos, sensores, clicks e outras fontes. Este fenômeno engloba também a velocidade da criação e capacidade de processamento dos dados, a variedade e diversidade dos tipos de dados processados (estruturados, semiestruturados e não estruturados, ou seja, desde banco de dados até textos, vídeos e imagens), e a capacidade de armazenar e analisar esses dados para extrair insights valiosos e tomar decisões informadas a partir deles[8]. Assim, só foi possível explorar a conjunção deste potencial de valor do *big data* por meio de algoritmos e da IA.

Uma subárea essencial dentro do campo mais amplo da Inteligência Artificial é o aprendizado de máquinas ou *machine learning*[9], que utiliza algoritmos para coletar informações, aprender com elas e, assim, chegar a algum resultado programado, o que pode ser, por exemplo, uma determinada projeção ou até mesmo uma tomada de decisões sobre alguma tarefa ou assunto. Novos dados vão moldando e alimentando o algoritmo com a mínima intervenção ou até mesmo sem a necessidade do direcionamento humano.

---

[4] Assistente virtual da Apple. Informações disponíveis em: https://www.apple.com/br/siri/. Acesso em: 10 jul. 2024.

[5] Assistente virtual da Amazon. Informações disponíveis em: https://www.amazon.com.br/b?ie=UTF8&node=19949683011. Acesso em: 10 jul. 2024.

[6] De acordo com Rouvroy e Berns (2018) a mineração de dados consiste "no tratamento automatizado de quantidades massivas de dados de modo a fazer emergir correlações sutis entre eles". Disponível em: ROUVROY, Antoniette: BERNS, Thomas. Governamentalidade algotítmica e perspecticas de emancipação: o díspar como condição de individuação pela relação. *In:* BRUNO, Fernanda *et al*. **Tecnopolíticas da vigilância:** perspectivas da margem. São Paulo: Boitempo, 2018.

[7] CONSTANTIOU, Ioanna; KALLINIKOS, Jannis. New games, new rules: big data and the changing context of strategy. **Journal of Information Technology**, v. 30, n. 1. [S.l.] 2015. Disponível em: https://eprints.lse.ac.uk/63017/1/Kallinikos_New%20Games%20New%20Rules.pdf . Acesso em 10 out. 2023.

[8] DE MAURO, Andrea; GRECO, Marco; GRIMALDI, Michele. A formal definition of Big Data based on its essential features. **Library Review**, v. 65 n. 3, p. 122-135. Disponível em: https://doi.org/10.1108/LR-06-2015-0061. Acesso em: 10 jul. 2024.

[9] ROUVROY, Antoniette: BERNS, Thomas. Governamentalidade algotítmica e perspecticas de emancipação: o díspar como condição de individuação pela relação. *In:* BRUNO, Fernanda *et al*. **Tecnopolíticas da vigilância**: perspectivas da margem. São Paulo: Boitempo, 2018.

Dentre deste campo de *machine learning,* foi desenvolvido o *deep learning*[10], ou aprendizado profundo, que utiliza algoritmos mais complexos e foi inspirado nas funções do cérebro humano, na forma de interconexão dos neurônios, surgindo a ideia de redes neurais artificiais. O *deep learning* utiliza redes neurais artificiais para modelar padrões complexos a partir de grandes volumes de dados. E, diferentemente dos algoritmos tradicionais de *machine learning*, que podem necessitar que algumas características sejam extraídas manualmente, o *deep learning* é capaz de descobrir automaticamente, sem intervenção, essas características.

> Esta solução permite que os computadores aprendam com a experiência e entendam o mundo em termos de uma hierarquia de conceitos, com cada conceito definido em termos de sua relação com conceitos mais simples. Ao reunir conhecimento a partir da experiência, essa abordagem evita a necessidade de operadores humanos especificarem formalmente todo o conhecimento que o computador precisa. A hierarquia de conceitos permite que o computador aprenda conceitos complicados construindo-os a partir de conceitos mais simples.[11]

Tecnicamente toda essa evolução é impressionante e fascinante, ocorre que o progresso científico não está necessariamente associado a um progresso social e é ilusório se pensar que a tecnologia, a Inteligência Artificial e as decisões tomadas por estas são neutras. Muito pelo contrário. Estes mecanismos de automação são capazes de gerar profundos impactos sociais e até mesmos antiéticos e perigosos, pondo em risco diversos direitos humanos fundamentais.

Ademais, muitas vezes algoritmos criados para facilitar a vida de quem os utiliza e com fins éticos acabam derivando de forma a tomar decisões indesejadas que sejam preconceituosas e discriminatórias.

No campo hipotético, não haveria nada de errado nem problemático em se utilizar dados disponíveis para embasar uma tomada de decisão. Intuitivamente se chega até mesmo à conclusão de que quanto maior o campo de informações mais bem tomada, objetiva e embasada poderá ser uma decisão.

Ocorre que o problema surge quando se tem informações enviesadas, distorcidas, dados privados ou referências sem os devidos filtros e estes são utilizados para alimentar a tomada de decisão. O que é agravado quando a decisão é automatizada e não se tem controle sobre os

---

[10] GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning.** Cambridge: MIT Press, 2016. Disponível em: http://imlab.postech.ac.kr/dkim/class/csed514_2019s/DeepLearningBook.pdf. Acesso em: 10 jul. 2024.

[11] *Ibidem*. p. 1 (Tradução nossa).

recortes que deveriam ser aplicados para garantir a qualidade do resultado ou *outcome* da decisão[12].

Considerando o avanço tecnológico como inevitável, e até mesmo desejável, intui-se que este deve ser centrado no bem-estar da pessoa. Para isso, o controle dos possíveis impactos sociais, econômicos e políticos da Inteligência Artificial deve ser cuidadosamente pensado, garantindo-se o controle desde a sua concepção até o fim da sua cadeia de uso para que os direitos fundamentais sejam respeitados e postos como prioridade.

Soma-se a isso as dificuldades que o indivíduo enfrenta em eventual tentativa de contestar alguma decisão tomada de forma automatizada sobre ele. Fatores como a opacidade e a complexidade no entendimento e linguagem técnica trazem limitações reais para a existência de efetivo "direito individual de defesa".

Dessa forma, o problema ao qual este trabalho se propõe a explorar é o poder de contestar decisões tomadas por meio de IA. Em um contexto em que algoritmos e sistemas automatizados estão cada vez mais presentes em processos decisórios, a capacidade dos indivíduos de questionar, revisar e, se necessário, reverter as decisões ou o impacto destas decisões se torna crucial. Surge, assim, uma busca por entender como as estruturas legais e regulamentares podem ser moldadas para assegurar que os direitos fundamentais dos indivíduos sejam protegidos contra decisões automatizadas que possam ser injustas, enviesadas ou prejudiciais. Sem possibilidade de contestação, parece não ser possível um efetivo direito de defesa, garantindo que os afetados por essas tecnologias tenham as ferramentas e os recursos necessários para se proteger de impactos significantes nas suas vidas.

Revela-se premente a compreensão a respeito de uma nova agenda humana que tem como fundo um novo paradigma tecnológico a demandar profundas reflexões sobre os princípios que devem nortear a sua aplicação.

Já se convive amplamente com algoritmos que "controlam" ou "induzem" a vida de milhões de indivíduos, determinando o caminho de carro que o sujeito deve adotar, que tipo de alimento este deve comer, onde ele deve trabalhar, morar etc.

Apesar de haver um racional explícito por detrás destes direcionamentos de comportamento humano, nem sempre este racional é conhecido e consentido. Ao usar um

[12] Caso de grande notoriedade foi o caso do Programa COMPAS, que com o objetivo de se ter um algoritmo capaz de auxiliar o sistema judiciário norte americano na tomada de decisões, calculando o grau de periculosidade de um criminoso e tirando subjetividades e possíveis erros humanos na análise, quando diante de perfis idênticos aumentava a pontuação em quase cinquenta porcento quando o sujeito era negro. Caso disponível em: https://noticias.uol.com.br/tecnologia/noticias/redacao/2018/04/24/preconceito-das-maquinas-como-algoritmos- tomam- decisoes-discriminatorias.htm. Acesso em: 10 abr. 2019.

aplicativo de carro para sair do engarrafamento (racional, conhecido e consentido), será que o caminho indicado tem apenas este fundamento? Será que o algoritmo não possui outros *inputs* que determinam e influenciam a tomada de decisão, como, por exemplo, interesses econômicos de que o seu veículo passe por determinado local – comércio local, pedágio, etc.?

E se a utilização destas tecnologias passasse a ser não apenas opcionais e sim obrigatórias? Que tipo de garantia teríamos sobre o campo de liberdade que poderia estar sendo sacado ou até mesmo que tipo de transparência se teria sobre o racional por detrás de eventuais restrições de direitos? Como poderíamos ter transparência e contestar eventuais decisões? Até que medida os segredos comerciais e industriais são barreiras para a transparência sobre o processo de tomada de decisão por meio de IA?

Emerge a consciência de se explorar juridicamente a concepção de uma *Trustworthy* AI, ou seja, é essencial que a Inteligência Artificial seja centrada na pessoa e no bem-estar desta, respeitando direitos fundamentais, regulamentações, princípios e voltada a um fim ético e justo. Complementar a isso, essa deve ser robusta em termos técnicos e sociais, sendo confiável e operando de forma transparente e "contestável".

Analisando o Direito neste processo, Boaventura[13] o distingue em dois vetores. O primeiro se refere à capacidade dele de regular as novas tecnologias, os riscos sociais decorrentes destas, bem como o agir repressivamente punindo as atividades danosas advindas do seu uso. O segundo ângulo, por sua vez, apesar de relacionado, com o primeiro não se confunde pois diz respeito ao impacto da expansão das novas tecnologias e a necessidade de minimizar impactos negativos.

Aqui, pautamos a questão no "impacto" e a partir daí miramos o vetor "regulatório", como trazido por Boaventura ao constatar que "inversamente a questão do impacto das novas tecnologias é o outro lado da constatação da incapacidade ou da ineficácia regulatória do direito"[14].

Muitas esferas normativas estão de certa medida passando pela construção de estruturas normativas que levam em conta a elaboração de matrizes de riscos sociais, analisando probabilidade/impacto e calibrando as necessidades regulatórias a partir dessa estrutura

---

[13] SANTOS, Boaventura de Sousa. Os tribunais e as novas tecnologias de comunicação e de informação. **Sociologias**, Porto Alegre, v. 7, n. 13, jan./jun., 2005, p. 82-109. Disponível em: http://www.boaventuradesousasantos.pt/media/Tribunais%20e%20novas%20tecnologias_Sociologias_2005(1).pdf. Acesso em: 10 out. 2023.

[14] *Ibidem*.

analítica[15]. Pretende-se trabalhar além de uma concepção preventiva a respeito da demanda, antecipando possíveis impactos que as decisões tomadas por meio de sistemas de Inteligência Artificial poderiam ter sobre os direitos fundamentais, mas igualmente pensando no momento pós concretização dos riscos, no que pode ser efetivamente feito em termos de proteção de indivíduo.

Não que o foco das propostas de regulamentação de Inteligência Artificial não deva se concentrar na prevenção, este certamente é um melhor caminho do que o da remediação. Mas, ocorrido o dano ou na suspeita da ocorrência do dado, o que resta ao indivíduo? Contestar? As pessoas devem ter o direito específico de "contestar" previsto por lei quando são submetidas à tomada de decisão por inteligência artificial?

E se a resposta for sim, como seria esse direito? Eles afetarão a promessa de eficiência de custo que as tomadas de decisão por meio de sistemas de Inteligência Artificial prometem tomar? E se o devido processo legal for mal desenhado ou implementado? Haveria uma contestação significativa?

Para a elaboração desta pesquisa, foi adotada uma abordagem metodológica qualitativa, caracterizada pela análise bibliográfica e documental. Foram analisados livros, artigos acadêmicos, periódicos, *websites*, legislações e relatórios de organizações nacionais e internacionais, com o objetivo de compreender o contexto teórico e prático da temática. Foram utilizados os métodos dedutivos, hipotético dedutivos e genealógicos, com análise das condições sociais, políticas e econômicas que influenciaram as transformações analisadas neste trabalho.

Quanto ao conteúdo do presente trabalho, no primeiro capítulo abordamos o tema de como os sistemas de tomada de decisão por meio de algoritmos automatizados ampliam o poder de empresas e governos, regulando cada vez mais aspectos de nossas vidas e como as suas dimensões de poder podem ser difíceis de entender. Exploramos como para essa engrenagem funcionar, são necessários dados sobre dinâmicas econômicas, sociais, políticas, culturais e sobre os indivíduos e como a ciência de dados tem impactado a sociedade pela rápida proliferação de regras de decisão orientadas por dados.

Ainda neste primeiro capítulo trazemos uma luz sobre algumas questões primordiais sobre a aplicação deste tipo de sistema, analisando as dinâmicas de poder no que se refere a "força" da decisão tomada por estes sistemas, a formação de perfis (influenciando inclusive no

[15] ZANATTA, Rafael A. F. Proteção de dados pessoais como regulação do risco: uma nova moldura teórica? *In*: **I Encontro da Rede de Pesquisa em Governança da Internet**. [on-line]. 2017. Disponível em: http://www.redegovernanca.net.br. Acesso em: 10 maio 2023.

direcionamento de informações), a obscuridade destas decisões e o fenômeno dos vieses discriminatórios.

O segundo capítulo, por sua vez, tem o objetivo de analisar o conceito que vem sendo debatido mundialmente da "*Trustworthy* AI", ou seja, da construção de uma Inteligência Artificial centrada na pessoa e no bem-estar desta, respeitando direitos fundamentais, regulamentações, princípios e voltada a um fim ético e justo. Para isso analisamos e selecionamos o que consideramos como os componentes essenciais para uma Inteligência Artificial de confiança, ou seja, os princípios éticos que deve estar presentes no seu desenvolvimento e uso (Ser humano como centro e Respeito a autonomia humana, Prevenção de danos, Equidade, Transparência e Explicabilidade e Prestação de contas e *Accountability*. Assim, ainda neste capítulo trabalhamos a ideia de como é parte fundamental da confiança a possibilidade real de contestação das decisões tomadas por meio destes sistemas e como isto tem íntima relação com os riscos e a governança algorítmica.

Por fim, no terceiro capítulo temos o objetivo de compreender as formas em que o direito de uma contestação de decisão de sistema automatizado por meio de algoritmos de Inteligência Artificial pode ser desenhado. Para isso traçamos relações de dependência entre a autodeterminação informativa e o direito de contestação das decisões que utilizam os dados pessoais dos sujeitos da decisão e, por fim, trabalhamos a ideia do devido processo legal e possíveis arquétipos de contestação.

## 2 DECISÕES ALGORÍTMICAS

O momento histórico e social tem sido objeto de discussão de estudiosos de áreas múltiplas. A depender do enfoque disciplinar, este contexto ganha nomes e prismas diversos. Alguns falam em capitalismo informacional[16], em sociedade ou civilização da informação[17], em capitalismo de vigilância[18], sociedade em rede[19], hipermodernidade[20] [21] , *data capitalism*[22] etc.

Para Marcos Dantas, o capitalismo informacional é a fase atual do capitalismo onde as tecnologias de informação e comunicação se tornaram os principais fatores de produção e de criação de valor e um recurso essencial para a acumulação do capital, tendo a capacidade cognitiva e a cooperação social como elementos chave para a valorização do capital[23].

Já na concepção do "*capitalismo de vigilância*" desenvolvido por Shoshana Zuboff, vivemos em um momento no qual é vista uma ascensão do domínio das grandes corporações tecnológicas sobre as relações humanas e as relações de poder, e este novo "modelo de capitalismo", o de vigilância, seria um modelo econômico e social que se baseia na coleta, análise massiva e na mercantilização de dados pessoais. Seria "uma nova ordem econômica que

---

[16] DANTAS, Marcos. Información, capital y trabajo: valorización y apropiación en el ciclo de la comunicación productiva, **Escribana**. Universidade de Manizales, jul.-dez., 2002. p. 21-48. Disponível em: https://dialnet.unirioja.es/servlet/articulo?codigo=6986986. Acesso em: 10 maio 2023.

[17] ZUBOFF, Shoshana. Big Other: capitalismo de vigilância e perspectivas para uma civilização informação. *In:* BRUNO, Fernanda *et al.* **Tecnopolíticas da vigilância:** perspectivas da margem. São Paulo: Boitempo, 2018.

[18] ZUBOFF, Shoshana. **A era do Capitalismo de Vigilância**: A luta por um futuro humano na nova fronteira do poder. Rio de Janeiro: Intrínseca, 2021.

[19] CASTELLS, Manuel. **A sociedade em rede**. v. 1 ed. 2. São Paulo: Paz e Terra, 1999.

[20] LIPOVETSKY, Gilles. **O império do efêmero:** a moda e seu destino nas sociedades modernas. Tradução Maria Lucia Machado. 7 ed. São Paulo: Companhia das Letras, 2004.

[21] Para o filosofo francês, a sociedade passa por um período denominado de hipermodernidade e este tem como um dos seus marcos o revivalismo ético ocorrendo por meio de um processo simultâneo que envolve a desorganização e a reorganização da ética.

[22] "No regime denominado por Mayer-Schönberger e Ramge (2018) de data capitalism, o preço perde sua centralidade, os agentes utilizam os dados para identificar better matches explorando várias dimensões, em uma transição do capitalismo financeiro para o capitalismo de dados. No primeiro, a informação, difícil e cara, convergia para o "preço"; no segundo, a informação é múltipla, complexa, rápida e barata. Os autores identificam três tecnologias-chave:15 (a) linguagem padrão para comparar e compartilhar os dados sobre os bens e as preferências, (b) capacidade para identificar "matches" em várias dimensões e selecionar os parceiros e transações adequados, e (c) capturar e usar as preferências de maneira eficaz (assertividade). Para os autores, os dados estão substituindo o preço como elemento estrutural da relação produtor e consumidor, e a moeda como meio de pagamento. Hoje, já pagamos vários serviços com dados (pesquisa no Google, benefícios do Facebook – relações sociais e plataforma de negócios) e, em breve, essa prerrogativa deve se estender às anuidades dos cartões de crédito, as taxas bancárias e aos custos da telefonia, setores que concentram grandes volumes de dados de seus clientes. KAUFMAN, Dora. *O protagonismo dos algoritmos da Inteligência Artificial*: observações sobre a sociedade de dados. Teccogs: **Revista Digital de Tecnologias Cognitivas**, TIDD | PUC-SP, São Paulo, n. 17, p. 44-58, jan.-jun. 2018. p. 50-51. Disponível em: https://revistas.pucsp.br/index.php/teccogs/article/view/48589/32069. Acesso em: 05 set. 2022.

[23] DANTAS, Marcos. Información, capital y trabajo: valorización y apropiación en el ciclo de la comunicación productiva, **Escribania.** Universidade de Manizales, jul.-dez., 2002. p. 21-48. Disponível em: https://dialnet.unirioja.es/servlet/articulo?codigo=6986986. Acesso em: 10 maio 2023.

reivindica a experiência humana como matéria-prima gratuita para práticas comerciais dissimuladas de extração, previsão e vendas”[24].

Contudo, independente do nome ou recorte feito, todos esses estudos reconhecem, em alguma medida, que existe o incentivo cada vez maior para a integração de decisões automatizadas em processos sociais e econômicos valiosos. Diariamente algoritmos automatizados tomam decisões que são capazes de ampliar o poder de empresas e governos e, à medida que os algoritmos passam a regular mais aspectos de nossas vidas, as dimensões de seu poder podem permanecer difíceis de entender.

E para essa engrenagem funcionar, são necessários dados sobre os mais diversos aspectos da dinâmica econômica, social, política, cultural e sobre os sujeitos objetos. Assim, a ciência de dados tem impactado a sociedade por meio da rápida proliferação de regras de decisão orientadas por dados.

De forma resumida, falar em decisões algorítmicas, significa falar em decisão automatizada e realizada por computadores usando regras, mas também significa tratar do fato de que as decisões são também orientadas por dados, e que observações feitas no mundo real formam o material bruto sobre o qual os algoritmos atuam. Assim, tanto a natureza dos dados quanto a estrutura do algoritmo são importantes na determinação das características de desempenho de uma regra de tomada de decisão, o que demanda para este estudo uma abordagem bifocal.

É assim que avaliadores de crédito, navegadores e sites de busca, aplicativos de locomoção, bancos, redes sociais e instituições públicas precisam construir as regras de decisão ao serem programados, mas também precisam coletar ou serem alimentados por dados sobre a população/usuários, estabelecem regras com base nas características e o acesso a estes dados e os convertem em pontuações, classificações, cálculos de risco e até mesmo listas de observações com consequências vitalmente importantes[25].

E aqui se evidencia um aspecto central e de debate urgente, os riscos em se delegar decisões relevantes para sistemas. Os sistemas são “inteligentes” o suficiente para estas tarefas? Quem pode legitimar um sistema para tomar determinada decisão? Os sistemas estão preparados? A população deve se submeter a estas decisões?

---

[24] ZUBOFF, Shoshana. **A era do Capitalismo de Vigilância:** A luta por um futuro humano na nova fronteira do poder. Rio de Janeiro: Intrínseca, 2021.

[25] PASQUALE, Frank. **The Black Box Society:** The Secret Algorithms That Control Money and Information. Cambridge, Harvard University Press, 2015.

As decisões tomadas por algoritmos podem se fundamentar em técnicas de aproximação, regras pré-estabelecidas, ou em análises de grandes conjuntos de dados. Essas regras podem ser definidas diretamente por programadores ou podem ser adaptáveis e flexíveis, ajustando-se com base no aprendizado automático dos dados[26].

Em certos casos, esses sistemas algorítmicos podem substituir completamente as decisões humanas, mas é possível que um ser humano ainda mantenha algum "controle" na tomada de decisão final no processo, havendo uma tomada de decisão hibrida. Porém, mesmo nesses casos, o algoritmo tende a influenciar – nem sempre de forma consciente - o operador, direcionando sua atenção para um conjunto específico de informações como observa Nicholas Diakopoulos:

> Decisões algorítmicas podem ser baseadas em heurísticas e regras, ou cálculos sobre imensas quantidades de dados. As regras podem ser articuladas diretamente pelos programadores, ou ser dinâmicas e flexíveis baseadas na aprendizagem de máquina dos dados. Às vezes, um operador humano mantém a autonomia e toma a decisão final em um processo, mas mesmo neste caso o algoritmo enviesa a atenção do operador para um subconjunto de informações.[27]

Eduardo Magrani[28], ao tratar da governança da Internet das Coisas (IoT) traz a ideia de que diferentemente de objetos cotidianos, que seriam tecnologias "domesticadas" pelos humanos, os algoritmos inteligentes utilizados nas tomadas de decisão ainda são tecnologias "não domesticadas", uma vez que o tempo de interação entre a tecnologia e a pessoa ainda não permite a previsibilidade dos riscos para fins de controle ou eliminação.

À medida que a autonomia na tomada de decisão e as responsabilidades crescem, as reflexões morais também vão aumentando. Os algoritmos deveriam ser programados também com "decisão moral"? Como lidar com a "moralidade das coisas"? Para Peter-Paul Verbeek[29], à medida que as máquinas operam cada vez mais em ambientes sociais considerados abertos, como esferas públicas conectadas, seria cada vez mais importante projetar um tipo de moralidade funcional que seja sensível a características eticamente relevantes e aplicável a situações pretendidas.

---

[26] O'NEIL, Cathy. **Algoritmos de destruição em massa**: como o big data aumenta a desigualdade e ameaça a democracia. 1 ed. São Paulo: Editora Rua do Sabão, 2020.

[27] DIAKOPOULOS, Nicholas. Algorithm accountability: journalistic investigation of computational power structures. **Digital Journalism**, v. 3, n. 3, p. 398, 2015. p. 3. (Tradução nossa).

[28] MAGRANI, Eduardo. New perspectives on ethics and the laws of artificial intelligence. **Internet Policy Review**, v. 8, n. 3. 2019. Disponível em: https://doi.org/10.14763/2019.3.1420. Acesso em: 10 set. 2023.

[29] VERBEEK, Peter. **Moralizing Technology**: Understanding and Designing the Morality of Things, Chicago - London, The University of Chicago Press, 2011. Disponível em: https://philpapers.org/rec/VERMTU. Acesso em: 04 fev. 2024.

Analisaremos como os algoritmos podem ter impactos profundos na vida das pessoas que, na maior parte, não têm consciência mínima sobre o fato de que eles estão tomando diversas decisões com reflexos direto em suas vidas.

Contudo, considerando que a tomada de decisão algorítmica tem as suas raízes no campo da estatística, por exemplo, no trabalho sobre teoria da decisão (que trata do raciocínio subjacente às escolhas de um agente)[30] e na ciência da computação, especialmente no subcampo da inteligência artificial, devemos antes, para entender a sua estrutura e desenho atual, devemos passar brevemente sobre a trajetória que trouxe até este momento.

Embora a criação de algoritmos não seja algo novo neste século, muitos dos desafios e riscos que enfrentamos atualmente são. É evidente que, ao definirmos algoritmo como uma sequência de passos para resolver um problema ou executar uma tarefa, podemos facilmente reconhecer que essa técnica tem sido empregada há décadas e inclusive serviu como base para o desenvolvimento dos sistemas de computação.

Nos últimos anos, porém, assiste-se a uma transição social movida por novos sistemas construídos a partir de avanços tecnológicos possuidores de conectividade que estão gerando uma revolução digital. Tem-se falado cada vez mais sobre o universo algorítmico e suas implicações na vida cotidiana dos indivíduos, o que naturalmente atraiu a atenção dos profissionais e estudiosos do Direito para o tema da possibilidade de regulação da atividade de desenvolvimento de algoritmos, o que constitui uma demanda em busca de segurança jurídica.

Como bem apontam Bruno Feigelson e Carolina Braga, a inexistência de um marco regulatório claro e específico possibilita que o tema seja "regulado e regulamentado por diferentes instrumentos normativos setoriais" e ainda sofra "reflexos de normas internacionais", algo que certamente provocará muito mais confusão do que certeza sobre a melhor forma de proceder no exercício da atividade de desenvolvimento algorítmico.

Assim, a interação das pessoas com os algoritmos e suas escolhas têm crescido consideravelmente, resultando em um aumento dos riscos associados. Riscos estes que não se limitam apenas à segurança física, mas também afetam o bem-estar psíquico.

---

[30] A teoria da decisão preocupa-se com o raciocínio subjacente às escolhas de um agente, quer se trate de uma escolha mundana entre apanhar um autocarro ou de um táxi, ou uma escolha mais abrangente sobre seguir ou não uma carreira política exigente. (Observe que "agente" aqui significa uma entidade, geralmente uma pessoa individual, que é capaz de deliberar e agir.) O pensamento padrão é que o que um agente escolhe fazer em qualquer ocasião é completamente determinado por suas crenças, desejos ou valores, mas isso não é incontroverso, como será observado a seguir. Em qualquer caso, a teoria da decisão é tanto uma teoria de crenças, desejos e outras atitudes relevantes quanto uma teoria da escolha; o que importa é como estas diversas atitudes (chamemos-lhes "atitudes de preferência") se articulam entre si". STEEL, Katie; STEFÁNSSON, H. Orri. **Decision Theory.** 2015; revisado em 2020. Disponível em: https://plato.stanford.edu/entries/decision-theory/. Acesso em: 04 fev. 2024. (Tradução nossa).

Mesmo quando os algoritmos são constituídos de forma objetiva para evitar a interferência do elemento subjetivo no processo de decisão, invariavelmente existe uma margem de subjetividade no processo de sua construção, porquanto conduzido por pessoas e suas escolhas diante de determinados cenários.

Um simples experimento é digitar no campo de busca do Google “cabelo feio” ou “trança feia”. Fazendo isso na presente data o usuário se depara com o destaque de fotografias de mulheres negras e, ao contrário, ao procurar por “cabelo bonito” ou “trança bonita”, aparecem em primeiro lugar imagens de mulheres brancas.

O algoritmo do Google é programado para elencar os resultados segundo determinados critérios, definidos por sujeitos que carregam consigo em seus comportamentos traços inequívocos de subjetividade e com base em dados produzidos ou outra carga de subjetividade, o que acaba dando ensejo a indesejáveis tratamentos discriminatórios.

Um exemplo é a coleta e o tratamento de dados para fins de programação de algoritmos para as atividades de marketing. É que a partir da coleta de dados de “perfis” diferentes de consumidor, dispositivos eletrônicos dotados de inteligência artificial podem definir determinados modelos (“personas”) destinatários de campanhas com o objetivo de comercializar produtos e/ou serviços.

É assim que informações sobre idade, escolaridade, o poder aquisitivo, local de residência, status de relacionamento, formação acadêmica, perfil familiar etc. podem ser muito úteis para fins de precificação, modelagem dos planos de contratação de produtos e/ou serviços ofertados no mercado, dentre outros. Há, pois, um direcionamento dos algoritmos e das vendas segundo determinados critérios fixados pelos seus desenvolvedores.

Plataformas de redes sociais, como o Instagram, oferecem aos usuários a opção de promover suas postagens para alcançar um público específico. O conteúdo promovido é distribuído de acordo com os algoritmos da plataforma. Não são os usuários que selecionam o conteúdo publicitário que veem, nem os anunciantes que escolhem individualmente quem verá seus anúncios. É o próprio programa da plataforma que decide quais peças publicitárias serão mostradas aos usuários, com base em seus perfis, os quais são definidos pelos dados coletados pelas empresas.

Igualmente são exibidas para cada usuário informações e conteúdos opinativos diferentes, de acordo com a experiência por eles vivenciada na rede ou fora dela, mas que de alguma forma foi integrado, revelando-se uma tendência à indicação de notícias sobre temas do seu habitual interesse, bem como de opiniões de pessoas com pensamento semelhante, o que termina por ampliar e fortalecer bolhas sociais.

Da mesma forma, ofertas de produtos à venda, oportunidades de emprego, divulgação de eventos, muitas vezes são direcionados a pessoas específicas com base em dados coletados por seus dispositivos eletrônicos. Os algoritmos são responsáveis por processar esses dados e distribuir o conteúdo correspondente. Esse cenário pode propiciar discriminações com base em características como idade, gênero, orientação sexual, raça, entre outros, conforme se verá mais adiante.

Surge, então, a questão crucial sobre o controle da natureza, do conteúdo e da abrangência dos dados coletados pelos dispositivos que não é conduzido pelo desenvolvedor do algoritmo. Como resultado, a decisão tomada pelo algoritmo com base nesses dados muitas vezes escapa ao controle do seu criador. Isso gera uma espécie de "autonomia" da linguagem codificada para tomar decisões, e uma imprevisibilidade em relação ao conteúdo dessas decisões, pois o desenvolvedor muitas vezes desconhece o conteúdo que alimenta as máquinas. Consequentemente, o algoritmo pode tomar decisões discriminatórias sem a intervenção do seu criador, que pode estar impossibilitado de evitar essas situações devido à falta de conhecimento sobre a veracidade dos dados ou à incerteza quanto à integridade dos dados necessários para uma tomada de decisão adequada.

Existem ainda sistemas que estão sendo desenvolvidos para a tomada de decisão que podem já estar decidido de forma direta sobre a integridade física dos usuários. É o caso por exemplo dos veículos autônomos. Haverá situações cotidianas em que os algoritmos responsáveis pela condução do veículo farão escolhas que podem vir a colocar em risco a segurança do passageiro, como por exemplo: (i) a ultrapassagem de um semáforo com a luz amarela acesa; (ii) a transição de uma faixa para outra; (iii) a identificação de irregularidades ou mesmo de buracos na pista em condições climáticas adversas, dentre outras.

Ou em situações ainda mais extremas, como na Máquina Moral[31] desenvolvida pelo MIT na qual a plataforma coleta a perspectiva humana em relação às decisões morais feitas pela inteligência das máquinas em carros autônomos. Para isso, são trazidos dilemas morais, onde um carro sem motorista deve escolher entre o menor dos males. São situações nas quais o julgador opta por privilegiar a vida de quem está dentro do veículo ou fora, se a vida de uma criança tem valor menor ou maior do que a de um idoso etc.

Importante observar que os casos de decisão automatizada por meio de algoritmos inteligentes não dizer respeito apenas à iniciativa privada, também a Administração Pública vem sendo utilizados sistemas de tomada de decisão algorítmica em cada vez mais larga escala

[31] Disponível em: https://www.moralmachine.net/hl/pt. Acesso em: 04 jan. 2024.

– muitas vezes por meio de contratos celebrados com empresas privadas que os fornecem no mercado.

A utilização destes sistemas pela Administração Pública na concepção e na execução de políticas públicas, por exemplo, possibilita em inestimável ganho de eficiência. Contudo, ocorre que, diante do desconhecimento da população e dos próprios gestores públicos sobre a metodologia de desenvolvimento do sistema de algoritmos contratado. Este fato põe em xeque, inclusive, o próprio princípio democrático que visa garantir a transparência e auditoria dos mecanismos de gestão adotados pela Administração muitas vezes em razão do sigilo que sobre ele se atribui.

Por meio desses sistemas, decisões importantes cabíveis ao Estado podem vir a ser delegadas, tais como (i) a definição de políticas públicas de segurança, com a realização de atividades em determinadas regiões, em vez de outras; (ii) a alocação e o direcionamento de verbas públicas para certas categorias de pessoas; (iii) a escolha do local onde serão prestados determinados serviços de saúde para a população; dentre outras, podem ser objeto direto da influência de algoritmos.

Isso porque os algoritmos indicarão, de forma preditiva, as medidas "mais adequadas" para a solução dos problemas que lhes foram postos, orientando a atividade dos gestores que são os efetivos responsáveis pelas escolhas políticas a serem realizadas. O potencial discriminatório dessas opções, no entanto, é evidente, visto que a análise feita por algoritmos não só limita os critérios no processo decisório àqueles segundo os quais eles são orientados, como exclui os dados não processados pelos dispositivos que controlam, cingindo-se, pois, ao estudo parcial da realidade sobre a qual os seus usuários se debruçam na gestão da coisa pública.

As consequências escapam ainda mais ao controle com o crescimento da autonomia das decisões tomadas com base em Inteligência Artificial e os fatos se mostram ainda mais delicados quando as empresas desenvolvedoras argumentam no sentido do sigilo sobre as técnicas de desenvolvimento algorítmico (realidade predominante nas chamadas "*Big Techs*" diante da ausência de regulação do desenvolvimento algorítmico), potencializando a insegurança face ao total desconhecimento dos riscos associados à sua atividade.

Nesse sentido, merece reprodução uma desconcertante colocação feita por Frank Pasquale na obra "*The Black Box Society*: *The secret Algorithms that control Money and information*", quando da análise a respeito de tentativas de se regular a atividade de empresas de tecnologia e financeiras: como vamos fazer isso sobre um objeto do qual não sabemos quase nada a respeito?

## 2.1 ALGORITMOS PARA LIDAR COM A INCERTEZA?

A incerteza é intrínseca e inevitável dentro de um processo de tomada de decisão no mundo contemporâneo de reflexos cada vez mais multifacetados e abrangentes. A falta de informações completas, a imprevisibilidade de eventos futuros, especialmente em razão de fatores como a complexidade e velocidade do avanço de novos sistemas e tecnologias, a volatilidade econômica, eventos climáticos imprevisíveis, mudanças políticas e a interconexão global, trazem um aumento ainda maior da incerteza nas tomadas de decisão.

Nesse cenário, decisões importantes são frequentemente tomadas diante de variáveis desconhecidas ou ininteligíveis, demandando um dinamismo desafiador. E lidar de forma eficaz com a incerteza requer a aplicação de abordagens flexíveis, estratégias adaptáveis e pode demandar o uso de tecnologias avançadas, refletindo a complexidade inerente à tomada de decisões em um mundo em constante evolução com uma infinidade de informações e eventos que se interrelacionam.

Soma-se a incerteza, a necessidade cada vez maior de decisões em massa, como uma resposta aparentemente inevitável à medida que a sociedade se torna mais globalizada, as informações se disseminam em quantidade e velocidades maiores e a cobrança por processos mais céleres e eficazes é feita nos mais diversos âmbitos sociais, econômicos e culturais.

Uma regra de decisão orientada por dados é uma forma de fazer uma previsão sobre algo que é atualmente desconhecido. O desenvolvimento de sistemas automatizados de tomada de decisão ou sistemas de suporte à decisão surge neste cenário como uma ferramenta valiosa com a promessa de trazer mais eficiência, precisão, escalabilidade, objetividade e consistência. A decisão algorítmica seria capaz de lidar melhor com as várias fontes de incerteza do que a decisão exclusivamente humana.

Assim, analisando os processos de tomada de decisão humano, Daniel Kahneman, Olivier Sibony e Cass R. Sunstein na obra "Ruido: Uma falha no julgamento humano"[32] identificam a convivência com um excesso de "ruídos" nos julgamentos humanos. Nas decisões humanas seriam possíveis se identificar dois fenômenos de erros: vieses e "ruídos", sendo que os vieses seriam uma inclinação sistemática nas decisões, o que resultaria em julgamentos que se desviam de forma consistente em uma direção específica. Já o "ruído" seria a variação aleatória em decisões que deveriam ser idênticas. Ou seja, o ruído seria disperso e aleatório, enquanto o viés sistemático.

[32] KAHNEMAN, Daniel; SUNSTEIN, Cass; SIBONY, Olivier. **Ruído:** Uma falha no julgamento humano. São Paulo: Objetiva, 2021.

Os mencionados autores reconhecem, assim, que quando falamos de algoritmo, estamos nos referindo a uma abordagem mecânica de decisão e que "todas as abordagens mecânicas são livres de ruído"[33], sendo que a combinação de padrões pessoais e ruído têm tamanho peso na qualidade do julgamento humano que a falta de ruído e a simplicidade são vantagens consideráveis na decisão.

E mais ainda, reconhecem que a Inteligência Artificial não "compreende", mas sim identifica padrões e que alguns algoritmos "são não só mais precisos que juízes humanos como também mais jutos"[34].

Neste sentido, considerando que a tomada de decisões na presença de incerteza é elemento central no campo da Inteligência Artificial, poderíamos pensar no processo de tomada de decisão algorítmica com base nos seguintes processamentos de incertezas: i) a incerteza de resultado, onde os efeitos das ações são incertos; ii) a incerteza de modelo, onde o modelo do problema é incerto; iii) a incerteza de estado, onde o verdadeiro estado do ambiente é incerto; e iv) a incerteza de interação, onde o comportamento de outros agentes interagindo no ambiente é incerto[35].

Para lidar com estas incertezas, surgem métodos para a tomada de decisão, tais como a "Programação Explícita", na qual o agente deve buscar antecipar os cenários nos quais ele pode se encontrar e programar explicitamente o que o agente deve fazer em resposta a cada um deles. Nessa abordagem de programação explícita pode funcionar bem para problemas simples, o que diante das diferentes incertezas mencionada acima, seria de improvável aplicação em situações complexas[36].

O "Aprendizado Supervisionado", por sua vez, seria indicado nos casos em que seria mais fácil mostrar a um agente o que fazer do que escrever um programa para que o agente siga. O designer fornece um conjunto de exemplos de treinamento, e um algoritmo de aprendizado automatizado deve generalizar a partir desses exemplos. Essa abordagem tem sido amplamente aplicada a problemas de classificação. Essa técnica pode também ser chamada de clonagem comportamental quando aplicada ao aprendizado de mapeamentos de observações para ações, e funciona bem quando um designer especializado realmente conhece a melhor ação para uma

---

[33] KAHNEMAN, Daniel; SUNSTEIN, Cass; SIBONY, Olivier. **Ruído:** Uma falha no julgamento humano. São Paulo: Objetiva, 2021. p. 124.

[34] *Ibidem*. p. 132.

[35] KOCHENDERFER, Mykel J.; WHEELER, Tim A.; WRAY, Kyle H. **Algorithms for decision making**. Cambridge: Massachusetts Institute of Technology, 2022. Disponível em: https://algorithmsbook.com/files/dm.pdf. Acesso em: 04 jan. 2024. p. 2.

[36] *Ibidem*, p. 5.

coleção representativa de situações. E geralmente eles não conseguem superar designers humanos em novas situações[37].

Outra abordagem é para o designer especificar o espaço de estratégias de decisão possíveis e uma medida de desempenho a ser maximizada, é a "Otimização", executando um conjunto de simulações. O algoritmo de otimização, então, realiza uma busca nesse espaço em busca da estratégia ótima. Aqui, o conhecimento de um modelo dinâmico não é utilizado de outra forma para orientar a busca, o que pode ser importante para problemas complexos[38].

Já o "Planejamento" é uma forma de otimização que utiliza um modelo da dinâmica do problema para ajudar a orientar a busca. Aqui uma ampla base de literatura explora vários problemas de planejamento, grande parte focada em problemas determinísticos. Assumir um modelo determinístico nos permite utilizar métodos que podem escalar mais facilmente para problemas de alta dimensionalidade. Para outros problemas, considerar a incerteza futura é crucial[39].

Por fim, o "Aprendizado por reforço" desconsidera a suposição, como no Planejamento, de que um modelo é conhecido antecipadamente. Em vez disso, a estratégia de tomada de decisões é aprendida enquanto o agente interage com o ambiente. O designer só precisa fornecer uma medida de desempenho; cabe ao algoritmo de aprendizado otimizar o comportamento do agente[40].

Contudo, de forma contraria, o uso dos algoritmos em tomadas de decisão traz novos tipos de incertezas, não pelos "ruídos", mas pelos vieses o a "perda do controle". E a imprevisibilidade é ainda maior quando estes algoritmos estão inseridos em sistemas sociotécnicos (redes que conectam as coisas com os humanos).

Eduardo Magrani[41], ao tratar deste tema traz as contribuições de Peter Kroes sobre os artefatos técnicos e os sistemas sociotécnicos. Os primeiros seriam coisas, objetos, produtos obtidos por meio de ação tecnológica, feitos pelas pessoas com uma função e um plano estabelecido de uso. Estes artefatos técnicos assim envolvem a necessidade de observar regras de uso, bem como a criação de parâmetros em relação aos papéis de indivíduos e instituições sociais em relação a eles e seu uso.

---

[37] KOCHENDERFER, Mykel J.; WHEELER, Tim A.; WRAY, Kyle H. **Algorithms for decision making**. Cambridge: Massachusetts Institute of Technology, 2022. Disponível em: https://algorithmsbook.com/files/dm.pdf. Acesso em: 04 jan. 2024. p. 6.

[38] *Ibidem*, p. 6.

[39] *Ibidem*, p. 6.

[40] *Ibidem*, p. 7.

[41] MAGRANI, Eduardo. New perspectives on ethics and the laws of artificial intelligence. **Internet Policy Review**, v. 8, n. 3. 2019. Disponível em: https://doi.org/10.14763/2019.3.1420. Acesso em: 10 set. 2023.

Os artefatos técnicos passam por uma avaliação se são bons ou não de acordo com a função e o plano de uso coincidem com o funcionamento dele. Quem cria este artefato assim, tem um parâmetro para que este não se desvie do seu propósito. E este propósito, inclusive, é inseparável das questões e decisões morais de quem o criou.

A influência da tecnologia na transformação do ambiente e na realização de objetivos individuais, sejam eles de natureza privada ou social, é evidente. E ao se considerar que os propósitos dos seres humanos ao desenvolverem artefatos técnicos estão intrinsecamente ligados às características desses objetos, torna-se claro que tais artefatos possuem uma dimensão moral inerente. Essa interseção entre os objetivos humanos e as características dos artefatos técnicos reflete um aspecto fundamental da ética e da responsabilidade na utilização da tecnologia para moldar o mundo.

Vale observar as palavras de Bruno Bioni e Maria Luciano sobre o conceito de "incerteza":

> Para além da falta de dados ou inadequação de modelos de avaliação de risco, ele também abarca a "indeterminação" (quando não se conhece todas as relações causais), a "ambiguidade" e a "ignorância" (unknow unkowns) (Science for Environment Policy, 2017). Os métodos tradicionais de regulação de risco (risk assessment, risk managment e análises de custo-benefício), que pressupõem algum conhecimento e estimativas de probabilidade na antecipação de riscos, parecem não dar conta do desconhecido.[42]

E à medida que os sistemas de inteligência artificial são aprimorados para se adaptarem e imitarem comportamentos humanos, sua previsibilidade diminui ainda mais. Eles deixam de ser apenas ferramentas com funções definidas, passando a desenvolver autonomia e comportamento próprio, o que resulta em um impacto concreto que escapa ao controle dos programadores ou usuários, já que os algoritmos podem evoluir e criar novas formas de realizar suas tarefas de maneiras imprevisíveis.

Essa crescente adaptabilidade dos algoritmos também aumenta os riscos associados às tomadas de decisão automatizadas, exigindo que os desenvolvedores estejam mais atentos às implicações éticas e legais envolvidas em suas atividades.

---

[42] BIONI, Bruno; LUCIANO, Maria. O princípio da precaução na regulação de inteligência artificial: seriam as leis de proteção de dados o seu portal de entrada. **Inteligência Artificial e Direito.** São Paulo: Thomson Reuters Brasil, p. 207-231, 2019. Disponível em: https://brunobioni.com.br/home/wp-content/uploads/2019/09/Bioni-Luciano_O-PRINCI%CC%81PIO-DA-PRECAUC%CC%A7A%CC%83O-PARA-REGULAC%CC%A7A%CC%83O-DE-INTELIGE%CC%82NCIA-ARTIFICIAL-1.pdf. Acesso em: 10 set. 2023. p. 4.

Assim, ao se olhar para os objetos mais complexos tecnologicamente que envolvem a tomada de decisão, como os sistemas sociotécnicos, a capacidade de interação e a incerteza dos resultados são ainda mais evidentes.

> Para uma análise regulatória, este conceito é ainda mais fundamental (Kroes, 2011). Precisamente por causa de sua complexidade encarnada em um conglomerado de 'atores' (em relação à concepção de teoria do ator-rede de Bruno Latour), causando sistemas sociotécnicos a terem consequências ainda menos previsíveis do que aquelas geradas por artefatos técnicos. Além disso, eles geram uma maior dificuldade para prevenir consequências não intencionais e para responsabilizar agentes em caso de dano, já que a ação tecnológica, refletida no sistema sociotécnico, é uma soma das ações dos atores, emaranhadas na rede em uma intra-relação.[43]

Fato é que independente da arquitetura, as decisões têm sido progressivamente terceirizadas para sistemas que envolvem tomada de decisão algorítmica, e mesmo em casos que a matéria a ser decidida possa envolver juízos valorativos complexos. E é neste campo que a substituição de decisões humanas por algorítmicas levanta questões profundas sobre transparência, viés algorítmico e a necessidade de explicabilidade das decisões e, neste trabalho, a possibilidade de contestação.

## 2.2 DADO PESSOAL COMO INSUMO

Correlações "objetivas" e automatizadas de dados pessoais utilizados na tomada de decisão automatizada necessitam de desconfiança a respeito dos seus efeitos, observando que "a escalada de pretensão à objetividade é precisamente, e muito concretamente, o esquecimento da escolha política".[44]

O custo decrescente da coleta, armazenamento e tratamento de dados, aliado a novas fontes de dados pessoais provenientes de sensores, câmeras e tecnologias de geolocalização, significa que a coleta de dados se tornou quase ubíqua, tendo a coleta e retenção de dados como um fenômeno praticamente permanente e a análise de dados cada vez mais realizada em velocidades próximas do tempo real.

E para falar dessa realidade, essencial é se falar em *big data*, termos que é utilizado para se referir a um conjunto de dados extremamente grande e complexo que não pode ser facilmente

---

[43] Barad, 2003 *apud* Magrani, 2019, p. 51 (tradução nossa). Disponível em: MAGRANI, Eduardo. New perspectives on ethics and the laws of artificial intelligence. **Internet Policy Review**, v. 8, n. 3. 2019. Disponível em: https://doi.org/10.14763/2019.3.1420. Acesso em: 10 set. 2023.

[44] ROUVROY, Antoniette; BERNS, Thomas. Governamentalidade algotítmica e perspecticas de emancipação*:* o díspar como condição de individuação pela relação. *In:* BRUNO, Fernanda *et al*. **Tecnopolíticas da vigilância**: perspectivas da margem. São Paulo: Boitempo, 2018. p. 113.

processado ou analisado com métodos tradicionais de processamento de dados. O *big data* é frequentemente utilizado para análises avançadas, como mineração de dados, aprendizado de máquina e análise preditiva, para obter insights valiosos, tomar decisões informadas e identificar padrões e tendências ocultas nos dados.

Esses conjuntos de dados geralmente possuem três características principais, conhecidas como os "3 V's", volume, variedade e velocidade[45].

O volume de Refere à enorme quantidade de dados gerados e acumulados continuamente a partir de diversas fontes, como dispositivos móveis, mídias sociais, sensores, transações comerciais, entre outros, a variedade à diversidade dos tipos e formatos de dados disponíveis (isso inclui dados estruturados como dados em bancos de dados tradicionais, dados semiestruturados, como arquivos XML e dados não estruturados, como texto, áudio e vídeo e, por fim, a velocidade se refere à rapidez na qual os dados são gerados e processados, podendo os dados serem gerados em tempo real, exigindo métodos de processamento e análise igualmente rápidos.

Constantiou e Kallinikos, por sua vez, identificam ainda algumas outras características, falando em volume (focando na grande quantidade de dados gerados e disponíveis na internet e nos ecossistemas de mídia digital), diversidade (para se referir a variedade de tipos de dados), velocidade (relacionando com à rapidez com que os dados crescem e são atualizados), heterogeneidade (para tratar da característica de que os dados podem ser ou não estruturados) e o foco em eventos em tempo real (o que mina as premissas padrão de tomada de decisão estratégica e desafia hábitos cognitivos e comportamentais enraizados em uma concepção linear do tempo e compromissos de longo prazo)[46].

> O Big data está se tornando um tópico destacado em pesquisas de SI (Segurança da Informação), gestão e ciências sociais. Como o próprio rótulo indica, big data é comumente utilizado para referir-se a grandes volumes de dados gerados e disponibilizados na internet e nos ecossistemas de mídia digital atual. Mas o volume de dados por si só nunca teria sido suficiente para encapsular a novidade do fenômeno (boyd e Crawford, 2012). Estreitamente associados a grandes volumes de dados estão a diversidade desses dados, a

[45] IBM Developer Blog. Wha tis big data? More than volume, velocity and Variety. 2017. Disponível em: https://developer.ibm.com/blogs/what-is-big-data-more-than-volume-velocity-and-variety. Acesso em: 10 set. 2023.

[46] CONSTANTIOU, Ioanna; KALLINIKOS, Jannis. New games, new rules: big data and the changing context of strategy. **Journal of Information Technology**, v. 30, n. 1. [S.l.] 2015. Disponível em: https://eprints.lse.ac.uk/63017/1/Kallinikos_New%20Games%20New%20Rules.pdf . Acesso em 10 out. 2023.

> frequência com que são atualizados e, mais genericamente, a velocidade com que crescem.[47]

Analisando o *big data*, Shoshana Zuboff propõe olhá-lo como um fenômeno de origem social, que se desenvolve dentro de uma lógica de acumulação de dados intencional, dentro do capitalismo de vigilância[48]. O capitalismo de vigilância teria se formado durante a última década em cima de um campo pouco desenhado e consolidado, criando uma forma de tentar prever e modificar o comportamento humano.

Apesar de reconhecer que o *big data* tem diversos âmbitos de aplicação, a autora considera que a sua origem teve como foco a produção de fórmulas para o controle de mercado por meio de um "projeto de extração fundado na indiferença formal em relação às populações que conformam tanto sua fonte de dados quanto seus alvos finais"[49].

Sessenta e sete porcento da população mundial, ou seja, cinco virgula quatro bilhões de pessoas[50] no mundo têm acesso à internet e têm as suas atividades diárias mediadas por computadores, o que cria para estas uma "nova dimensão simbólica à medida que eventos, objetos, processos e pessoas se tornam visíveis, cognoscíveis e compartilháveis de uma nova maneira"[51].

A nova lógica de acumulação, no qual o mundo "renasce" em dados,

> organiza a percepção e molda a expressão das capacidades tecnológicas em sua origem, sendo aquilo que já é tomada como dado em qualquer modelo de negócio. Suas suposições são amplamente tácitas e seu poder de moldar o campo das possibilidades é, então, amplamente invisível. Ela define objetivos, sucessos, fracassos e problemas, além de determinar o que é mensurado e o que é ignorado, o modo como recursos e pessoas são alocados e organizados, quem – e em quais funções – é valorizado, quais atividades são realizadas e

---

[47] O'Reilly, 2012; Davenport, 2014 *apud* Constantiou; Lallinikos, 2015, p. 2. (Tradução nossa). Disponível em: CONSTANTIOU, Ioanna; KALLINIKOS, Jannis. New games, new rules: big data and the changing context of strategy. **Journal of Information Technology**, v. 30, n. 1. [S.l.] 2015. Disponível em: https://eprints.lse.ac.uk/63017/1/Kallinikos_New%20Games%20New%20Rules.pdf . Acesso em 10 out 2023.

[48] ZUBOFF, Shoshana. Big Other: capitalismo de vigilância e perspectivas para uma civilização informação. *In:* BRUNO, Fernanda *et al*. **Tecnopolíticas da vigilância**: perspectivas da margem. São Paulo: Boitempo, 2018. p. 18

[49] ZUBOFF, Shoshana. Big Other: capitalismo de vigilância e perspectivas para uma civilização informação. *In:* BRUNO, Fernanda *et al*. **Tecnopolíticas da vigilância**: perspectivas da margem. São Paulo: Boitempo, 2018. p. 18

[50] Informação disponível em: https://www.telesintese.com.br/26-bilhoes-de-pessoas-ainda-nao-tem-acesso-a-internet-no-mundo-aponta-a-uit/. Acesso em: 20 mar. 2020.

[51] ZUBOFF, Shoshana. Big Other: capitalismo de vigilância e perspectivas para uma civilização informação. *In:* BRUNO, Fernanda *et al*. **Tecnopolíticas da vigilância**: perspectivas da margem. São Paulo: Boitempo, 2018. p. 24.

> com que propósitos. A lógica da acumulação produz suas próprias relações sociais e com elas suas concepções e seus usos de autoridade e poder.[52]

E para a compreensão de onde vem as informações que recriam o mundo nessa nova dimensão, essencial compreender um pouco mais esse processo de "extração", que envolve dados de transações econômicas mediadas por computadores, dados vindo de sensores incorporados em objetos, corpos e lugares, banco de dados governamentais e corporativos (com a intermediação de pagamentos eletrônicos -, o pix por exemplo, até mesmo pagamentos feitos em uma feira de rua ou uma gorjeta para um guardador de carro se tornou registrável -, os censos, planos de saúde, farmácias etc.), as câmeras de vigilância, tanto pública quanto privadas, e as cotidianidades ou *data exhaust* (os pequenos conjuntos de dados coletados das ações cotidianas, do dia a dia, como fotos, músicas e cliques).[53]

Destas diferentes pontes, provavelmente as advindas de sensores, Internet das Coisas e as *data exhaust* sejam as que trazer maiores dificuldade de transparência:

> Os novos investimentos da Google em *machine learning*, drones, dispositivos vestíveis, carros automatizados, nanopartículas que patrulham o corpo procurando sinais de doença e dispositivos inteligentes para o monitoramento do lar são componentes essenciais dessa cada vez maior rede de sensores inteligentes e dispositivos conectados à internet destinados a formar uma nova infraestrutura inteligente para corpos e objetos.[54]

No caso dos *data exhaust*, muitos dos riscos à privacidade estão no fato de que o titular do dado geralmente não enxerga qualquer valor ou potencial de risco pela presunção da "irrelevância" daquele seu dado.

> "Exaustão de dados" refere-se a dados ambientais que são coletados passivamente, dados não essenciais com valor limitado ou nulo para o parceiro original da coleta de dados. Esses dados foram coletados para um propósito diferente, mas podem ser recombinados com outras fontes de dados para criar novas fontes de valor. Quando os indivíduos adotam e usam novas tecnologias (por exemplo, telefones móveis), eles geram dados ambientais como subprodutos de suas atividades cotidianas. Os indivíduos também podem estar emitindo passivamente informações à medida que prosseguem com suas vidas diárias (por exemplo, quando fazem compras, mesmo em mercados informais; quando acessam cuidados básicos de saúde; ou quando interagem com outros). Outra fonte de exaustão de dados é o comportamento de busca de informações, que pode ser usado para inferir as necessidades, desejos ou intenções das

---

[52] ZUBOFF, Shoshana. Big Other: capitalismo de vigilância e perspectivas para uma civilização informação. *In:* BRUNO, Fernanda *et al*. **Tecnopolíticas da vigilância**: perspectivas da margem. São Paulo: Boitempo, 2018. p. 22.

[53] *Ibidem*.

[54] *Ibidem*, p. 27.

> pessoas. Isso inclui buscas na Internet, linhas telefônicas de emergência ou outros tipos de centrais de atendimento privadas.[55]
>
> Nada é trivial ou efêmero em excesso para essa colheita: as "curtidas" do Facebook, as buscas no Google, e-mails, textos, fotos, músicas e vídeos, localizações, padrões de comunicação, redes, compras, movimentos, todos os cliques, palavras com erros ortográficos, visualizações de páginas e muito mais. (...) Esses dados foram rotulados pelos tecnólogos de "data exhaust". Presumidamente, uma vez que os dados são redefinidos como resíduos, a contestação da sua extração e eventual monetização é menos provável.[56]

Outro fato importante – que sustenta a relevância do *data exhaust* - nessa lógica de acumulação de dados é a "indiferença formal"[57] do processo de extração dos dados, no qual o relevante é a quantidade e não a qualidade do dado que é extraído. Aqui há uma objetividade mecânica tanto na extração quanto na análise:

> Análises do tipo Big Data reivindicam uma forma de objetividade – não uma forma crítica de objetividade baseada no conhecimento das circunstâncias, do contexto e das causas dos fenômenos e, portanto, no reconhecimento de sua natureza contingente, mas uma objetividade mecânica, baseada por um lado na automação dos sistemas de processamento de dados e no desrespeito pela subjetividade (seletividade, pontos de vista específicos, percepção, interpretação) e, por outro, na aparente independência da modelagem algorítmica vis-à-vis categorizações politicamente instituídas ou socialmente experienciadas.[58]

Se optarmos por aprofundar ainda mais no uso dos dados pessoais por mecanismos de Inteligência Artificial e o uso dos dados pessoais por eles, a privacidade que já estava reduzida pela grande quantidade de interações com dispositivos tecnológicos acessados pelos indivíduos como celulares e computadores vem sendo minada ainda mais por dispositivos de IoT.

Na Internet das coisas, objetos cotidianos tem as suas utilidades transformadas ao serem dotados de conexão com a internet. Surgem novas funcionalidades, capacidade de comunicação e processamento de sensores, troca de informações, interações, entre outras funcionalidades que permitem a sua utilização em cidades inteligentes, casas inteligentes dentre outras aplicações.

---

[55] GEORGE, Gerard; HAAS, Martine R.; PENTLAND, Alex. Big data and management. **Academy of Management Journal**, v. 57, n. 2, p. 321-326. 2014. Disponível em: https://ink.library.smu.edu.sg/lkcsb_research/4621 Acesso em: 05 abr. 2021. (Tradução nossa).

[56] ZUBOFF, Shoshana. Big Other: capitalismo de vigilância e perspectivas para uma civilização informação. *In:* BRUNO, Fernanda *et al*. **Tecnopolíticas da vigilância**: perspectivas da margem. São Paulo: Boitempo, 2018. p. 31-32.

[57] *Ibidem*, p. 33.

[58] ROUVROY, Antoinette. **Of Data and Men**: Fundamental Rights and Freedoms in a World of Big Data. Council of europeu. Strasbourg, 2016. Disponível em: https://rm.coe.int/16806a6020. Acesso em: 03 jan. 2024. (Tradução própria).

Para a *International Telecommunication Union* (ITU), a Internet das Coisas pode ser definida como uma infraestrutura global para a sociedade da informação, permitindo serviços avançados interligando coisas (físicas e virtuais) com base em tecnologias de informação e comunicação interoperáveis existentes e em evolução[59].

No Brasil, o Decreto nº 9.854 de 2019[60], que institui o Plano Nacional de Internet das Coisas no Brasil, define ela como:

> a infraestrutura que integra a prestação de serviços de valor adicionado com capacidades de conexão física ou virtual de coisas com dispositivos baseados em tecnologias da informação e comunicação existentes e nas suas evoluções, com interoperabilidade.

As cidades inteligentes poderiam se beneficiar de sensores instalados em postes de luz, lixeiras, semáforos, câmeras de monitoramento, medidores de água ou de energia, para coletar dados sobre o tráfego, qualidade do ar, uso de determinados recursos públicos e muito mais, otimizando o planejamento urbano.

Nessa análise sobre as cidades inteligentes e o uso de dados pessoais dos cidadãos, Brian Fabrègue e Andrea Bogoni[61] sinalizam sobre os impactos negativos que os programas inteligentes que vem sendo desenvolvidos podem trazer para a privacidade e a segurança da informação. Neste sentido, os autores sinalizam a existência de uma hiper "datificação" e quantificação da vida humana por meio de informação digital e o uso destas informações por diversos atores com interesses particulares como interesses públicos, de eficiência política, de lucratividade empresarial, *marketing*, da conveniência de consumidores etc. e como a privacidade não entra nas preocupações dos designers de programas de cidades inteligentes.

E, apesar do objeto dos autores ter a Itália e a Suíça, os *insights* trazidos podem ser aplicados de forma ampla para comunidades urbanas inteligentes e digitais. Assim, os riscos de violações de privacidade existe neste cenário devido à coleta e processamento massivo de dados defendida como necessária para o funcionamento eficiência dos programas, o desafios em informar e obter eventualmente consentimento dos cidadãos em relação ao uso de novas

---

[59] INTERNATIONAL TELECOMUNICATION UNION -T. New ITU standards define the Internet of Things and provide the blueprints for its development. [S.l.]: ITU, 2012. Disponível em: https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=11559&lang=en. Acesso em: 03 jan. 2024.

[60] BRASIL. Decreto nº 9.854, de 25 de junho de 2019. Institui o Plano Nacional de Internet das Coisas e dispõe sobre a Câmara de Gestão e Acompanhamento do Desenvolvimento de Sistemas de Comunicação Máquina a Máquina e Internet das Coisas. Diário oficial da União, Brasília, 2019. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/d9854.htm.

[61] FABRÈGUE, Brian F. G.; BOGONI, Andréa. Privacy and Security Concerns in the Smart City. **Smart Cities**, v. 6, p. 586–613, 2023. Disponível em: https://doi.org/10.3390/smartcities6010027. Acesso em: 20 jun. 2024.

ferramentas de cidades inteligentes, a necessidade de garantir a conformidade legal com acordos de nível de serviço em relação à operação do sistema, dos dados gerados e o uso e compartilhamento de informações dos cidadãos, a disponibilidade ampliada de dados e ferramentas analíticas robustas, o monitoramento invasivo e até mesmo riscos relativos à liberdade individual decorrentes do monitoramento frequente do consumo de recursos (como água, energia) em ambientes inteligentes[62].

Já no âmbito mais privado e doméstico, são exemplos de IoT domésticos os lâmpadas e interruptores inteligentes, fechaduras inteligentes, câmeras de segurança conectadas, assistentes virtuais como o Amazon Echo e o Google Home, *smartwatches*, sensores de estacionamento e computador de bordo etc.

E quanto mais esses objetos vão se tornando comuns e vai se criando uma dependência neles, mais eles se tornam invisíveis. E quanto mais invisíveis, menos se questiona e menores são as preocupações com a privacidade. Provavelmente uma pessoa idosa irá se incomodar mais de trocar de roupas em um closet com uma câmera do que uma pessoa jovem que já nasceu em uma casa com circuito interno de vigilância.

Contudo nem sempre se está consciente de quais tipos de dados pessoais podem estar sendo coletados e menos ainda de todas as possíveis finalidades de uso – seja um uso autorizado ou um acesso indevido em razão de algum vazamento. Por exemplo, um *smartwatche* é capaz de coletar informações sobre a saúde de alguém, sobre quais lugares ela frequenta, sobre seus hábitos de consumo etc. E se essas informações foram parar nas mãos de lojas que querem vender determinados produtos ou serviços, nas mãos de planos de saúde ou nas mãos de um criminoso que tem como objetivo realizar um sequestro?

E quando vamos mais além do que isso e falamos de dispositivos da Internet dos Corpos de *wearables* “fixos”, incluindo neste estudo objetos que não apenas coletam dados pessoais dos seus usuários, mas estão fisicamente ligados de forma fixa ao usuário e tomam decisões sobre diversas questões, inclusive de saúde. É o caso dos dispositivos de monitoramento de glicose ou marcapassos.

## 2.3 A APLICAÇÃO DA TECNOLOGIA

Para entender como estes sistemas funcionam e os impactos que eles podem ter no objeto deste estudo, ou seja, na construção de uma Inteligência Artificial de confiança com a

[62] *Ibidem*.

contestação de decisões automatizadas, é fundamental se debruçar sobre a sua aplicação e as características das decisões tomadas por meio de sistemas de Inteligência Artificial que influenciam fortemente neste processo.

### 2.3.1 A “força” da decisão

Antoinette Rouvroy, ao falar da "força prescritiva" do resultado do processamento automático de dados por meio de algoritmos, se referindo à medida em que a tecnologia molda e prevê comportamentos que são obrigatórias, persuasivos ou incentivados observa que isso pode variar a depender de fatores relacionados com o propósito do sistema. Se é um sistema projetado para ajudar no processo de tomada de decisão, se é um sistema que tem como foco oferecer recomendações ou se é um sistema que substitui a tomada de decisão humana[63].

> Por exemplo, poder-se-ia imaginar um carro "inteligente" que "se recusasse" a iniciar até que todos os passageiros tenham afivelado seus cintos de segurança. Esse tipo de sistema é intrinsecamente prescritivo: desobedecer a ordem (de afivelar o cinto de segurança) significaria não poder usar o objeto (o carro). Outro exemplo seria a detecção automática de comportamento suspeito em aeroportos que, ao invés de "apenas" alertar a equipe de segurança, imediatamente desligaria todos os elevadores e escadas rolantes e fecharia todas as portas para áreas abertas ao público. Se a equipe em questão ignorasse o alerta em tal caso, isso seria equivalente a fechar o aeroporto completamente. Quando o procedimento é acionado, eles são, portanto, compelidos, quase fisicamente, a tomar uma ação, independentemente de qual poderia ter sido sua própria avaliação da situação.[64]

No entanto, Rouvroy sinaliza que mesmo no caso de sistemas projetados apenas para fazer recomendações e não de fato uma substituição da tomada de decisão humana, fatores relacionados com o contexto de gestão no qual o processo de tomada de decisão automatizado ocorre, combinado com fatores relacionados à psicologia dos operadores humanos podem aumentar a força de se seguir as decisões automatizadas.

Isso ocorre pois nestes casos vai caber ao operador humano acatar ou ignorar a recomendação, e o apetite ao risco, a chance de assumir alguma responsabilidade individual, as metas de produtividade ou até mesmo a inclinação daquele sujeito para obedecer ou não decisões, são fatores chaves.

---

[63] ROUVROY, Antoinette. **Of Data and Men**: Fundamental Rights and Freedoms in a World of Big Data. Council of europeu. Strasbourg, 2016. Disponível em: https://rm.coe.int/16806a6020. Acesso em: 03 jan. 2024. (Tradução nossa).

[64] *Ibidem*, p. 31.

> Portanto, é concebível que, em vários casos, o operador humano achará difícil desconsiderar a recomendação automática, porque, por um lado, isso provavelmente reduzirá seu nível de produtividade e, por outro, o obrigará a assumir responsabilidade pessoal pela decisão e suas consequências e justificá-la em caso de um resultado negativo, enquanto uma decisão de acordo com a recomendação teria permitido a ele ou ela transferir a culpa para o sistema de computador. Consequentemente, a própria existência de uma recomendação automática gera um dever, para aqueles que decidem ignorá-la, de justificar sua não conformidade não "em sua honra e consciência" ou com base em qualquer conceito que possam ter de justiça ou equidade, levando em conta as circunstâncias reais nas quais sua decisão teve que ser tomada, mas com base em argumentos que sejam pelo menos tão quantificáveis quanto as previsões algorítmicas.[65]

Isso para Rouvroy pode gerar um fenômeno de perda da habilidade de operadores humanos avaliar os casos e situações em que se encontram por conta dos sistemas automatizados de tomada de decisão que os tornam profundamente dependentes de suas ferramentas, com o risco - em caso de falha técnica - de que sejam até mesmo incapazes de tomar qualquer decisão.

Fenômeno relativamente fácil de se perceber é que os seres humanos as vezes apresentam um "viés de automação", criando uma confiança excessiva nas decisões tomadas por máquinas. E este excesso de confiança, o *overtrust* na tecnologia e nas suas decisões, pode levar a decisões e usos descuidados.

Alguns estudos chegam a falar na presença de uma "vergonha" que alguns titulares que por não compreender direito o funcionamento do algoritmo, não tendo conhecimento nem transparência sobre o mesmo, aderem de forma integral e sem reflexão ao que é sugerido, não discordam pois não se sentem aptos para tal[66].

> Este tipo de aderência manifesta um excesso de Confiança, também chamado "overtrust" ou "benevolency" (CHIEN et al., 2014; GLIKSON, 2020; HANCOCK et al., 2011; MILLER, 2015; SUTTON; HOLT; ARNOLD, 2016) e pode, entre outras coisas, levar o usuário a supor e aceitar que o modelo "sabe o que faz" (ainda que o usuário não compreenda), cabendo-lhe apenas aderir à recomendação feita.[67]

---

[65] ROUVROY, Antoinette. **Of Data and Men**: Fundamental Rights and Freedoms in a World of Big Data. Council of europeu. Strasbourg, 2016. Disponível em: https://rm.coe.int/16806a6020. Acesso em: 03 jan. 2024. p. 31-32 (Tradução nossa).

[66] SOUZA, Gustavo Henrique Costa. **Análise da relação entre a transparência da inteligência artificial e a tomada de decisões gerenciais**. 2023. Tese (Doutorado em Ciências Contábeis) – Universidade Federal de Pernambuco, Recife, 2023. Disponível em: https://repositorio.ufpe.br/bitstream/123456789/51312/1/TESE%20Gustavo%20Henrique%20Costa%20Souza.pdf. Acesso em: 15 set. 2023.

[67] *Ibidem*.

De forma aparentemente contraditória, contudo, é que nem sempre mais transparência sobre a tecnologia significa uma maior confiança.

> Não se pode desconsiderar também que, ao fornecer mais Transparência ao usuário do modelo de IA, abrem-se – para o usuário – mais possibilidades de interpretação acerca do modelo – algumas das quais podem não favorecer a Aderência (PREECE, 2018; SHRESTHA; BEN-MENAHEM; VON KROGH, 2019). Em cenário de alta Transparência, eventual desconfiança (distrust) em relação ao volume de informações que foram comunicados também não pode ser descartada: os detalhes, quando julgados excessivos, podem – de acordo com a leitura de alguns usuários – colocar o modelo de IA sob suspeição. Assim, por paradoxal que seja, mais Transparência pode, sim, significar menos Aderência. [68]

Neste sentido, Gustavo Souza, desenvolveu pesquisa examinando se a aderência dos gestores às recomendações de um modelo de inteligência artificial era afetada pela transparência do modelo e pelo tipo de decisão envolvida. As hipóteses da pesquisa eram três: i) para avaliar o impacto individual da transparência sobre a aderência; ii) para avaliar o impacto individual da decisão sobre a aderência; iii) e para examinar eventual influência conjunta de transparência e decisão sobre a aderência.

Ao testar o impacto individual da transparência sobre a aderência hipótese se verificou que de fato a transparência afeta de forma significativa a aderência à decisão da IA. Contudo, contrariando a predição, a relação é de natureza inversa – quanto maior a transparência menor a aderência.

> Como possibilidade de explicação deste resultado, pode-se aventar a ocorrência de dois fenômenos já mapeados pela literatura acadêmica: overtrust (excesso de confiança), para justificar a alta aderência em cenários de baixa transparência; e, paralelamente, distrust (desconfiança) para justificar a baixa aderência em cenário de alta transparência.[69]

Assim, a desconfiança em relação a grande quantidade de informações comunicadas ao usuário, como os detalhes julgados excessivos, pode colocar o modelo sob suspeição.

Para Paulo Schwartz os outputs dos computadores/máquinas, possuem uma "precisão sedutora". As respostas que o computador fornece seriam aceitas como completas porque as pessoas em grande parte acreditam que a máquina ofereceria uma solução entre mente e corpo:

---

[68] SOUZA, Gustavo Henrique Costa. **Análise da relação entre a transparência da inteligência artificial e a tomada de decisões gerenciais**. 2023. Tese (Doutorado em Ciências Contábeis) – Universidade Federal de Pernambuco, Recife, 2023. Disponível em: https://repositorio.ufpe.br/bitstream/123456789/51312/1/TESE%20Gustavo%20Henrique%20Costa%20Souza.pdf. Acesso em: 15 set. 2023. p. 72.

[69] *Ibidem*.

“a razão humana é limitada pela paixão humana, mas o computador supera essa limitação ao tornar a análise cognitiva disponível sem o efeito *obscurecedor* do corpóreo"[70].

> A sedutora precisão do computador resulta da diferença entre o poder que lhe atribuímos e suas capacidades e limitações reais. Essa disparidade causa uma superestimação da precisão e aplicabilidade do computador. Ela incentiva a crença de que atividades sociais díspares são capazes de uma gestão técnica pura, e desencoraja a análise das incongruências entre a gestão técnica e os objetivos reais do serviço social.[71]

O autor ainda complementa evidenciando como designs de interfaces amigáveis, como “carinhas felizes” nas telas servem para suavizar essa diferença na percepção objetiva, matemática e fria da decisão da máquina.

### 2.3.2 Formação de perfis

Uma análise dos sistemas de recomendação, amplamente utilizado no comércio eletrônico, é capaz de elucidar algumas características relevantes das tomadas de decisão algorítmica. Quando o usuário visita um site de compras, são exibidos produtos que podem ser do seu interesse, juntamente com diversos outros produtos que o usuário pesquise explicitamente. Esses produtos recomendados são frutos de uma identificação por algoritmos e podem estar relacionados às buscas atuais ou passadas do usuário, ou podem também ser itens que outras pessoas com perfil demográfico semelhante ao do usuário compraram.

Essas sugestões de recomendação, embora visem atender aos interesses dos usuários, também podem ser tendenciosas a favor de determinados produtos que as plataformas gostariam especialmente de vender. Assim, enquanto um algoritmo de recomendação pode estar fornecendo ao usuário um serviço valioso, o algoritmo também está trabalhando para promover os interesses do varejista.

Essas recomendações automáticas e orientadas por dados geralmente visam maximizar a probabilidade de que um usuário acabe fazendo uma compra ou passe mais tempo em determinada plataforma. E a estratégia para alcançar esses objetivos geralmente envolve a coleta de informações sobre a atividade passada do usuário – ou usuários “semelhantes” - na plataforma ou em plataforma semelhantes, como produtos visualizados ou pesquisados no

---

[70] SCHWARTZ, Paul. Data Processing and Government Administration: The Failure of the American Legal Response to the Computer. **Hastings Law Journal**, v. 43, n. 5, 1992. p. 1341. Disponível em: https://repository.uclawsf.edu/cgi/viewcontent.cgi?article=3086&context=hastings_law_journal. Acesso em: 10 abr. 2022. (Tradução nossa).

[71] *Ibidem* (Tradução nossa).

passado. Ademais, pode fazer parte da estratégia utilizar também informações sobre o comportamento offline do usuário.

O *profiling* pode ser definido como a formação de modelos gerais com base em dados de vários usuários/sujeitos individuais[72]. E quando este processo envolve algoritmos, como aprendizado de máquina, alguma ferramenta de inteligência artificial e a mineração de dados, estaremos diante da versão automatizada do *profiling*. De forma mais técnica, o *profiling* é uma questão de reconhecimento de padrões, que é comparável à categorização, generalização e *estereotipação*.

O *profiling*, assim, apesar dos efeitos individuais, não é uma questão particular, mas um fenômeno supraindividual. De forma bastante prática, Pedro Martins[73] traz uma referência metafórica utilizada por Martjin van Otterlo inspirada no livro de ficção "Nós", demonstrando o lado da dimensão individual da privacidade e da dimensão coletiva:

> No mundo da obra de ficção em que todas as casas são de vidro, ao andar na rua seria possível observar o dia-a-dia das pessoas em suas casas e começar a observar padrões de comportamento. Se, nesse mundo, uma pessoa decide criar uma casa de madeira para que seu comportamento não possa ser observado por terceiros, ainda assim seria razoavelmente fácil de, a partir dos padrões observados em seus vizinhos, inferir informações sobre essa pessoa.

Na era da informação, o advento da governança algorítmica redefine radicalmente as dinâmicas de poder, destacando a transição do foco individual para a complexidade das relações e do ambiente digital. Este fenômeno está intimamente relacionado com as decisões algorítmicas automatizadas, que desempenham um papel central na formação de perfis individuais.

A automação de processos decisórios, por meio de algoritmos, tornou-se uma prática comum em diversas esferas, desde recomendações de produtos até decisões mais complexas, como concessão de crédito ou avaliação de desempenho de funcionários. Esses processos

---

[72] OTTERLO, Martijn van. A machine learning view on profiling. *In:* HILDEBRANDT, Mireille; VRIES, Katja. **Privacy, Due process and the Computational Turn:** The Philosophers of Law meet Philosophers of Technology. London: Routledge, 2013. Disponível em: https://www.taylorfrancis.com/chapters/edit/10.4324/9780203427644-4/machine-learning-view-profiling-martijn-van-otterlo. Acesso em: 20 ago. 2022. p. 5.

[73] MARTINS, Pedro Bastos Lobo. **A regulação do profiling na lei geral de proteção de dados**: o livre desenvolvimento da personalidade em face da governamentalidade algorítmica. 2021. Dissertação (Mestrado em Direito) – Universidade Federal de Minas Gerais. Belo Horizonte, 2021. Disponível em: https://repositorio.ufmg.br/bitstream/1843/43900/4/Pedro%20Martins%20-%20Disserta%C3%A7%C3%A3o%20-%20A%20REGULA%C3%87%C3%83O%20DO%20PROFILING%20NA%20LEI%20GERAL%20DE%20PROTE%C3%87%C3%83O%20DE%20DADOS%20o%20livre%20desenvolvimento%20da%20personalidade%20em%20face%20da%20governamentalidade%20algor%C3%ADtmica.pdf. Acesso em: 20 ago. 2022.

automatizados têm, por sua vez, contribuído significativamente para a formação de perfis detalhados, resultando em “retratos” (mesmo que muitas vezes distorcidos) dos comportamentos e preferências individuais.

Este emprego crescente de métodos automatizados engloba também um objetivo de analisar e antecipar comportamentos humanos, o que também pode gerar riscos à autonomia humana. Essas técnicas, especialmente a de *profiling*, podem constituir ameaças estruturais ao incorporar premissas que minam a agência humana e prejudicam o processo de subjetivação em contextos de decisões automatizadas.

Henrique Parra, analisa fenômeno semelhante por meio de Antoinette Rouvroy, Fernanda Bruno e Pablo Esteban Rodríguez que lançaram a luz sobre essa transformação, evidenciando como o controle no nível do "dividual" reconfigura as fronteiras do poder, operando além das regulamentações jurídicas tradicionais[74].

Ao deslocar a disputa pelo controle para a esfera do “dividual”, a identidade civil e biológica do indivíduo perde relevância. O cerne da governança algorítmica reside na modulação existencial, na produção e gestão de dados informáticos, e na extração de valor a partir destas informações. Essas práticas ocorrem na dualidade do pré-individual e do supraindividual, onde a "relação" supera a centralidade do próprio indivíduo. O perfil torna-se a unidade de produção e controle, deslocando o poder para a capacidade de gerir o ambiente mediado digitalmente.

Nesse cenário, a subjetividade é eclipsada pelos rastros informacionais, os dados descontextualizados que adquirem uma suposta objetividade utilizada no processo de tomada de decisão algorítmico. Resultados de site de busca como o Google, sugestões em redes sociais como o TikTok, perfis de mercado e consumo e até mesmo conteúdo jornalístico atribuídos baseiam-se em uma estatística preditiva, transcendendo as abordagens tradicionais centradas na média e normalidade.

Seria o fenômeno da problematização da *superpersonalização*[75] resultante dessa dinâmica, alertando para o surgimento do efeito "bolha". Na *superpersonalização* o que se verifica são informações acessadas, as oportunidades apresentadas e a disponibilização e oferta de serviços e produtos parecem sempre se adequar ao perfil do indivíduo.

---

[74] PARRA, Henrique Zoqui Martins. Experiências com tecnoativistas: resistências na política do dividual? *In:* BRUNO, Fernanda *et al*. **Tecnopolíticas da vigilância**: perspectivas da margem. São Paulo: Boitempo, 2018.

[75] ROUVROY, Antoniette: BERNS, Thomas. Governamentalidade algotítmica e perspecticas de emancipação: o díspar como condição de individuação pela relação. *In:* BRUNO, Fernanda *et al.* **Tecnopolíticas da vigilância:** perspectivas da margem. São Paulo: Boitempo, 2018.

A customização de filtros algorítmicos cria uma esfera privada hipertrofiada, limitando o contato com visões contraditórias e diferentes. Este fenômeno, ao minar a experiência comum, reforça a polarização e a radicalização de opiniões, alimentando um ambiente em que a diversidade de perspectivas é suprimida. Assim, as decisões algorítmicas vão de forma sutil, mas não por isso menos poderosa, moldando as nossas percepções e comportamentos.

> E, o pior, desejamos cada vez mais o resultado que nos é oferecido por essas máquinas. Aquilo que percebemos como nossa liberdade de expressão online acaba por produzir todo um ambiente em que, na realidade, nem percebemos como nossas escolhas estão sendo conduzidas. Por isso a imagem do Big Brother e do panóptico não é mais suficiente ou adequada. "Não se trata mais de excluir o que sai da média, mas de evitar o imprevisível, de tal modo que cada um seja verdadeiramente si mesmo". Quando o poder informacional se desloca para a produção do ambiente e se combina com a modulação existencial, o que está em jogo é a possibilidade de produzir e gerenciar tendências. Em suma, produzir futuros.[76]

Quando o poder informacional se integra à produção do ambiente e à modulação existencial, a capacidade de gerenciar tendências e, em última análise, produzir futuros, torna-se o cerne da governança algorítmica. Com a datificação da vida dos indivíduos, a "aplicação do poder desloca-se do indivíduo para a gestão dos perfis potenciais e para a modelização dos ambientes em que a ação humana se desenvolverá"[77].

> A governança algorítmica não produz qualquer subjetivação, ela contorna e evita os sujeitos humanos reflexivos, ela se alimenta de dados "infraindividuais" insignificantes neles mesmos, para criar modelos de comportamento ou perfis supraindividuais sem jamais interpelar o sujeito, sem jamais convocá-lo a dar-se conta por si mesmo daquilo que ele é, nem daquilo que ele poderia se tornar. [...] A força bem como o perigo da generalização das práticas estatísticas à qual nós assistimos residiriam não em seu caráter individual, mas, pelo contrário, em sua autonomia ou mesmo em sua indiferença para com o indivíduo.[78]

A dimensão supraindividual das decisões automatizadas voltadas para o *profiling* introduz desafios nas regras de proteção de dados, que tradicionalmente se concentram na

---

[76] PARRA, Henrique Zoqui Martins. Experiências com tecnoativistas: resistências na política do dividual? *In:* BRUNO, Fernanda *et al*. **Tecnopolíticas da vigilância**: perspectivas da margem. São Paulo: Boitempo, 2018. p. 349.

[77] PARRA, Henrique Zoqui Martins. Experiências com tecnoativistas: resistências na política do dividual? *In:* BRUNO, Fernanda *et al*. **Tecnopolíticas da vigilância**: perspectivas da margem. São Paulo: Boitempo, 2018. p. 351.

[78] ROUVROY, Antoniette: BERNS, Thomas. Governamentalidade algotítmica e perspecticas de emancipação: o díspar como condição de individuação pela relação. *In:* BRUNO, Fernanda *et al.* **Tecnopolíticas da vigilância**: perspectivas da margem. São Paulo: Boitempo, 2018.

salvaguarda de indivíduos identificados ou identificáveis. Pedro Martins[79] ainda adiciona um ponto de tensão que surge ao considerar a utilização de dados agregados e anonimizados, uma vez que as leis de proteção de dados excluem o tratamento de dados anonimizados de sua abrangência. O *profiling* não necessariamente se concentra em indivíduos específicos e pode ser conduzido sem identificar qualquer pessoa, utilizando dados anonimizados.

> Os modelos preditivos ou perfis supraindividuais atribuídos a indivíduos são baseados em dados infraindividuais decorrentes de um grande número de indivíduos. Nesse processo, os dados de um indivíduo são tão válidos quanto os de qualquer outro – seus dados são tão bons quanto os de seu vizinho.

Hipoteticamente seria viável demandar a possibilidade de que indivíduos impactados de maneira significativa por uma decisão proveniente do processamento automático de dados expressem suas próprias perspectivas. Contudo, tudo sugere que essas perspectivas teriam uma influência limitada em comparação com os resultados objetivos e algorítmicos antecipados dos sistemas automáticos.

> Além disso, em uma fase anterior do processo, a influência que os indivíduos podem ter sobre o perfilamento é muito limitada. Os modelos preditivos ou perfis supra-individuais atribuídos aos indivíduos são baseados em dados infra-individuais derivados de um grande número de indivíduos. Neste processo, os dados de qualquer indivíduo são tão válidos quanto os de qualquer outro – seus dados são tão bons quanto os do seu vizinho – o que significa que são necessários muito poucos dados para inferir novos conhecimentos. Qualquer informação estatística relevante sobre os indivíduos já terá sido incluída no modelo bem antes de eles poderem enviar qualquer informação sobre si mesmos. O modelo literalmente esquece que indivíduos "pessoas" estão envolvidos.[80]

O perfil resultante não representa uma reprodução exata do indivíduo, mas uma tentativa de antecipar seu comportamento para propósitos específicos, derivada de uma agregação massiva de dados. Para a efetividade da decisão algorítmica, bastaria que o indivíduo se comportasse de maneira suficientemente semelhante ao restante do grupo para pode ser tratado como um membro daquele grupo.

---

[79] MARTINS, Pedro Bastos Lobo. **A regulação do profiling na lei geral de proteção de dados**: o livre desenvolvimento da personalidade em face da governamentalidade algorítmica. 2021. Dissertação (Mestrado em Direito) – Universidade Federal de Minas Gerais. Belo Horizonte, 2021. Disponível em: https://repositorio.ufmg.br/bitstream/1843/43900/4/Pedro%20Martins%20-%20Disserta%C3%A7%C3%A3o%20-%20A%20REGULA%C3%87%C3%83O%20DO%20PROFILING%20NA%20LEI%20GERAL%20DE%20PROTE%C3%87%C3%83O%20DE%20DADOS%20o%20livre%20desenvolvimento%20da%20personalidade%20em%20face%20da%20governamentalidade%20algor%C3%ADtmica.pdf. Acesso em: 20 ago. 2022.

[80] ROUVROY, Antoinette. **Of Data and Men**: Fundamental Rights and Freedoms in a World of Big Data. Council of europeu. Strasbourg, 2016. Disponível em: https://rm.coe.int/16806a6020. Acesso em: 03 jan. 2024. p. 33 (Tradução nossa).

Dessa forma, a questão da utilização de dados pessoais não se restringe apenas a informações pessoais de um único indivíduo na construção de um modelo de decisão automatizado, o que demanda uma luz também sobre a possibilidade de exercício dos direitos individuais de controle sobre um perfil e os dados agregados provenientes de diversos indivíduos, que podem até ser anonimizados, usados para formar esse perfil[81]. É a necessidade da ampliação do foco dessa luz também para a dimensão coletiva da proteção de dados pessoais.

Neste sentido, já um movimento substancial argumentando a favor da ideia de privacidade de grupo e um reconhecimento mais robusto da dimensão coletiva da proteção de dados, Rafael Zanata que propõe a abordagem de risquificação da proteção de dados pessoais como parte do processo de reformatação jurídica a partir da ampliação da tutela coletiva [82]. Tudo isso está dentro de um processo "que supera a tradicional concepção bilateral entre sujeito de direito e aquele que processa dados pessoais"[83].

O impacto coletivo do *profiling* é evidente nos efeitos discriminatórios criados e amplificados como é o caso retratado abaixo por Pedro Martins[84]:

> No contexto brasileiro, os pesquisadores Ramon Vilarino e Renato Vicente demonstram como, na fase de testes de um sistema de pontuação de crédito alimentado com uma base de dados brasileira, o uso dos primeiros 3 dígitos do CEP enquanto variável de entrada funciona como um proxy para gerar vieses raciais. Os autores sugerem, inclusive, que o uso da variável de "proporção de branquitude" equivalha ao uso do CEP. Os pesquisadores usaram os mesmos dados (idade, histórico financeiro e de crédito, hábitos de pagamento), alterando apenas os 3 dígitos do CEP, "movendo" pessoas de São Paulo, em que 36% das pessoas se autodeclaram não-brancas, para a Bahia,

---

[81] MARTINS, Pedro Bastos Lobo. **A regulação do profiling na lei geral de proteção de dados**: o livre desenvolvimento da personalidade em face da governamentalidade algorítmica. 2021. Dissertação (Mestrado em Direito) – Universidade Federal de Minas Gerais. Belo Horizonte, 2021. Disponível em: https://repositorio.ufmg.br/bitstream/1843/43900/4/Pedro%20Martins%20-%20Disserta%C3%A7%C3%A3o%20-%20A%20REGULA%C3%87%C3%83O%20DO%20PROFILING%20NA%20LEI%20GERAL%20DE%20PROTE%C3%87%C3%83O%20DE%20DADOS%20o%20livre%20desenvolvimento%20da%20personalidade%20em%20face%20da%20governamentalidade%20algor%C3%ADtmica.pdf. Acesso em: 20 ago. 2022.

[82] ZANATTA, Rafael A. F. Proteção de dados pessoais como regulação do risco: uma nova moldura teórica? *In*: **I Encontro da Rede de Pesquisa em Governança da Internet**. [on-line]. 2017. Disponível em: http://www.redegovernanca.net.br. Acesso em: 10 maio 2023.

[83] *Ibidem*, p 184.

[84] MARTINS, Pedro Bastos Lobo. **A regulação do profiling na lei geral de proteção de dados**: o livre desenvolvimento da personalidade em face da governamentalidade algorítmica. 2021. Dissertação (Mestrado em Direito) – Universidade Federal de Minas Gerais. Belo Horizonte, 2021. Disponível em: https://repositorio.ufmg.br/bitstream/1843/43900/4/Pedro%20Martins%20-%20Disserta%C3%A7%C3%A3o%20-%20A%20REGULA%C3%87%C3%83O%20DO%20PROFILING%20NA%20LEI%20GERAL%20DE%20PROTE%C3%87%C3%83O%20DE%20DADOS%20o%20livre%20desenvolvimento%20da%20personalidade%20em%20face%20da%20governamentalidade%20algor%C3%ADtmica.pdf. Acesso em: 20 ago. 2022. p 126.

> em que 78% das pessoas se autodeclaram não-brancas. Em 99,8% dos casos houve uma diminuição absoluta na pontuação de crédito da pessoa.

Importante observar que a definição do que constitui uma discriminação da decisão algorítmica no processo de *profiling* gera diversas tensões, qual a linha entre um tratamento desigual ética e juridicamente justificado e uma discriminação abusiva.

Naturalmente a função de uma decisão automatizada para a pontuação de crédito é aplicar um tratamento desigual às pessoas, separando-as entre possivelmente boas ou más pagadoras. Mas quais critérios poderiam ser utilizados para isso? Nem sempre o abuso é tão claro quanto o uso de dados raciais.

### 2.3.3 Obscuridade

Uma das características mais preocupantes das decisões tomadas por meio de mecanismos de inteligência artificial é a obscuridade. Nessa linha, Pasquale desenvolve a ideia da "*Black Box Society*" ou “Sociedade da Caixa Preta”, se referindo ao fenômeno em que algoritmos e sistemas complexos, especialmente aqueles baseados em inteligência artificial, operam muitas vezes de maneira opaca e inacessível, sem oferecer explicações claras sobre o processo de construção das suas decisões[85].

Essa característica é intimamente relacionada ao fato de que muitas vezes as pessoas afetadas por essas decisões não conseguem entender o funcionamento interno desses sistemas, o que pode impossibilitar a contestação das decisões tomadas por eles e resultar em uma sociedade em que as práticas de tomada de decisão permanecem obscuras e fora do alcance do escrutínio público.

E, ironicamente, faz parte dessa dinâmica de decisões automáticas obscuras a busca incessante por transparência e conhecimento sobre as pessoas objeto das decisões. É assim que um processo seletivo baseado em decisões feitas por Inteligência Artificial ao mesmo tempo em que se verifica uma busca pelo máximo de informações sobre a vida de um candidato, para imputar o máximo de dados no sistema, vemos, concomitantemente, camadas de complexidade que tornam incompreensível como os dados pessoais foram processados e o resultado final da seleção.  A assimetria na obscuridade é neste contexto até mesmo um ato de poder.

Sobre este poder, Frank Pasquasle dispõe:

---

[85] PASQUALE, Frank. **The Black Box Society:** The Secret Algorithms That Control Money and Information. Cambridge, Harvard University Press, 2015.

> Conhecimento é poder. Escrutinar os outros enquanto se evita o escrutínio sobre si mesmo é uma das formas mais importantes de poder. Empresas buscam detalhes íntimos da vida de potenciais clientes e empregados, mas fornecem aos reguladores o mínimo de informações possível sobre suas próprias estatísticas e procedimentos. Empresas de internet coletam cada vez mais dados sobre seus usuários, mas lutam contra regulamentações que permitiriam a esses mesmos usuários exercer algum controle sobre os dossiês digitais resultantes.[86]

A obscuridade pode se referir a diversos fatores como os critérios, as variáveis consideradas e as fontes dos dados que alimentam o algoritmo na tomada de decisão. A falta de transparência sobre estes tantos fatores tem tornado cada vez mais - à medida que o aprendizado de máquinas se expande - as decisões tecnicamente inescrutável e, portanto, difícil de compreender e contestar.

> Por exemplo, o Google sabe muito mais sobre sua população de usuários do que estes sabem sobre si mesmos. De fato, não há meios pelos quais as populações possam atravessar essa divisão dados os obstáculos materiais, intelectuais e proprietários necessários para a análise de dados e a ausência de *feedback loops*. Outra assimetria assenta no fato de que o usuário típico tem pouco ou nenhum conhecimento sobre as operações comerciais da Google, sobre a ampla gama de dados pessoais com que contribui para os servidores da Google ou sobre a retenção de dados ou, ainda, como eles são instrumentalizados e monetizados. Já é bem sabido que os usuários têm poucas opções significativas para a autogestão de privacidade. O capitalismo de vigilância prospera na ignorância do público.[87]

Soma-se a este fenômeno o fato de que muitas vezes as decisões tomadas por Inteligência Artificial se tornam tão familiares que sequer o indivíduo afetado percebe que os seus dados estão sendo utilizados e processados para tomar decisões que irão lhe afetar de forma adversa. E, eventualmente, é neste ponto de invisibilidade que os riscos florescem.

Em 1980, Mark Weiser, cientista da computação, já falava que "as tecnologias mais importantes são aquelas que desaparecem. Elas se integram à vida do dia a dia, ao nosso cotidiano e tornam-se indistinguíveis". Hoje, com o avanço da Internet das Coisas, percebemos a clareza e atualidade dessa ideia, mais de quarenta anos depois. Máquinas e robôs (sejam carros, relógios, elevadores etc.) passaram a integrar a rotina das pessoas de modo que estas sequer percebem, mas eventual retirada destas tecnologias desestruturaria a dinâmica dos seus cotidianos pessoais e profissionais.

---

[86] *Ibidem*, p 3 (Tradução nossa).

[87] ZUBOFF, Shoshana. Big Other: capitalismo de vigilância e perspectivas para uma civilização informação. *In:* BRUNO, Fernanda *et al*. **Tecnopolíticas da vigilância**: perspectivas da margem. São Paulo: Boitempo, 2018. p. 50.

O conceito de "computação ubíqua"[88] trata deste fenômeno no qual a tecnologia recua para o pano de fundo da vida das pessoas, trazendo a ideia de uma universalização da computação: "a computação ubíqua é uma integração muito difícil de fatores humanos, ciência da computação, engenharia e ciências sociais".

Aqui o computador deveria ser tão incorporado, tão apropriado, tão natural, que usamos sem nem mesmo pensar. Para Mark Weiser, "um computador com o qual eu preciso conversar, dar comandos ou ter um relacionamento (muito menos ser íntimo), é um computador que está demasiadamente no centro das atenções"[89].

E alimentando e tornando essa computação funcional, conforme tratado anteriormente, estão principalmente dados pessoais derivados de transações econômicas, de sensores incorporados aos objetivos, lugares ou corpos ou até mesmo de cotidianidades.

A obscuridade está também neste processo de extração das informações que alimentam a tomada de decisão. A extração de dados, especialmente quando feita por estas tecnologias "ubíqua", é um processo unidirecional, que contrasta com a natureza recíproca dos relacionamentos. Ao contrário de interações baseadas em trocas mútuas, a extração de dados aqui é conduzida de maneira unilateral, frequentemente sem a participação ativa ou o consentimento informado dos indivíduos envolvidos[90].

Os processos extrativos necessários para viabilizar o funcionamento destas decisões geralmente ocorrem à margem do diálogo significativo ou da obtenção de consentimento explícito. Isso é particularmente preocupante, pois esses procedimentos não apenas capturam fatos objetivos, mas também abrangem subjetividades ligadas às vidas individuais. A ausência dessa via de mão dupla destaca ainda mais a assimetria de poder, onde a coleta de dados ocorre à revelia da vontade e conhecimento das pessoas, muitas vezes resultando em uma desconexão entre aqueles que fornecem os dados e as entidades que os coletam.

Na mencionada lógica do capitalismo de vigilância:

> as receitas dependem de ativos de dados apropriados por meio de ubíquas operações automatizadas. Essas operações constituem uma nova classe de ativos: os ativos de vigilância. Os críticos do capitalismo de vigilância podem caracterizar tais ativos como "bens roubados" ou "contrabando" na medida

---

[88] WEISER, Mark. The computer for the 21st Century. Scientific American. p. 94-10. 1991 Disponível em: https://ics.uci.edu/~corps/phaseii/Weiser-Computer21stCentury-SciAm.pdf. Acesso em: 10 set. 2023 (Tradução nossa).

[89] WEISER, Mark. The Invisible Interface: Increasing the Power of the Environment through Calm Technology. Conference paper, 1998. Disponível em: https://link.springer.com/chapter/10.1007/3-540-69706-3_1. Acesso em: 10 set. 2023. (Tradução nossa).

[90] ZUBOFF, Shoshana. Big Other: capitalismo de vigilância e perspectivas para uma civilização informação. *In:* BRUNO, Fernanda *et al*. **Tecnopolíticas da vigilância**: perspectivas da margem. São Paulo: Boitempo, 2018.

> em que foram tomados, em vez de fornecidos, e não produzem, como argumentarei a seguir, as devidas reciprocidades.[91]

Essa falta de reciprocidade levanta questões éticas significativas sobre privacidade, a autonomia dos sujeitos objeto das decisões e o controle sobre informações pessoais. A reflexão sobre essas práticas extrativas é crucial para promover discussões informadas, regulamentações eficazes e especialmente o desenvolvimento de tecnologias que respeitem os direitos individuais, mitigando as preocupações relacionadas à exploração não autorizada de dados.

Neste sentido, Frank Pasquasle observa que até mesmo situações "banais" como direcionamento de conteúdos de entretenimento, como acontece em plataformas de vídeo como o Netflix e o YouTube podem ter agendas ocultas de direcionamento que impactam na nossa autonomia de forma aleatória ao nosso conhecimento:

> De forma mais benigna, talvez, essas empresas influenciam as escolhas que faríamos por nós mesmos. Motores de recomendação na Amazon e no YouTube afetam uma familiaridade automatizada, sugerindo delicadamente ofertas que acham que vamos gostar. Mas não subestime a importância desse "talvez". As agendas econômicas, políticas e culturais por trás de suas sugestões são difíceis de desvendar. Como intermediários, eles se especializam em mudar alianças, às vezes avançando os interesses dos clientes, às vezes dos fornecedores: tudo para orquestrar um mundo online que maximize seus próprios lucros.[92]

Por outro lado, conforme observou Danilo Doneda e Virgílio Almeida[93], ao mesmo tempo em que vai aumentando a consciência sobre os riscos da obscuridade e opacidade das decisões, principalmente sobre direitos e garantias fundamentais, existem justificativas técnicas e não técnicas para o não uso de algoritmos abertos ou ausência de transparência. São, por exemplo, argumentos de proteção à concorrência, à propriedade intelectual ou até mesmo evitar que as pessoas tentem "enganar" o algoritmo.

Danielle Keats Citron e Frank Pasquale, ao tratar dos impactos que a prática de score de crédito social traz e dos questionamentos sobre a justiça neste sistema, sinalizam que a falta de transparência nos sistemas de pontuação de crédito levanta preocupações significativas, pois os indivíduos não conseguem entender, desafiar ou auditar completamente os processos por trás de suas pontuações de crédito. Além disso, as agências de crédito frequentemente negam

---

[91] ZUBOFF, Shoshana. Big Other: capitalismo de vigilância e perspectivas para uma civilização informação. *In:* BRUNO, Fernanda *et al*. **Tecnopolíticas da vigilância**: perspectivas da margem. São Paulo: Boitempo, 2018. p. 40.

[92] PASQUALE, Frank. **The Black Box Society:** The Secret Algorithms That Control Money and Information. Cambridge, Harvard University Press, 2015. p 5 (Tradução nossa).

[93] DONEDA, Danilo; ALMEIDA, Virgílio A. F. O que é governança de algoritmos. *In:* BRUNO, Fernanda *et al.* **Tecnopolíticas da vigilância:** perspectivas da margem. São Paulo: Boitempo, 2018.

pedidos de detalhes sobre seus sistemas de pontuação, impedindo que os indivíduos e os reguladores conduzam auditorias dos algoritmos preditivos subjacentes.

Sabe-se que o comportamento individual impacta mas o indivíduo não sabe como.

> A falta de transparência dos sistemas de pontuação de crédito deixa os consumidores confusos sobre como e por que suas pontuações mudam. FICO e as agências de crédito não explicam a extensão na qual o comportamento individual afeta certas categorias. Os consumidores não podem determinar o comportamento de crédito ótimo ou mesmo o que fazer para evitar uma queda em suas pontuações.
>
> No entanto, FICO e as agências de crédito anunciam o peso relativo de certas categorias em seus sistemas de pontuação. Por exemplo, pode ser utilizado o "uso do crédito" (quanto das linhas de crédito atuais de um tomador estão sendo usadas). Mas a estratégia ótima de utilização de crédito é incerta. Ninguém sabe se, por exemplo, usar vinte e cinco por cento do limite de crédito é melhor ou pior do que usar quinze por cento. Um consumidor ambicioso poderia tentar reverter a engenharia das pontuações de crédito, mas tais esforços seriam caros e pouco confiáveis.[94]

Assim, a opacidade torna desafiador para os consumidores determinar o comportamento de crédito ideal ou como evitar impactos negativos em suas pontuações.

### 2.3.4 Vieses algorítmicos discriminatórios

O recorte que se faz aqui é o dos vieses algorítmicos discriminatórios e não dos vieses algorítmicos em geral. O viés algorítmico de forma genérica se refere a uma propriedade do algoritmo de inteligência artificial em si e, para os cientistas de dados, "o viés, juntamente com a variância, descreve uma propriedade de algoritmo que influencia o desempenho da previsão."[95]

> O viés e a variação são interdependentes e os cientistas de dados geralmente buscam um equilíbrio entre os dois.
>
> Modelos com alta variação tendem a se flexionar para encaixar os dados de treinamento. Eles podem acomodar mais facilmente a complexidade, mas também são mais sensíveis ao ruído e podem não ser bem generalizados para dados fora do conjunto de treinamento.

---

[94] CITRON, Danielle Keats; PASQUALE, Frank A. The Scored Society: Due Process for Automated Predictions. **Washington Law Review**, n. 1, 2014, p. 2-27. Disponível em: https://digitalcommons.law.umaryland.edu/fac_pubs/1431/. Acesso em: 04 fev. 2024. p. 11 (Tradução nossa).

[95] GOMES, Pedro César Tebaldi. Ética e Inteligência Artificial: viés em machine learning. Data Geeks. 2019. Disponível em: https://www.datageeks.com.br/etica-e-inteligencia-artificial/. Acesso em: 10 jan. 2024.

> Modelos com alto viés são rígidos. Eles são menos sensíveis a variações nos dados e podem perder complexidades subjacentes. Ao mesmo tempo, eles são mais resistentes ao ruído.
>
> Encontrar o equilíbrio apropriado entre essas duas propriedades para um determinado modelo em um determinado ambiente é um conjunto de habilidades críticas para a ciência de dados.
>
> Reduzir erros de previsão no aprendizado de máquina através do trade-off de viés é uma etapa bem compreendida pelos bons profissionais, mas ainda podem ocorrer falhas. Portanto, é preciso atenção para evitar o viés de algoritmo. [96]

Já quando falamos em vieses algorítmicos discriminatórios, o foco da luz passa a ser os resultados discriminatórios, sejam eles programados para tal ou não. E a decisão automatizada, especificamente, pode simultaneamente sistematizar e ocultar a discriminação.

Principalmente em razão da dificuldade de prever os efeitos de uma regra complexa e derivada de algoritmo antecipadamente, reguladores e os próprios sujeitos impactados podem ser incapazes de perceber os efeitos discriminatórios de alguma regra decisória.

Contudo, antes de adentrar nos vieses discriminatórios em decisões automatizadas, é importante que seja aberto um parêntese para a constatação de que esses vieses não são uma particularidade das decisões automatizadas tomadas por sistemas de Inteligência Artificial, muito pelo contrário, muitas vezes eles são exatamente o reflexo de vieses presentes em decisões humanas. E neste sentido, entender o viés cognitivo é necessário.

Conforme trazido em pesquisa sobre Inteligência artificial, vieses cognitivos e decisões judiciais, Maurício Requião[97] explica que os vieses cognitivos são distorções sistemáticas no processo de tomada de decisão humana, decorrentes do uso inadequado de atalhos mentais, conhecidos como heurísticas. Esses atalhos são estratégias cognitivas que facilitam o processamento rápido de informações, especialmente em contextos de alta complexidade ou sob pressão temporal. No entanto, quando aplicadas sem a devida reflexão crítica, essas heurísticas podem levar a julgamentos tendenciosos e conclusões incorretas.

A ocorrência de vieses cognitivos nas decisões humanas deve-se a várias razões. Em primeiro lugar, há uma limitação inerente à capacidade cognitiva humana, que não permite o processamento exaustivo de todas as informações relevantes de maneira consciente e racional. Isso leva ao uso de heurísticas como uma forma de simplificação cognitiva. Em segundo lugar, fatores subjetivos como experiências passadas, crenças pessoais e emoções desempenham um

---

[96] *Ibidem*.

[97] REQUIÃO, Maurício. (no prelo). Inteligência artificial, vieses cognitivos e decisões judiciais.

papel significativo na forma como as informações são interpretadas e avaliadas, muitas vezes sem que o indivíduo esteja ciente dessa influência. Por fim, o autor sinaliza que até mesmo a pressão por decisões rápidas, como no contexto do judiciário, exacerba a dependência de processos automáticos de pensamento, que são altamente suscetíveis a vieses.

Assim, sendo o viés uma característica do próprio pensamento humano, a Inteligência Artificial poderia ser uma ferramenta utilizada para diminuir a ocorrência de vieses cognitivos ao conseguir analisar um volume muito maior de dados e até mesmo um controle sobre quais dados alimentam aquela decisão. Contudo, apesar de teoricamente preciso esse racional, na prática, o que se verifica nas decisões analisadas é diferente.

Uma das características teorizadas das decisões automatizadas é a retirada dos elementos subjetivos de decisão e o consequente resultado de decisões mais objetivas, imparciais e neutras, conforme observam Marco Aurélio Marrafon e Filipe Medon[98]:

> Com o mencionado avanço da tecnologia, mais decisões essenciais sobre a vida de uma pessoa vem sendo tomadas automaticamente a partir de algoritmos comandados por Inteligência Artificial (IA). Em tese, esses algoritmos são programados para produzir um resultado melhor, a partir de técnicas como o machine learning e o deep learning.
>
> O machine learning ou aprendizagem de máquina, "faz com que a máquina aprenda certas funções a ponto de conseguir agir sem a interferência humana.". Isto é, a máquina aprende com base em suas experiências pretéritas, podendo chegar, por isso, a resultados sequer previsíveis pelos seus programadores.
>
> Com o crescente desenvolvimento, chegou-se à subespécie do deep learning, ou aprendizagem profunda, que envolve a criação de redes neurais artificiais que permitem dotar a máquina de estruturas similares ao cérebro humano. Ainda que baseada em uma racionalidade formal e probabilística, a máquina seria capaz de realizar análises cada diz mais complexas e aprender com a experiência.

Essas decisões podem, assim, desfrutar de uma suposição (irreal) de imparcialidade ou objetividade. Contudo, contradizendo essa crença na prática, o *design* de sistemas de decisão automatizadas podem apresentar determinadas falhas e tendências, com a identificação de vieses algorítmicos:

> Entretanto, o que tem se visto é que, em verdade, a neutralidade é aparente: as máquinas herdam o conteúdo a que possuem contato, seja por carregamento

[98] MARRAFON, Marco Aurélio, MEDON, Filipe. Importância da revisão humana das decisões automatizadas na Lei Geral de Proteção de Dados. Consultor Jurídico. 2019. Disponível em: https://www.conjur.com.br/2019-set-09/constituicao-poder-importancia-revisao-humana-decisoes-automatizadas-lgpd. Acesso em: 10 set. 2023.

> inicial de programadores, seja por aprendizado na interação humana, inclusive o preconceito.
>
> Sem capacidade de análise crítica, elas podem ainda aprender por si própria, nas técnicas mais avançadas de machine learning. Isso porque os dados desempenham o papel de combustível para a decisão dos algoritmos que, com base nas técnicas descritas acima, vão gerando novos conhecimentos, numa inteligência formal própria da IA.
>
> Sendo assim, como esperar a neutralidade de um algoritmo, se o banco de dados que o alimenta for enviesado?[99]

Neste sentido, Maurício Requião e Diego Carneiro[100] sinalizam a falta de substância no argumento de que os sistemas de decisão algorítmica trariam maior objetividade e neutralidade e que:

> tem-se verificado que a IA e o processo algorítmico não apenas são incapazes de corrigir o erro subjetivo humano, como também podem replicar e até reforçar os preconceitos existentes sociedade, ocasionando distinções, preferências ou exclusões capazes de afetar a igualdade de tratamento entre os indivíduos, sobretudo os grupos vulneráveis. É o que se chama de viés discriminatório do algoritmo ou, simplesmente, discriminação algorítmica.[101]

Cathy O'Neil observa que muitas das aplicações matemáticas que fomentam a economia dos dados são baseadas em escolhas feitas por humanos e muitas vezes humanos com as melhores das intenções, contudo o resultado acabava sendo preconceitos, equívocos e vieses nos sistemas de software.

> Como deuses, esses modelos matemáticos opacos, seus mecanismos invisíveis a todos exceto os altos sacerdotes de seus domínios: os matemáticos e os cientistas da computação. Suas decisões, mesmo quando erradas ou danosas, estavam para além de qualquer contestação. E elas tendiam a punir os pobres e oprimidos da sociedade enquanto enriqueciam ainda mais os ricos.[102]

Vieses algorítmicos se referem a distorções sistemáticas que podem surgir nos resultados produzidos por algoritmos devido a vários fatores, como a forma como os dados são coletados, processados ou utilizados para treinar o algoritmo. Esses vieses podem assim resultar

---

[99] *Ibidem*.

[100] REQUIÃO, Maurício; COSTA, Diego Carneiro. Discriminação algorítmica: ações afirmativas como estratégia de combate. **Civilistica.com**, Rio de Janeiro, v. 11, n. 3, p. 1-24, 2022. Disponível em: https://civilistica.emnuvens.com.br/redc/article/view/804. Acesso em: 07 jul. 2024.

[101] REQUIÃO, Maurício; COSTA, Diego Carneiro. Discriminação algorítmica: ações afirmativas como estratégia de combate. **Civilistica.com**, Rio de Janeiro, v. 11, n. 3, p. 1-24, 2022. Disponível em: https://civilistica.emnuvens.com.br/redc/article/view/804. Acesso em: 07 jul. 2024. p. 4.

[102] O'NEIL, Cathy. **Algoritmos de destruição em massa**: como o big data aumenta a desigualdade e ameaça a democracia. 1 ed. São Paulo: Editora Rua do Sabão, 2020. p 8.

em decisões injustas ou desiguais, criando, perpetuando ou ampliando desigualdades existentes na sociedade.

Estes vieses podem ser vieses i) de seleção de dados, o que ocorre quando os conjuntos de dados usados para treinar os algoritmos não são representativos da população ou do fenômeno que estão tentando modelar; ii) da programação, quando o próprio algoritmo exibe padrões discriminatórios devido à maneira como foi projetado ou aos dados usados para treiná-lo; iii) de implementação, quando há falhas na implementação prática do algoritmo, como erros de codificação ou configuração incorreta, que resultam em resultados injustos ou imprecisos; iv) de retroalimentação, quando os resultados produzidos pelo algoritmo são usados para tomar decisões subsequentes, criando um ciclo de retroalimentação que reforça e amplifica quaisquer vieses presentes nos dados ou no próprio algoritmo.

> Ainda assim, mesmo quando, ou talvez particularmente quando, são *autoaprendizáveis*, os algoritmos incorporam certas "visões de mundo", incluindo aquelas que toleram discriminação, e também permitem diferenças de tratamento no campo do emprego a serem realizadas de acordo com fatores e critérios altamente opacos (mesmo aos olhos daqueles que usam algoritmos para fins de seleção) ou dentro deles mesmos, individualmente, por razões não relacionadas às exigências do cargo ou do trabalho.
>
> No entanto, a impenetrabilidade dos processos algorítmicos e o fato de estarem cobertos pelo segredo industrial tornam qualquer discriminação muito difícil de provar, especialmente porque, na maioria das vezes, a intenção de discriminar não reside de forma alguma nas pessoas que simplesmente usam esses sistemas automáticos para tornar suas próprias decisões mais objetivas. A discriminação indireta resultante da operação de um sistema automatizado de recomendação decorre não tanto da pessoa que decide seguir a recomendação (na verdade, pode-se dizer que a disposição dessa pessoa em tornar suas decisões mais objetivas reflete um desejo de anular seus próprios preconceitos) quanto da existência prévia na sociedade de uma mentalidade discriminatória (um "apetite" por ou aceitação da discriminação) variando em escopo, mas refletida passivamente em conjuntos de dados e, portanto, adquirindo o status de um fato objetivo, apolítico, neutro e não problemático.[103]

Assim, exemplos históricos que refletem preconceitos ou viés implícito ou dados estatísticos distorcidos ao alimentarem algoritmos, especialmente os que incluem algum tipo de aprendizado de máquina, podem levar a resultados discriminatórios[104]. Da mesma forma,

---

[103] ROUVROY, Antoinette. **Of Data and Men**: Fundamental Rights and Freedoms in a World of Big Data. Council of europeu. Strasbourg, 2016. Disponível em: https://rm.coe.int/16806a6020. Acesso em: 03 jan. 2024. p. 33 (Tradução nossa).

[104] Dados de treinamento contaminados seriam um problema, por exemplo, se um programa para selecionar entre candidatos a emprego fosse treinado com base nas decisões de contratação anteriores feitas por humanos, e essas decisões anteriores fossem tendenciosas. Distorção estatística, mesmo sem maldade, pode produzir efeitos igualmente problemáticos: considere, por exemplo, um algoritmo que instrui a polícia a parar e revistar

modelos de aprendizado de máquina podem incorporar discriminação por meio de escolhas na forma como os modelos são construídos, a seleção de características por exemplo (escolhas sobre quais dados os modelos devem considerar podem inclusive revelar uma visão de mundo do programador).

Contudo, é possível que exista uma discriminação intencional, sendo essa disfarçada como uma das formas de discriminação não intencional. Um programador poderia distorcer os dados de treinamento ou escolher proxies específicas para gerar resultados discriminatórios.[105]

Na obra "Algoritmos de destruição em massa", Cathy O'Neil traz diversos exemplos de como o uso de dados pessoais e o *big data* trazem ricos e aumento de desigualdade. Para isso a autora cria o conceito de "Armas de Destruição Matemáticas", ou "ADMs". As ADMs, contudo, apesar de serem modelos nocivos, não necessariamente foram desenhadas para produzir este resultado.

É o caso do *software* de previsão de crimes feito de uma startup de big data da California, a PredPol, que utilizava dados de histórico criminal e calculava onde seria mais provável que os crimes acontecessem. A lógica do software era possibilitar uma melhor alocação do corpo policial, e foi com esse intuito que a polícia de Reading, após a redução do seu quadro, resolveu começar a utilizar essa tecnologia em 2013. A depender das áreas indicadas no mapa pelo software, era lá que a polícia ia para inibir os criminosos. O resultado esperado era que os policiais fossem posicionados nos locais mais prováveis de os crimes acontecerem.

Essa tecnologia era baseada em um software sísmico: "ele vê um crime numa área, o incorpora em padrões de histórico, e faz a previsão de onde e quando pode ocorrer novamente"[106]. De forma prévia, inclusive em relação ao uso de dados pessoais e impacto aos direitos individuais, o PredPol parecia razoavelmente justo: não tem como foco o indivíduo, e

---

pedestres. Se este algoritmo foi treinado em um conjunto de dados que sobrerrepresenta a incidência de crime entre certos grupos (porque esses grupos historicamente foram o alvo de fiscalização desproporcional), o algoritmo pode direcionar a polícia a deter membros desses grupos a uma taxa desproporcionalmente alta (e não membros a uma taxa desproporcionalmente baixa). Esse foi o caso com o programa de parar e revistar do Departamento de Polícia de Nova York, para o qual dados de 2004 a 2012 mostraram que 83% das paradas foram de pessoas negras ou hispânicas e 10% foram de pessoas brancas em uma população residente que era 52% negra ou hispânica e 33% branca. Observe que a sobrerrepresentação de pessoas negras e hispânicas nesta amostra pode levar um algoritmo a associar características tipicamente negras ou hispânicas com paradas que levam à prevenção do crime, simplesmente porque essas características estão sobrerrepresentadas na população que foi parada (Kroll *et al*., 2017, p. 681-681). KROLL, Joshua A. *et al.* Accountable Algorithms. Forthcoming, **Fordham Law Legal Studies Research Paper**, v. 165, n. 2765268. University of Pennsylvania Law Review, 2017. Disponível em: https://ssrn.com/abstract=2765268. Acesso em: 05 fev. 2024. p. 681-681 (Tradução nossa).

[105] *Ibidem*.

[106] O'NEIL, Cathy. **Algoritmos de destruição em massa**: como o big data aumenta a desigualdade e ameaça a democracia. 1 ed. São Paulo: Editora Rua do Sabão, 2020. p. 135.

sim a localização geográfica, os principais inputs eram o tipo e o local de cada crime e o resultado final acabaria beneficiando as áreas indicadas (aumentando o policiamento delas).

Mas a realidade foi diversa uma vez que a polícia tinha a opção de, ao configurar o software, se concentrar exclusivamente nos chamados crimes Parte 1 (crimes violentos, incluindo os de homicídio, a agressão ou incêndios criminosos) ou também incluir os crimes Parte 2, que englobam alguns crimes de perturbação que dificilmente seriam registrados se não houvesse um polícia lá presente (como vadiagem, mendicância mais agressiva e consumo e venda de pequenas quantidades de drogas).

> Esses crimes de perturbação são endêmicos em muitos bairros empobrecidos. (...) Infelizmente, incluí-los no modelo ameaça distorcer a análise. Uma vez que os dados de perturbação flutuam para dentro de um modelo de previsão, mais policiais são atraídos para aqueles bairros, onde é mais provável que prendam mais pessoas. (...) É da natureza do patrulhamento.[107]

O *feedback* assim é fator de viés e cria um ciclo. Os crimes Parte 2 inundam os inputs dados pela polícia no sistema e o resultado é a polícia sempre voltando para os mesmos bairros. E com isso reflexos de ordem socioeconômicas e raciais acabam surgindo:

> E nossos presídios se enchem de centenas de milhares de pessoas condenadas por crimes sem vítima. A maioria delas vem de bairros empobrecidos, e a maioria é negra ou hispânica. Então mesmo que o modelo não enxergue a cor da pele, o resultado o faz. Em nossas cidades amplamente segregadas, a localização geográfica é um proxy altamente eficaz de raça.
>
> (...)
>
> Mas e os crimes removidos dos mapas do PredPol, aqueles cometidos pelos ricos?[108]

No Brasil o resultado do PredPol não seria diferente, e mesmo sem este software, é percebido que a polícia faz escolhas sobre concentrar a atenção nas regiões mais pobres. E com os cientistas de dados estão criando e perpetuando esses status quo social em seus modelos, "o resultado é que criminalizamos a pobreza acreditando o tempo todo que nossas ferramentas não são apenas científicas, mas justas"[109].

O *European Parliamentary Research Service* no trabalho "*A Governance framework for algorithmic accountability and transparency*"[110], traz a reflexão de que quando se tem

---

[107] *Ibidem*, p. 137.
[108] *Ibidem*, p. 137.
[109] *Ibidem*, p. 144.
[110] EUROPEAN PARLIMENT RESEARCH SERVICE. A governance framework for algorithmic accountability and transparency. Scientific Foresight Unit, 2019. Disponível em:

múltiplas fontes de discriminação, são necessárias igualmente múltiplas soluções, o que cria um desafio ético para os reguladores. Aqui entram demandas por maior transparência no processo de desenvolvimento do algoritmo, por diversidade entre os desenvolvedores, de construir banco de dados de treinamento não tendenciosos e mais inclusivos com diferentes grupos demográficos, na acomodação de eventuais vieses já conhecidos como raciais ou estereótipos de gênero mesmo que por meio de distorções, entre outros. Oura abordagem seria:

> Regulamentar a opacidade dos algoritmos, que é principalmente estabelecida por meio de acordos de confidencialidade, estabelecendo agências públicas de supervisão e até mesmo incentivando a auto-regulação setorial (como visto na indústria de publicidade). Um bom exemplo é o Projeto de Lei sobre Responsabilidade Algorítmica de Nova York. Os termos atuais da proposta buscam estabelecer uma agência responsável pela justiça, responsabilidade e transparência dos algoritmos que são usados pelas autoridades públicas. Os cidadãos podem solicitar ação da agência para buscar explicação e eventualmente contestar decisões guiadas por algoritmos por essas autoridades. A agência também seria responsável por policiar práticas discriminatórias dentro dos sistemas de decisão algorítmicos e fornecer informações sobre como um algoritmo funciona e impacta a cidade.[111]

Os desafios aqui são muitos uma vez que, como dito, a transparência nem sempre é suficiente para determinar se estamos diante de um sistema algorítmico são tendenciosos e nem sempre há uma linha reta dos valores de design para os resultados. Diferença de valores entre as partes interessadas podem também impactar, "a eficiência pragmática de uma pessoa pode ser o racismo tecnocrático de outra, e muitos resultados tendenciosos ou discriminatórios das tecnologias de decisão algorítmica começaram com boas ou neutras intenções"[112].

No âmbito do uso de decisões automatizadas que poderiam auxiliar magistrados a julgarem, Maurício Requião evidencia que as decisões por meio de sistemas de Inteligência Artificial poderiam aumentar vieses de adesão e ancoragem.

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

---

https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf. Acesso em: 10 out. 2023.

111 EUROPEAN PARLIMENT RESEARCH SERVICE. A governance framework for algorithmic accountability and transparency. Scientific Foresight Unit, 2019. Disponível em: https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf. Acesso em: 10 out. 2023. (Tradução nossa).

112 *Ibidem* (Tradução nossa).

> Já o viés de ancoragem "ocorre quando o indivíduo é exposto a uma informação ou experiência previamente à decisão que servirá de base (ou âncora) a seu raciocínio ao considerar estimativas e realizar julgamentos" (TABAK, AMARAL, 2018. p. 478). Assim, se uma pessoa é exposta a um número qualquer, antes de ter que fazer certa estimativa para uma quantidade desconhecida, ela tenderá a realizá-la de modo a aproximá-la deste número (KAHNEMAN, 2012, p.152-153).[113]

Assim, os sistemas de Inteligência Artificial, dentro do contexto da atividade-fim dos magistrados, auxiliando no próprio ato de julgar, poderiam ter como consequência tornar os magistrados meros repetidores das sugestões do sistema. Essa análise parece guardar estrita relação com o que identificado na pesquisa sobre a "força" da decisão, analisado no item 2.3.1 neste trabalho.

Ao estudar o uso de decisões automatizadas no setor financeiro, Frank Pasquale observou que:

> Assim como os setores de reputação e busca, a indústria financeira caracterizou cada vez mais decisões como procedimentos computáveis e programáveis. Big data possibilita técnicas complexas de reconhecimento de padrões para analisar conjuntos de dados massivos. Métodos algorítmicos de reduzir o julgamento a uma série de etapas deveriam racionalizar as finanças, substituindo intermediários tendenciosos ou autocentrados por estruturas de decisão sólidas. E eles de fato reduziram algumas ineficiências. Mas também acabaram solidamente incorporando alguns padrões duvidosos antigos de castas de crédito e irresponsabilidade corporativa. As caixas-pretas das finanças substituíram problemas antigos familiares por um triplo golpe de complexidade técnica, segredo real e leis de segredo comercial.[114]

Ou seja, mesma áreas nas quais era comum se verificar vieses indesejados nas tomadas de decisão humana, a substituição por decisões automatizadas pode ter significado não a retirada dos vieses com decisões mais justas, mas apenas decisões igualmente enviesadas mas com uma camada de complexidade adicional de constatação e contestação.

Importante considerar que muitos sistemas algorítmicos de tomada de decisão foram criados exatamente para superar vieses nas tomadas de decisão por humanos, como é o exemplo de sistemas utilizados para recrutamento e seleção que expressam que eles servem para evitar discriminação e vieses do recrutador nos processos de recrutamento da empresa. Mas avaliar se

---

[113] REQUIÃO, Maurício. (no prelo). Inteligência artificial, vieses cognitivos e decisões judiciais. p. 8-9.

[114] PASQUALE, Frank. **The Black Box Society:** The Secret Algorithms That Control Money and Information. Cambridge, Harvard University Press, 2015. p. 15.

a decisão tomada pelo algoritmo é melhor do que a que seria tomada pelo humano é difícil de se verificar. Bastaria que eles fossem apenas "menos injusto" ou "menos discriminatórios"?

Solução interessante para essa questão é trazida no mencionado artigo de Maurício Requião e Diego Carneiro, uma solução para os sistemas de tomada de decisão algorítmica que não está no próprio algoritmo, mas sim em ações afirmativas:

> É por isso que a proposta de solução ventilada neste trabalho se encontra fora do algoritmo, visto que, para garantir que os sistemas algoritmos não discriminem, não será suficiente apenas o controle de resultados indesejáveis, mesmo quando isso seja possível. É preciso avançar em ações afirmativas para que as empresas diversifiquem as suas equipes, possibilitando um controle interno mais rigoroso por meio dos próprios funcionários membros de grupos minoritários, a fim de que mesmo os dados coletados em rede, que necessariamente refletem uma sociedade injusta, tenham cada vez menos impacto no resultado das decisões automatizadas.[115]

O racional seria o de que a existência de uma maior participação de populações mais vulneráveis nas equipes de desenvolvimento ou entre os responsáveis pela tomada de decisão nas empresas poderia ser uma forma de diminuir os vieses uma vez que "pessoas que vivenciam as dificuldades e discriminação passadas pelos grupos vulneráveis, terão maior capacidade de desenvolver e aprovar o uso de inteligência artificial que seja menos prejudicial a estes"[116].

Essa solução proposta é ao mesmo tempo simples e inovadora, uma vez que encara o viés algorítmico como parte de um problema prévio ao desenvolvimento do algoritmo, presente na própria sociedade e amplia as possibilidades de busca por soluções. Abordagens como estas, inclusive, trazem para a mesa a possibilidade de contribuições mais plurais e multidisciplinares de profissionais de outras áreas.

---

[115] REQUIÃO, Maurício; COSTA, Diego Carneiro. Discriminação algorítmica: ações afirmativas como estratégia de combate. **Civilistica.com**, Rio de Janeiro, v. 11, n. 3, p. 1-24, 2022. Disponível em: https://civilistica.emnuvens.com.br/redc/article/view/804. Acesso em: 07 jul. 2024.

[116] *Ibidem*, p. 19.

# 3 TRUSTWORTHY AI

Em março de 2023 um grupo de pesquisadores da área da IA, incluindo nomes notáveis como o historiador Yuval Noah Harari e Elon Musk, fundador da SpaceX, assinaram uma carta pública que foi publicada pelo *Future of Life Institute* – FLI[117], pedindo uma pausa de seis meses no desenvolvimento dos sistemas de IA.

Na carta são sinalizados como os sistemas de Inteligência Artificial vêm sendo desenvolvidos sem planejamento de gestão, os riscos que estão sendo criados, principalmente em sistemas que nem mesmos os desenvolvedores conseguem entender, ter previsibilidade ou um controle confiável. A sugestão foi que essa pausa de seis meses servisse para os especialistas e laboratórios de Inteligência Artificial implementarem protocolos de segurança, desenvolvam IAs auditáveis e recuem "da perigosa corrida para modelos de caixa-preta imprevisíveis cada vez maiores com capacidades emergentes"[118].

O foco da pesquisa durante essa "pausa" concerne ao desenvolvimento de sistemas de Inteligência Artificial mais precisos, seguros, interpretáveis, transparentes, resilientes, alinhados, confiáveis e leais:

> Sistemas de IA contemporâneos agora estão se tornando competitivos com relação aos humanos em tarefas gerais, e devemos nos perguntar: devemos deixar as máquinas inundar nossos canais de informação com propaganda e inverdades? Devemos automatizar todos os empregos, inclusive os gratificantes? Devemos desenvolver mentes não humanas que poderiam acabar nos substituindo, nos tornando obsoletos e nos superando em números e inteligência? Devemos arriscar a perda do controle da nossa civilização? Tais decisões não devem ser delegadas a líderes tecnológicos não eleitos. Sistemas de IA poderosos devem ser desenvolvidos somente quando estivermos confiantes que seus efeitos serão positivos e seus riscos serão geríveis.[119]

O que se verificou após a publicação dessa carta, contudo, é que na prática este apelo não parece ter sido maior do que os objetivos dos chamados laboratórios de Inteligência Artificial. Independentemente das respostas acima, os canais de comunicação continuam sendo inundados com informações inverídicas, permanecemos, na medida da tecnologia existente,

---

[117] O Future of Life Institute – FLI é uma das principais organizações no mundo focadas na diminuição de riscos catastróficos globais e existenciais. Ela tem como missão direcionar a tecnologia transformadora para beneficiar a vida e afastar riscos extremos em grande escala. Future of Life Institute. 2021. Disponível em: https://futureoflife.org/. Acesso em: 04 jan. 2024.

[118] Pause os Experimentos de IA Gigantes: uma Carta Aberta do Future of Life Institute 2023. Disponível em: https://80000horas.com.br/pause-os-experimentos-de-ia-gigantes-uma-carta-aberta-do-future-of-life-institute/.

[119] *Ibidem*.

buscando automatizar os empregos e a substituição humana e arriscando a mencionada perda de controle. E tudo isso sim, feito por "líderes tecnológicos não eleitos".

Contudo, apesar da ausência da "pausa", em paralelo, estudos com esses objetivos vêm sendo desenvolvidos, como é o caso dos realizados pelo Grupo Independente de Peritos de Alto Nível Sobre a Inteligência Artificial (GPAN), que vem há anos discutindo o tema e propondo orientações éticas para uma Inteligência Artificial de confiança, ou a *Trustworthy* IA[120].

Em 2018 este grupo sinalizava que havia uma importante janela de oportunidade para moldar o desenvolvimento da Inteligência Artificial e que o foco deveria ser garantir que os sistemas sociotécnicos[121] em que a Inteligência Artificial eram incorporadas eram confiáveis. Anos depois, a janela parece ainda existir, mas certamente de forma mais estreita.

De forma mais cautelosa, estes estudos apontam que para aumentarmos a dependência e delegarmos cada vez mais decisões aos sistemas de Inteligência Artificial, é essencial garantir que eles impactem a vida das pessoas de maneira minimamente justa, que estejam alinhados com valores fundamentais e que possam agir de acordo com esses valores. Além disso, é crucial que se estabelecem processos de responsabilização adequados para garantir isso, o que envolvem sistemas que sejam robustos em termos técnicos e sociais, confiáveis e operando de forma transparente e "contestável".

## 3.1 COMPONENTES PARA UMA IA DE CONFIANÇA

O Grupo Independente de Peritos de Alto Nível Sobre a Inteligência Artificial (GPAN) apresenta como sugestão que um sistema de Inteligência Artificial para ser considerado de confiança deve possuir três componentes necessários: i) deve ser Legal, o que envolve garantir o respeito de toda a legislação e regulamentação aplicáveis a ele; ii) deve ser Ético, devendo observar assim não apenas a lei mas também princípios e valores éticos; e iii) deve ser Sólido, tanto tecnicamente como do ponto de vista social.

O sistema de Inteligência Artificial será considerado Legal se cumprir todas as legislações e as regulamentações aplicáveis a ele. Contudo, atualmente ainda existem poucas

---

[120] RELATÓRIO DA COMISSÃO EUROPEIA. Orientações éticas para uma IA de confiança. European Commission. 2019. Disponível em: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. Acesso em: 13 jan. 2024.

[121] Sistemas sociotécnicos incluem seres humanos, intervenientes estatais, empresas, infraestruturas, software, protocolos, normas, governação, legislação em vigor, mecanismos de supervisão, estruturas de incentivo, procedimentos de auditoria, comunicação de melhores práticas e outros elementos. Disponível em: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. Acesso em: 13 jan. 2024.

legislações no mundo desenvolvidas com o objetivo de disciplinar legalmente o uso da Inteligência Artificial, o que traz pouca segurança jurídica e expectativa a respeitos da (i)legalidade de muitos sistemas de Inteligência Artificial. Adicionalmente, não é suficiente para se ter uma Inteligência Artificial confiável a simples análise de subsunção dela às legislações de determinado local (se houver). A legislação dificilmente acompanha a rapidez da evolução tecnológica, ou é capaz de abranger de forma completa as diferentes situações. A legislação pode, inclusive, estar afastada de normas éticas – o processo legislativo nem sempre funciona com este guia.

Ademais, a própria "tropicalização" de regulamentações e legislações aplicáveis a esses sistemas, como foi o caso da lei 13.709/18, a Lei Geral de Proteção de Dados, que importou toda a lógica da regulamentação de uso de dados pessoais do Regulamento Geral sobre a Proteção de Dados (RGPD) europeu, é outro ingrediente que pode trazer uma insuficiência dos textos legislativos.

O componente Ético, assim, é capaz de trazer reflexões ainda mais amplas que o Legal, até mesmo porque nem sempre o que está previsto na legislação pode ser a opção mais ética. Compreender a forma como o planejamento, o desenvolvimento, a implantação e o uso dos sistemas de Inteligência Artificial ocorre é fundamental para a compreensão dos riscos e da potência destas tecnologias do ponto de vista ético, tanto para os indivíduos impactados quanto para a sociedade como um todo.

Assim, falar em ética da Inteligência Artificial é falar em ética aplicada às questões de como os sistemas de Inteligência Artificial impactam a vida das pessoas, melhoram ou pioram a qualidade de vida individual e social tendo como referência princípios éticos e valores humanos. Isso inclui questões de privacidade, de princípios democráticos, promoção da justiça e equidade, a promoção do bem-estar coletivo, garantias de transparência, entre outros aspectos.

Nesse sentido, o Código de Ética da Associação de Máquinas de Computação[122] separa a necessidade legal da ética, indicando que independente de regulamentação legal existente, os profissionais devem realizar avaliações dos sistemas de computador e os seus impactos incluindo os riscos éticos. É estabelecido, nesse sentido, inclusive um peso maior para a avaliação ética do que a moral:

> Profissionais da computação devem seguir estas regras, a menos que haja uma justificativa ética convincente para agir de outra forma. Regras que são consideradas antiéticas devem ser contestadas. Uma regra pode ser antiética

[122] Código de Ética da Associação de Máquinas de Computação. Association for Computing Machinery. 2018. Disponível em: https://www.acm.org/code-of-ethics. Acesso em: 04 jan. 2024.

> quando tem uma base moral inadequada ou causa danos reconhecíveis. Um profissional da computação deve considerar contestar a regra por meio dos canais existentes antes de violá-la. Um profissional da computação que decide violar uma regra porque ela é antiética, ou por qualquer outro motivo, deve considerar as potenciais consequências e aceitar responsabilidade por essa ação.[123]

É importante desde já esclarecer que as pesquisas na área da ética aplicada a Inteligência Artificial não parecem que devem ter como objetivo a construção de um "código de ética" universal e aplicável à Inteligência Artificial que resolverá todos os dilemas. O objetivo principal deve ser a construção de uma mentalidade ética, o que sempre demandará raciocínios éticos, debates políticos, educação e um olhar sensível à realidade daquele contexto de aplicação. Não se defende neste trabalho, assim, que os guias éticos para um sistema de Inteligência Artificial no Brasil devem ser construídos seguindo exatamente os mesmos guias que um sistema de Inteligência Artificial que vá ser aplicado a países economicamente mais desenvolvidos e sem a mesma realidade social, racial, cultural e econômica que o Brasil.

Por fim, o último componente, a Solidez, envolve um duplo fator, a solidez técnica e a social. Para se ter uma Inteligência Artificial de confiança, os destinatários e os impactados por ela deveriam poder confiar que os sistemas não causarão danos não intencionais.

> Tais sistemas devem funcionar de forma segura e fiável, e devem prever-se salvaguardas para evitar impactos negativos não intencionais. Por conseguinte, é importante garantir a solidez dos sistemas de IA, tanto do ponto de vista técnico (assegurando a solidez técnica do sistema exigida em determinado contexto, tal como o domínio de aplicação ou a fase do ciclo de vida), como do ponto de vista social (tendo devidamente em conta o contexto e o ambiente em que o sistema opera).[124]

A solidez técnica em um sistema de Inteligência Artificial de tomada de decisão, assim, refere-se à capacidade da tecnologia de tomar decisões precisas e confiáveis com base em algoritmos robustos e bem desenvolvidos. A solidez social, por sua vez, envolve as considerações de seus impactos na sociedade, incluindo questões éticas, justiça social e equidade. Uma Inteligência Artificial sólida socialmente é construída levando em consideração os valores culturais e direitos humanos, garantindo que as suas decisões não discriminem ou prejudiquem grupos específicos.

---

[123] Código de Ética da Associação de Máquinas de Computação. Association for Computing Machinery. 2018. Disponível em: https://www.acm.org/code-of-ethics. Acesso em: 04 jan. 2024. (Tradução nossa).

[124] RELATÓRIO DA COMISSÃO EUROPEIA. Orientações éticas para uma IA de confiança. European Commission. 2019. Disponível em: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. Acesso em: 13 jan. 2024.

Este último componente, é o que mais pontos possui de relação com o objeto deste trabalho. Uma Inteligência Artificial tecnicamente sólida é menos propensa a cometer erros graves ou a tomar decisões injustas, o que reduz a necessidade de contestação das decisões tomadas por ela - uma relação "preventiva". A transparência, a responsabilidade e a justiça no uso desses sistemas não só são essenciais para a concretização do direito de contestar uma decisão tomada por uma IA, como também são essenciais para garantir a solidez do sistema.

Conforme se estudará de forma mais detalhada ao longo do próximo capítulo, mesmo em sistemas altamente sólidos, o direito de contestação é fundamental como salvaguarda contra possíveis falhas.

## 3.2 PRINCÍPIOS ÉTICOS

Diversos são os estudos que atualmente refletem sobre quais princípios éticos deveriam ser aplicados aos sistemas de Inteligência Artificial e às decisões algorítmicas tomadas por eles. Neste trabalho, com o objetivo de condensar e sistematizar os princípios, foi feita uma curadoria dos princípios que foram compreendidos como os que possuíam uma relação mais íntima e necessária com o objeto deste estudo, quais sejam: i) Ser humano como centro e Respeito a autonomia humana; ii) Prevenção de Danos; iii) Equidade; iv) Transparência e Explicabilidade; v) Prestação de contas e *Accountability*. Vale se ter atenção também aos grupos vulneráveis (crianças, idosos, pessoas com deficiência, grupos historicamente menos favorecidos ou em risco de exclusão ou situações de assimetria de poder ou informação conforme sinalizado quanto aos riscos das decisões).

### 3.2.1 Ser humano como centro e Respeito a autonomia humana

Este princípio envolve principalmente os subprincípios do valores centrados no ser humano e equidade[125], desenvolvido pela OECD, e do respeito à autonomia humana, trabalhado pelo GPAN[126].

Este princípio envolve a diretriz de que ao longo de todo o ciclo de vida de um sistema de Inteligência Artificial, desde a concepção até os resultados e impactos finais, a tecnologia

[125] OECD LEGAL INSTRUMENTS. RECOMMENDATION of the Council on Artificial Intelligence. 2019. Disponível em https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Acesso em: 13 jan. 2024.

[126] RELATÓRIO DA COMISSÃO EUROPEIA. Orientações éticas para uma IA de confiança. European Commission. 2019. Disponível em: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. Acesso em: 13 jan. 2024.

deve ter como guia o Estado de direito, os direitos humanos e valores democráticos. Isso inclui a liberdade, a dignidade e a autonomia humana, as questões relativas à privacidade e proteção de dados, a não discriminação e igualdade, a diversidade, equidade, e até mesmo temas como a justiça social e os direitos trabalhistas.

A autonomia humana aparece aqui intimamente relacionada com a ideia da autodeterminação, mas não apenas a autodeterminação informativa, mas autodeterminação das suas vidas, do espaço para a escolha humana:

> Os sistemas de IA não devem subordinar, coagir, enganar, manipular, condicionar ou arregimentar injustificadamente os seres humanos. Em vez disso, os sistemas de IA devem ser concebidos para aumentar, complementar e capacitar as competências cognitivas, sociais e culturais dos seres humanas.[127]

Para a concretização deste princípio, é essencial que no processo decisório da Inteligência Artificial exista espaço para a supervisão e o controle por parte dos seres humanos.

Pode-se dividir a supervisão de sistemas de Inteligência Artificial em três diferentes abordagens, *a Human In The Loop* (HITL) (intervenção humana em todos os ciclos de decisão do sistema, ou seja, é a capacidade do humano interferir no procedimento em cada processo de decisão), a *Human On The Loop (*HOTL) (intervenção humana durante o ciclo de design do sistema e monitoramento da operação do sistema, pela visualização no sistema) e a *Human In Command* (HIC) (supervisão humana na atividade geral do sistema de IA, na qual o humano decide quando e como usá-lo).

Na HITL, automação da decisão não significa ausência do envolvimento humano, mas a presença seletiva da sua participação humana em algumas fases:

> O resultado seria um processo que aproveita a eficiência da automação inteligente, permanecendo passível de feedback humano, mantendo ao mesmo tempo um maior senso de significado. A abordagem human-in-the-loop objetiva reformular a automação dos entes inteligentes artificialmente ao estabelecer uma relação Human-Computer Interaction (HCI), para descobrir como podemos incorporar a interação humana de forma mais útil, segura e significativa no sistema de IA.[128]

---

[127][125] RELATÓRIO DA COMISSÃO EUROPEIA. Orientações éticas para uma IA de confiança. European Commission. 2019. Disponível em: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. Acesso em: 13 jan. 2024.

[128] DIVINO, Sthefano Bruno Santos; MAGALHÃES, Rodrigo Almeida. Inteligência Artificial e Direito Empresarial: Mecanismos de Governança Digital para Implementação e Confiabilidade. **Revista dos Tribunais Online**, [s.l.], v. 1021, p. 191-212, 2020. Disponível em: https://www.thomsonreuters.com.br/content/dam/ewp-m/documents/brazil/pt/pdf/other/rt-1021-inteligencia-artificial-e-direito-empresarial-mecanismos-de-governanca-digital-para-implementacao.pdf. Acesso em: 04 fev. 2024.

Aqui é possível se ter fluxos de trabalho em que os algoritmos da Inteligência Artificial aprendem com o operador humano e como, em contrapartida, o trabalho humano se torna mais eficiente ao mesmo tempo em que traz conhecimento tácito para o sistema. A máquina ao executar uma ação, solicita informações a um especialista humano e aprende com a resposta que recebe[129].

Um exemplo simples e cotidiano de HITL em um sistema de Inteligência Artificial são os de detecção de *spam* de e-mail. Nesses sistemas, algoritmos de Inteligência Artificial são utilizados para identificar e filtrar mensagens de e-mail indesejadas, classificando-as como *spam* ou não *spam*. No entanto, mesmo com algoritmos avançados, pode ocorrer que mensagens legítimas sejam erroneamente identificadas como *spam* ou que algumas mensagens de *spam* passem pelo filtro. Assim, um sistema HITL garante maior precisão na classificação das mensagens.

Quando uma mensagem é marcada como *spam* pelo algoritmo de Inteligência Artificial, ela não é automaticamente descartada e, em vez disso, é enviada para revisão por um humano, o próprio usuário do e-mail, que pode examinar a mensagem e decidir se ela realmente é *spam* ou não. Se for considerada uma falsa detecção de *spam*, o operador pode corrigir a classificação e, assim, melhorar o desempenho do algoritmo de Inteligência Artificial no futuro.

No *Human On The Loop* (HOTL), por sua vez, como concepção, não se verifica a intervenção do operador humano na rodagem do sistema, a capacidade decisória do humano, assim, diferente nessas duas abordagens. Na HOTL o humano é menos um decisor, e mais um supervisor.

Dessa forma, a interferência humana se daria nas hipóteses em que a Inteligência Artificial seria incapaz de solucionar a situação problema apresentada ou quando os riscos da atividade delegada para a Inteligência Artificial são tão grandes que se tomadas de forma totalmente automatizadas poderiam causar um prejuízo alto demais.

São exemplos de aplicação deste tipo de abordagem o controle de tráfego aéreo e no campo da medicina, áreas que envolvem uma intervenção de forma seletiva e estratégica no funcionamento do sistema, diminuindo certos riscos para garantir que ele opere de forma eficaz, eficiente e alinhada com os objetivos do usuário, mesmo sem uma intervenção constante ou em tempo real.

---

[129] WANG, Ge. **Humans in the Loop:** The Design of Interactive AI Systems. Human-centered Artificial Intelligence. Stanford University. 2019. Disponível em: https://hai.stanford.edu/news/humans-loop-design-interactive-ai-systems Acesso em: 04 fev. 2024.

Mary "Missy" Cummings, acadêmica e engenheira espacial, ao tratar do conceito do HOTL identificou que o tema da supervisão humana envolve uma abordagem interdisciplinar, incluindo:

> a psicologia da tomada de decisão humana, que é crítica em sistemas de alto risco e sob pressão de tempo, e muitas vezes o fator limitante no sucesso do sistema, 1) ciência da computação, especificamente o design de algoritmos (e a automação resultante), bem como as interfaces que se comunicam com o operador (incluindo auditivas, visuais e táteis), e 2) a engenharia do sistema que executa a tarefa (por exemplo, é importante que um designer de um cockpit de UAV entenda como latências em sistemas de comunicação podem impactar negativamente o entendimento humano e a execução de um mecanismo de controle).[130]

Esse seria o caso dos sistemas de diagnóstico assistido por inteligência artificial, que são utilizados para analisar dados de saúde, como imagens de radiografia, ressonância magnética ou tomografia computadorizada, para auxiliar os médicos no diagnóstico de doenças ou condições médicas. A abordagem HOTL entra em cena quando os médicos utilizam esses sistemas para revisar e validar os diagnósticos sugeridos pela Inteligência Artificial.

Assim o médico pode revisar a análise do sistema de Inteligência Artificial e pode concordar com o diagnóstico sugerido, mas também pode decidir examinar ou solicitar exames adicionais para confirmar o diagnóstico. Nesse cenário, a Inteligência Artificial atua como uma ferramenta de apoio ao diagnóstico, fornecendo informações úteis que podem ajudar os médicos a tomarem decisões mais informadas. No entanto, os médicos permanecem no *loop*, supervisionando, avaliando e interpretando as informações fornecidas pela Inteligência Artificial.

Por fim, tem-se o *Human In Command* (HIC), que evidencia o humano com autoridade direta e primária sobre o controle e a operação de um sistema de IA, com o poder de decisão final sobre as ações do sistema. Isso significa que, embora a Inteligência Artificial possa fornecer sugestões ou assistência, os humanos ainda têm autoridade para revisar, modificar ou anular as decisões tomadas.

A abordagem que tem como objetivo garantir que o humano permaneça no controle pode ajudar a evitar erros graves, reduzir o risco de comportamento inadequado da Inteligência Artificial e garantir que as decisões finais estejam alinhadas com as normas éticas e sociais.

---

[130] CUMMINGS, Mary. **Supervising automation**: humans on the loop. Aero-Astro Magazine Highlight: MIT Department of Aeronautics and Astronautics. 2008. Disponível em: http://web.mit.edu/aeroastro/news/magazine/aeroastro5/cummings.html. Acesso em: 02 fev. 2024. (Tradução nossa).

Tudo isso deve ser levado em conta quando se pensa, inclusive, em reflexos de responsabilidade jurídica.

O *European Economic and Social Committee* (EESS) propõe essa abordagem e, assim, Catelijne Muller esclarece que:

> Humanos podem e devem também estar no comando de se, quando e como a IA é usada em nossas vidas cotidianas - quais tarefas transferimos para a IA, quão transparente ela é, se ela será um agente ético. Afinal, cabe a nós decidir se queremos que certos trabalhos sejam realizados, cuidados sejam prestados ou decisões médicas sejam tomadas por IA, e se queremos aceitar IA que possa comprometer nossa segurança, privacidade ou autonomia.[131]

Contudo, é importante observar que o desenvolvimento de sistemas de tomada de decisão automatizadas tem sido também construído com base na abordagem do *Human out of the Loop*, no qual o humano não faz parte do processo de tomada de decisão do sistema. Uma dessas aplicações seriam os *Human out of the Loop Weapons*, no qual robôs seriam capazes de selecionar e direcionar uma arma sem qualquer interação humana.

Nesse sentido, Danielle Keats Citron e Frank Pasquale observam que em alguns casos, como o desse exemplo, esse tipo de sistema já poderia estar violando normas já existentes, como as de direito internacional aplicáveis aos casos de guerra em razão da impossibilidade de sistemas de decisão completamente automatizada distinguirem combatentes de não combatentes e de aplicarem valores humanos como o da proporcionalidade. Haveria nessas situações a necessidade de um julgamento holístico e não algorítmico para supervisionar situações difíceis e bastante complexas como essas.

Outro relevante exemplo de identificação de risco de retirada do controle humano é o campo de aplicação dos Neurodireitos, que, na classificação de Rafael Yuste, Jared Genser, e Stephanie Herrmann[132], incluem direitos como o direito à privacidade mental, o direito à identidade e à autonomia pessoal e o direito à proteção contra a discriminação algorítmica. Seria possível, assim, com base no que já se tem de tecnologia desenvolvida, cogitar de processos de tomada de decisão automatizada com a falsa percepção de controle humano.

---

[131] EUROPEAN ECONOMIC AND SOCIAL COMMITTEE. Artificial Intelligence: Europe needs to take a human-in-command approach, says EESC. Press Release, n. 27, 2017. Disponível em: https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fwww.eesc.europa.eu%2Fsites%2Fdefault%2Ffiles%2Fresources%2Fdocs%2Fcp-27-artificial-intelligence.docx&wdOrigin=BROWSELINK. Acesso em: 04 jan. 2024. (Tradução nossa).

[132] YUSTE, R.; GENSER, J.; HERRMANN, S. It's time for neuro--rights: new human rights for the age of neurotechnology. Horizons J. Int. Relat. Sustain. Dev. v. 18, p. 154–164, 2021. Disponível em: https://www.cirsd.org/en/horizons/horizons-winter-2021-issue-no-18/its-time-for-neuro--rights. Acesso em: 20 jul. 2024.

A preservação da autonomia humana decorreria do foco em proteger as pessoas contra influências externas que possam manipular ou controlar seus pensamentos e comportamentos e a proteção da autonomia cognitiva ao garantir que as tecnologias não sejam usadas para influenciar ou controlar os processos mentais sem o consentimento da pessoa, reforçando a capacidade de tomada de decisão autônoma. Buscaria se garantir com os Neurodireitos que as decisões de um indivíduo sejam verdadeiramente suas.

Ademais, a capacidade de tomar decisões autênticas está intrinsecamente ligada à manutenção da identidade e da integridade mentais, a preservação da liberdade cognitiva e a capacidade de escolha, elementos centrais da autonomia humana.

Hipoteticamente, poderíamos cogitar que tecnologias como a “tiara neural” das escolas da China[133] poderiam ser utilizadas não apenas para "monitorar e melhorar" a concentração e desempenho dos alunos, mas também manipular o estado emocional e mental. Por exemplo, a Inteligência Artificial pode sugerir uma pausa para relaxamento com base em padrões que não visam realmente o bem-estar do aluno, mas sim aumentar a produtividade de uma maneira que beneficie exclusivamente os interesses terceiros. Além disso, o sistema pode fazer com que os alunos sintam que estão optando por sugestões de maneira autônoma, enquanto na verdade as escolhas são projetadas para induzir estados mentais específicos que maximizam a concentração ou o aprendizado.

### 3.2.2 Prevenção de Danos

Por tudo quanto exposto, é certo que o desenvolvimento de algoritmos para a tomada de decisão que impacta os sujeitos pode gerar riscos à integridade ou direitos das pessoas, o que aliás é natural de diversas atividades, sendo muitas delas ligadas à necessidades essenciais, como a de transportes e saúde, por exemplo. Não cabe, portanto, utilizar tal conclusão como argumento suficiente para inibir ou até mesmo desestimular a exploração dessas tecnologias, o que inclusive iria de encontro ao princípio da livre iniciativa econômica, garantido na Constituição Federal brasileira. O que se deve buscar, é uma conciliação entre a liberdade de iniciativa empresarial e a proteção da integridade da pessoa humana.

---

[133] VALOR ECONÔMICO. *Somos experimenta ferramenta que mede atenção do aluno*. Valor Econômico, 1 out. 2020. Disponível em: https://valor.globo.com/empresas/noticia/2020/10/01/somos-experimenta-ferramenta-que-mede-atencao-do-aluno.ghtml. Acesso em: 12 ago. 2024.

A vulnerabilidade dos indivíduos diante da capacidade crescente dos algoritmos de tomarem decisões autonomamente e que impactam a sua liberdade, dignidade e suas oportunidades é real, colocando-as em risco de sofrer danos pessoais relevantes. Trata-se, contudo, de danos que muitas vezes não são "visíveis", não são imediatos, ou até mesmo não são compreendidos, dependendo eventualmente da sua verificação do reconhecimento da tutela integral da pessoa humana[134], compreendida em toda a sua complexidade de relações sociais, econômicas, afetivas, pessoais, dentre outras e de condições e características individuais relacionadas à classe social, raça, etnia, gênero, orientação sexual, idade dentre outros aspectos.

Aqui, ao se falar em prevenção de danos, é dada luz aos sujeitos de direito. O sentido de direito, portanto, perpassa pela apreensão do conteúdo da dignidade da pessoa humana[135], núcleo irradiador de todos os demais direitos[136] e parâmetro interpretativo do teor destes, na medida em que consagra o indivíduo como centro das preocupações dos juristas.

Nesse sentido, conforme trazido anteriormente, os algoritmos, a partir do desenvolvimento da inteligência artificial, podem causar prejuízos à dignidade de pessoas humanas, seja privando-as da liberdade de escolha com a aplicação de um filtro sobre a realidade fática a partir do qual essa poderá ser conhecida, sujeitando-as a tratamento discriminatório, ou algo que as coloca em posição de vulnerabilidade. Diante das consequências advindas destes danos, falar em prevenção, em ações para evitar a concretização dos dados é uma abordagem valiosa.

Bruno Bioni e Maria Luciano[137], tratam do princípio da precaução, traçando paralelos entre a aplicação desse princípio na esfera ambiental, com a possibilidade de abordagem pela

---

134 O autor se refere à "obrigatoriedade da proteção máxima à pessoa", com a garantia de "respeito absoluto ao indivíduo", de modo a lhe proporcionar "uma existência plenamente digna e protegida de qualquer espécie de ofensa [...]". ALMEIDA NETO, Amaro Alves de. Dano existencial e tutela da dignidade da pessoa humana. **Revista de Direito Privado**, [s.l.], v. 6, n. 24, p. 21-53, out./dez. 2005.

135 É esse o caso do ordenamento jurídico brasileiro, que tem como ponto de partida (e prioridade) a tutela integral da pessoa humana, fundada no respeito à sua dignidade e na garantia do livre e pleno desenvolvimento da sua personalidade, os quais pressupõem um inequívoco reconhecimento da titularidade de um patrimônio imaterial. Inclusive, a Constituição Federal, no seu art. 1º, consagra a dignidade da pessoa humana como princípio fundamental a ser observado, elevando-a ao patamar de macroprincípio, que alicerça materialmente os direitos fundamentais, representando um poderoso instrumento de defesa dos cidadãos contra arbitrariedades cometidas pelo Estado ou por particulares.

136 Ingo Wolfgang Sarlet, ao tratar da matéria, sustenta que os direitos fundamentais prescritos no ordenamento jurídico brasileiro (seu raciocínio, nesse particular, é perfeitamente adaptável a outras experiências normativas) correspondem, em alguma medida, a explicitações do princípio da dignidade da pessoa humana. SARLET, Ingo Wolfgang. **A eficácia dos direitos fundamentais:** uma teoria geral dos direitos fundamentais na perspectiva constitucional. 6 ed. Porto Alegre: Livraria do Advogado, 2006, p. 129.

137 BIONI, Bruno; LUCIANO, Maria. O princípio da precaução na regulação de inteligência artificial: seriam as leis de proteção de dados o seu portal de entrada. **Inteligência Artificial e Direito.** São Paulo: Thomson Reuters Brasil, p. 207-231, 2019. Disponível em: https://brunobioni.com.br/home/wp-content/uploads/2019/09/Bioni-Luciano_O-PRINCI%CC%81PIO-DA-PRECAUC%CC%A7A%CC%83O-

análise de riscos no âmbito de sistemas de inteligência artificial. Um ponto condutor de tais aplicações é a dificuldade de obtenção de evidências científicas quanto aos dados possíveis.

Pela lógica da precaução seria possível se desenhar graus de força de aplicação do princípio da precaução como: i) fraca: a incerteza não justificaria uma inação; ii) moderada: a incerteza na avaliação do risco justifica ação; iii) forte: quando houver ameaça de dano, medidas de precaução devem ser tomadas; diante da incerteza, inverte-se o ônus da prova.

Isso levaria a decisões sobre a natureza das ações de precaução, por exemplo. Nos casos de "força fraca", haveria um pressuposto de gerenciamento de risco, no de "força moderada, o pressuposto de gerenciamento de risco seria implícito, sendo indicadas medidas sujeitas à revisão quando novas informações ou evidências científicas surgirem, enquanto nos casos de "força forte", o pressuposto seria o de evitar o risco.[138]

Para o Grupo Independente de Peritos de Alto Nível Sobre a Inteligência Artificial – GPAN, pelo princípio da prevenção de danos, os sistemas de Inteligência Artificial não devem afetar de forma negativa os seres humanos, não devendo causar ou agravar qualquer dano.

Os sistemas de inteligência artificial e os ambientes nos quais eles operam devem ser seguros e protegidos com foco preventivo e para isso é essencial que eles sejam tecnicamente robustos e que sejam tomadas medidas para garantir que não sejam explorados para propósitos maliciosos.

Ainda para o GPAN, é importante dedicar uma atenção especial às pessoas vulneráveis, incluindo-as no desenvolvimento e implementação desses sistemas de Inteligência Artificial e se considerar cuidadosamente situações em que os sistemas de Inteligência Artificial possam causar ou intensificar impactos negativos, especialmente quando há desequilíbrios de poder ou de informação, como entre empregadores e trabalhadores, empresas e consumidores, ou governos e cidadãos.

A prevenção de danos também envolve a consideração do meio ambiente natural e do bem-estar de todos os seres vivos.

Intimamente ligado a esse princípio está o da solidez técnica, que exige:

> que os sistemas de IA sejam desenvolvidos seguindo uma abordagem de prevenção dos riscos e de forma a que se comportem fiavelmente conforme o previsto, minimizando os danos não intencionais e inesperados, e prevenindo os danos inaceitáveis. Tal deverá também ser aplicado a eventuais alterações do ambiente em que operam ou à presença de outros agentes (humanos e

PARA-REGULAC%CC%A7A%CC%83O-DE-INTELIGE%CC%82NCIA-ARTIFICIAL-1.pdf. Acesso em: 10 set. 2023.

138 *Ibidem*

> artificiais) que possam interagir com o sistema de forma antagónica. Além disso, deve assegurar-se a integridade física e mental dos seres humanos.[139]

Aqui a prevenção da ameaça à privacidade exige uma governança adequada dos dados, que assegure a qualidade, assim como a integridade dos dados utilizados pelos sistemas nas tomadas de decisão, a sua relevância para o domínio em que os sistemas de Inteligência Artificial serão implantados, bem como os seus protocolos de acesso e a capacidade de tratar os dados de modo a proteger a privacidade dos titulares e usuários.

Por fim, a abordagem *ex ante* de atuação sobre os riscos gerados pelas decisões de algoritmos impõem breves reflexões sobre a Governança Algorítmica conforme se verá mais adiante.

### 3.2.3 Equidade

O uso dos sistemas de Inteligência Artificial para a tomada de decisão que impactam a vida das pessoas deve ser justo e imparcial. O princípio da equidade representa um compromisso em garantir uma distribuição justa e equitativa dos benefícios e custos, evitar qualquer viés injusto, discriminação ou estigmatização contra indivíduos e grupos e promover a igualdade de oportunidades no acesso à saúde, educação, bens, serviços e tecnologia.

Além disso, o princípio da equidade engloba a garantia de que os usuários desses sistemas ou sujeitos das decisões automatizadas não sejam enganados ou privados da sua liberdade de escolha.

Por outro lado, a equidade também requer que os profissionais de Inteligência Artificial ajam de acordo com o princípio da proporcionalidade, pesando cuidadosamente os meios e os fins, e equilibrando os interesses e objetivos em jogo.

Por fim, em uma vertente processual, o princípio da equidade também implica que haja mecanismos para contestar e buscar recursos eficazes contra as decisões tomadas tanto pelos sistemas de Inteligência Artificial quanto pelos humanos que os utilizam. Para isso, é fundamental que a entidade responsável pela decisão seja identificável e que os processos de tomada de decisão sejam transparentes e explicáveis, o que será trabalhado mais adianta no próximo capítulo.

---

[139] RELATÓRIO DA COMISSÃO EUROPEIA. Orientações éticas para uma IA de confiança. European Commission. 2019. Disponível em: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. Acesso em: 13 jan. 2024.

Em um estudo publicado na conferência WWW 2018 (*The 2018 Web Conference*), realizada em 2018 em Lyon, França[140], com o tema "*Human Perceptions of Fairness in Algorithmic Decision Making: A Case Study of Criminal Risk Prediction*", um grupo de estudiosos se dedicou a explorar como as pessoas percebem e raciocinam sobre a justiça na tomada de decisões algorítmicas, com foco na previsão de risco criminal.

Os pesquisadores propuseram um *framework* que identificou oito propriedades-chave que informam os julgamentos morais das pessoas sobre a justiça do uso de características em algoritmos de tomada de decisão, quais sejam:

i. Confiabilidade: se informações sobre a característica podem ser avaliadas de forma confiável.
ii. Relevância: se informações sobre a característica são relevantes para a tomada dessa decisão.
iii. Privacidade: se informações sobre a característica são privadas.
iv. Volicionalidade: se uma pessoa pode alterar a característica fazendo uma escolha ou decisão.
v. Causa do Resultado: se a característica pode fazer com que a pessoa viole sua liberdade condicional.
vi. Causa de um Ciclo Vicioso: se tomar essa decisão com base em informações sobre a característica pode causar um ciclo vicioso.
vii. Causa Disparidade nos Resultados: se tomar essa decisão com base em informações sobre a característica pode ter efeitos negativos em certos grupos de pessoas protegidos por lei.
viii. Causada pela Associação a um Grupo Sensível: se a característica pode ser causada pelo pertencimento a um grupo protegido por lei (por exemplo, raça, gênero, idade, religião, origem nacional ou status de deficiência).

A conclusão do estudo destaca várias descobertas importantes sobre essas propriedades, entre elas as preocupações das pessoas sobre a injustiça de usar determinada característica na tomada de decisão, que se estendem além da discriminação, incluindo considerações das outras propriedades elencadas acima, como a relevância da característica para o cenário de tomada de decisão e a confiabilidade com que a característica pode ser avaliada.

Um exemplo sobre o uso de dados pessoais e a percepção de que a invasão da privacidade representa um critério para que a tomada de decisão automática seja considerada inválida é o do uso da história de abuso de substâncias de um réu como uma característica em um algoritmo de previsão de risco de reincidência criminal[141].

---

[140] GRGIĆ-HLAČA, Nina *et al.* **Human Perceptions of Fairness in Algorithmic Decision Making:** A Case Study of Criminal Risk Prediction. *In:* WWW 2018: The 2018 Web Conference, 2018. p. 903-912. Disponível em: https://doi.org/10.1145/3178876.3186138. Acesso em: 03 fev. 2024.

[141] *Ibidem*.

Nesse caso, a sensibilidade da informação sobre o uso de drogas na juventude pode levar as pessoas a considerarem essa característica como injusta de ser utilizada no algoritmo, devido à sua natureza privada e potencialmente invasiva.

Foi também identificado que existem consideráveis discordâncias sobre quais características diferentes pessoas percebem como injustas de serem utilizadas como critério na tomada de decisão algorítmica. Ademais, a falta de consenso pode ser atribuída às discordâncias na forma como as pessoas avaliam as propriedades das características, especialmente aquelas concernentes às relações causais entre características de entrada e a sua influência causal nos resultados.

É evidenciando assim a complexidade do julgamento humano e a suscetibilidade a ruídos, conforme discutido no mencionado livro[142], podendo a eliminação do ruído por meio de decisões algorítmicas ser algo positivo em situações que demanda uma estrita igualdade de decisão. Como em algumas hipóteses no campo de decisões judiciais nas quais situações exatamente idênticas recebem tratamentos absolutamente divergentes. Não são raros os relatos de jurisdicionados que passaram pelo mesmo dano, ingressam com ações judiciais idênticas, e saem com sentenças opostas.

Este tema do uso de algoritmos pelo Poder Judicial também dialoga com o preste objeto de estudo, uma vez que caso se torne usual a utilização de inteligência artificial para a tomada de decisão em processos judiciais, igualmente será de grande relevância garantir que essa inteligência artificial seja *trustworthy,* "digna de confiança", e passível de efetiva contestação.

Em trabalho dedicado ao tema de questões trazidas pela Inteligência Artificial no Sistema Judiciário Brasileiro, Saulo José Casali Bahia[143] traz um surpreendente levantamento de como o uso de sistemas de Inteligência Artificial já é uma realidade nos mais diversos sistemas judiciais no mundo, inclusive o brasileiro.

O autor descreve a evolução da justiça digital no Brasil em três fases principais: a digitalização de processos judiciais (substituindo documentos físicos por eletrônicos), a automação de tarefas administrativas e judiciais, e, mais recentemente, o uso de Inteligência Artificial mais avançada, como aprendizado de máquina e *deep learning*. Casali menciona ainda várias iniciativas de Inteligência Artificial implementadas em diferentes tribunais brasileiros, incluindo plataformas como VICTOR (utilizado para o auxílio da resolução de recursos

---

[142] KAHNEMAN, Daniel; SUNSTEIN, Cass; SIBONY, Olivier. **Ruído:** Uma falha no julgamento humano. São Paulo: Objetiva, 2021.

[143] BAHIA, Saulo José Casali. Digital justice, robot judges and new challenges and perspectives for the justice: issues posed by AI in the Brazilian court system. **European Review of Public Law**, v. 36, n. 1, 2024.

extraordinários, identificando os temas de Repercussão Geral dentro do Supremo Tribunal Federal) e o SIGMA (sistema inteligente que utiliza modelos para a elaboração de minutas de decisões). [144]

Inclusive, já existe uma plataforma nacional de armazenamento, treinamento supervisionado, controle de versionamento, distribuição e auditoria dos modelos de Inteligência Artificial dentro do âmbito do Poder Judiciário, a Plataforma Sinapses, cuja responsabilidade por manter é do Departamento de Tecnologia da Informação do CNJ.

Mas este tema, contudo, por si só, já é um tema complexo e que demanda análises profundas de impacto, motivo pelo qual não está sendo foco desta pesquisa.

Essas observações destacam a necessidade de abordar preocupações de equidade além da discriminação e a importância de reunir dados objetivos sobre propriedades para eventualmente informar um heurística comum de justiça. Assim, o estudo destaca a importância de considerar as percepções humanas de justiça para melhorar os processos de tomada de decisão algorítmica.

### 3.2.4 Transparência e Explicabilidade

Se fosse necessário escolher dentre os princípios para uma Inteligência Artificial de confiança o que possui maior relevância, peso e destaque nos estudos sobre o tema, o princípio da transparência provavelmente seria a opção mais óbvia. Esee fato se dá especialmente em razão da transparência ser fundamental para se alcançar grande parte dos outros princípios.

É assim que este princípio passa: i) pelo entendimento do funcionamento da IA, uma vez que a transparência permite que os desenvolvedores, reguladores e usuários entendam como um sistema de Inteligência Artificial toma decisões e chega a suas conclusões. Isso é crucial para avaliar a confiabilidade do sistema e garantir que suas operações sejam compreensíveis e previsíveis; ii) pela possibilidade de detecção de viés e discriminação, facilitando a exposição das características dos dados e dos processos de tomada de decisão, tornando, teoricamente, possível examinar se o sistema está agindo de maneira justa e imparcial; iii) processo de responsabilização, uma vez que a transparência é essencial para estabelecer a responsabilidade pelos resultados produzidos por sistemas de Inteligência Artificial. Quando algo dá errado ou ocorre um resultado indesejado, é necessário que haja clareza sobre como e por que isso aconteceu, para que as partes responsáveis possam ser identificadas e responsabilizadas; iv)

[144] *Ibidem*.

pela confiabilidade, uma vez que se acredita que os sistemas de Inteligência Artificial transparentes tendem a ser mais confiáveis para os usuários, pois eles podem entender melhor como as decisões são tomadas e confiar na consistência e na justiça do sistema; e v) pela possibilidade de se facilitar processos de auditoria e a melhoria contínua.

Neste sentido, Paulo Sá Elias bem explica que:

> Até mesmos os maiores defensores desses sistemas, admitem essa fraqueza. Embora as redes neurais profundas (DNN – Deep Neural Networks) tenham demonstrado uma grande eficácia em uma ampla gama de tarefas, quando eles falham, muitas vezes falham espetacularmente, catastroficamente, produzindo resultados inexplicáveis e incoerentes que podem deixar o ser humano perplexo, sem conseguir entender a razão pela qual o sistema tomou tais decisões.[145]

Com isso, a falta de transparência em sistemas com processos de decisão em redes neurais profundas ainda é um grande obstáculo para a sua adoção em segmentos nos quais as decisões devem ser altamente confiáveis com pouca tolerância ao erro, como no campo da saúde e segurança[146].

O princípio da transparência, assim, passa pela compreensão da transparência e conhecimento sobre os dados, os sistemas e os modelos de negócio. O direito à explicação, por sua vez, decorre igualmente do princípio da transparência. Esse é o caso por exemplo das legislações voltadas à proteção de dados pessoais, como a LGPD, que traz no art. 6º que as atividades de tratamento de dados pessoais deverão observar o princípio da transparência e conceitua esses como a "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". No seu art. 20, garante-se o direito de revisão de decisões tomadas unicamente por tratamento automatizado.

Para o Grupo Independente de Peritos de Alto Nível Sobre a Inteligência Artificial (GPAN) o princípio da transparência inclui a rastreabilidade, a explicabilidade e a comunicação. A rastreabilidade aplicada às decisões tomadas por sistemas de Inteligência Artificial se refere ao fato de que os dados e os processos que produzem as decisões, o que inclui os processos de coleta e etiquetagem dos dados, bem como os algoritmos utilizados, devem ser documentados para permitir a rastreabilidade. Assim, com a rastreabilidade:

---

[145] ELIAS, Paulo Sá. Algoritmos, inteligência artificial e o direito. Consultor Jurídico, 2017. Disponível em: https://www.conjur.com.br/dl/algoritmos-inteligencia-artificial.pdf. Acesso em: 04 fev. 2021. p. 5.

[146] *Ibidem*.

> [...] é possível identificar os motivos por que uma decisão de Inteligência Artificial foi errada, o que, por sua vez, poderá ajudar a evitar erros futuros. A rastreabilidade facilita, assim, a auditabilidade e a explicabilidade.[147]

A auditabilidade, especificamente, é uma importante forma de verificar o funcionamento de sistemas de computador ao tratar o processo de tomada de decisão como uma caixa preta, na qual as entradas e saídas são visíveis, mas o funcionamento interno não. Além disso, ela serve também para avaliar a conformidade das decisões com as regulamentações aplicáveis e detectar possíveis usos vedados como discriminações.

Contudo, as limitações da auditoria não são irrelevantes. Ao tratar a decisão de um algoritmo como uma caixa preta, os detalhes internos do funcionamento do sistema não são visíveis, o que limita a capacidade de compreender completamente como as decisões são tomadas e quais os critérios que foram efetivamente aplicados. Ademais, a transparência completa nem sempre é alcançada, o que pode dificultar a compreensão do porquê certos comportamentos diferenciados foram observados em um sistema[148].

Por fim, mesmo em auditorias estruturadas de sistemas de software, em que são fornecidos *inputs* relacionados e analisados para comportamentos diferenciais, a cobertura completa do comportamento de um programa pode não ser alcançada. Isso ocorre porque a metodologia não explica o que acontece com inputs que não foram testados, mesmo que sejam ligeiramente diferentes.

Assim, em razão desses desafios, ao se pensar em auditoria destas decisões, é importante se considerar outras estratégias e abordagens para garantir a conformidade, detectar problemas e promover a transparência nos processos automatizados, conforme se verá mais adiante.

A explicabilidade, por sua vez, diz respeito à capacidade de explicar tanto os processos técnicos de um sistema de Inteligência Artificial como as decisões humanas com eles relacionadas:

> A explicabilidade técnica exige que as decisões tomadas por um sistema de IA possam ser compreendidas e rastreadas por seres humanos. Além disso, poderá ser necessário adotar soluções de compromissos entre o reforço da

---

147 RELATÓRIO DA COMISSÃO EUROPEIA. Orientações éticas para uma IA de confiança. European Commission. 2019. Disponível em: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. Acesso em: 13 jan. 2024.

148 KROLL, Joshua A. *et al.* Accountable Algorithms. Forthcoming, **Fordham Law Legal Studies Research Paper**, v. 165, n. 2765268. University of Pennsylvania Law Review, 2017. Disponível em: https://ssrn.com/abstract=2765268. Acesso em: 05 fev. 2024.

> explicabilidade de um sistema (o que poderá reduzir a sua exatidão) ou o aumento da sua exatidão (à custa da sua explicabilidade).[149]

Para o GPAN, sempre que um sistema de Inteligência Artificial tenha um impacto significativo, deverá ser possível solicitar uma explicação – oportuna, adequada à parte interessada - sobre o processo de tomada de decisões. A explicação assim, seria diferente para um usuário leigo, um regulador ou um investigador.

> Além disso, devem ser disponibilizadas explicações sobre o grau de influência e de intervenção de um sistema de IA no processo decisório da organização, as opções de conceção do sistema e os fundamentos da sua implantação (assegurando assim a transparência do modelo de negócio).[150]

Por fim, faz parte do princípio da transparência a comunicação, considerando que os usuários têm direito a serem informados de que estão interagindo com um sistema de Inteligência Artificial. A decisão por um sistema de Inteligência Artificial deve ser identificada como tal e deveria ser possível se optar pela interação humana. Ademais, também faz parte da comunicação a necessidade de se comunicar aos utilizadores informações relevantes, como o nível de exatidão do sistema de Inteligência Artificial e eventuais limitações.

Esses fatores são importantes principalmente porque nem sempre "ver" um sistema é o mesmo que saber como ele funciona e muito menos governá-lo. É assim que Mike Ananny e Kate Crawford questionam que a abertura das "caixas-pretas" não é necessariamente suficiente e que a transparência é muitas vezes inadequada para entender e governar esses sistemas de tomada de decisão por meio de Inteligência Artificial[151].

Entretanto, vale observar que os segredos comercial e industrial constituem em diversos momentos objeções à transparência. Quando estamos falando da regulamentação do uso de dados pessoais, por exemplo, elemento central para o funcionamento de parte relevante dos sistemas de Inteligência Artificial, a LGPD apresenta de forma objetiva opções ao princípio da transparência privilegiando interesses sociais ou individuais dos agentes[152] de tratamento de dados. O racional e o rigor técnico dessas opções, contudo, não estão isentos de dúvidas, inseguranças jurídicas e questionamentos.

---

149 RELATÓRIO DA COMISSÃO EUROPEIA. Orientações éticas para uma IA de confiança. European Commission. 2019. Disponível em: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. Acesso em: 13 jan. 2024.

150 *Ibidem*.

151 ANANNY, Mike; CRAWFORD, Kate. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. **New media & society**, [s.l.], v. 20, n. 3, 2018.

152 Entende-se por "agente de tratamento de dados" a pessoa natural ou jurídica que desempenha o papel de controlador ou operador nos termos do art. 5º da LGPD.

Esse é o caso das constantes relativizações de direitos e princípios em razão dos termos "observados", "pela observância" e "respeitados" os "segredos industriais e comerciais", que aparecem em treze diferentes momentos na LGPD.

É assim que o princípio da transparência, que estabelece a garantia de informações claras, precisas e acessíveis aos titulares, bem como o consequente direito de acesso à informação sobre a forma e duração do tratamento dos seus dados, devem ser vistos com a observância dos segredos comercial e industrial.

Igualmente, na solicitação pela Autoridade Nacional de Proteção de Dados, o relatório de impacto à proteção de dados pessoais (documento que contém os processos de tratamento e as garantias de segurança das informações e mecanismos de mitigação de riscos), também tem como garantia do controlador do dado a observância dos seus segredos comercial e industrial.

> A abordagem da precaução parece ser útil na definição dos contornos desse debate. De um lado, ela colabora na construção de espaços de deliberação para se discutir o que seria "informação qualificada" ou como mitigar os problemas em decisões futuras. De outro, ela possibilita endereçar questionamentos a respeito dos segredos comercial e industrial. Ao exigir informações sobre a racionalidade de uma decisão específica, o direito à explicação não se confunde com a transparência pura e simples. Variações nos dados de raça, por exemplo, já poderiam fornecer o impacto e a maneira como esse tipo de dado impacta uma decisão, sem, contudo, demandar a revelação de todo o sistema automatizado envolvido naquela decisão.[153]

Certo parece ser, contudo, a necessidade de equilibrar a transparência, segredos comerciais, mas também políticas de segurança e a proteção de interesses como a privacidade pessoal. Ademais a transparência completa nem sempre é viável ou desejável em certos contextos, como na auditoria de impostos ou na segurança aeroportuária, em que a opacidade parcial pode ser necessária para evitar abusos.

Nas decisões sobre auditoria de impostos, a opacidade parcial pode sim ser necessária para evitar que indivíduos manipulem o sistema de auditoria de impostos. Bem como nos processos de triagem de segurança no aeroporto, por exemplo, no qual a divulgação completa

---

[153] Doshi-Velez; Kortz, 2017 *apud* Bioni; Luciano, 2019, p. 13. Disponível em: BIONI, Bruno; LUCIANO, Maria. O princípio da precaução na regulação de inteligência artificial: seriam as leis de proteção de dados o seu portal de entrada. **Inteligência Artificial e Direito.** São Paulo: Thomson Reuters Brasil, p. 207-231, 2019. Disponível em: https://brunobioni.com.br/home/wp-content/uploads/2019/09/Bioni-Luciano_O-PRINCI%CC%81PIO-DA-PRECAUC%CC%A7A%CC%83O-PARA-REGULAC%CC%A7A%CC%83O-DE-INTELIGE%CC%82NCIA-ARTIFICIAL-1.pdf. Acesso em: 10 set. 2023.

dos critérios de seleção pode comprometer a eficácia da segurança, permitindo que potenciais ameaças contornem o sistema.[154]

No entanto, qual o limite da opacidade? Quem define a parcialidade? Como se questiona o que não se sabe da existência?

> Manter aspectos de uma política de decisão em segredo pode ajudar a prevenir a manipulação estratégica de um sistema. Por exemplo, o IRS (Receita Federal) pode procurar por sinais em declarações de imposto que estão altamente correlacionados com evasão fiscal com base em declarações previamente auditadas. Mas se o público souber exatamente quais itens em uma declaração de imposto são tratados como sinais evidentes de fraude, os fraudadores podem ajustar seu comportamento e os sinais podem perder seu valor preditivo para a agência.[155]

Ademais, Bruno Bioni e Maria Luciano acrescentem que para mitigar os custos envolvidos em sistemas de explicação, os espaços deliberativos, representados pela participação de diversos atores podem ajudar, evitando que empresas menores sejam afetadas de forma desproporcional.

Assim, a *trustworthiness*, ou a confiabilidade nos sistemas de tomada de decisão algoritmos de Inteligência Artificial se relaciona com a confiabilidade (referente ao comportamento previsível em condições normais de uso), a robustez (a capacidade de manter a previsibilidade do sistema mesmo em condições inesperadas) e a resiliência (a capacidade de recuperar o comportamento confiável do sistema após uma interrupção).

> A transparência e/ou explicabilidade dos sistemas algorítmicos beneficia esses fatores de confiabilidade ao ajudar a entender melhor como os sistemas se comportam além dos pontos de dados discretos fornecidos pelos testes de treinamento/teste/validação. Os requisitos de justiça e responsabilidade apoiam mais indiretamente a confiabilidade devido ao rigor aumentado da inspeção do comportamento do sistema que é necessário para controlar a justiça e estabelecer responsabilidade.[156]

A confiança, assim, no comportamento de um sistema de tomada de decisão automático refere-se à percepção humana do sistema como sendo digno de confiança, o que inclui julgamento baseado em avaliações sobre os valores éticos percebidos e incorporados ao sistema

---

[154] KROLL, Joshua A. *et al*. Accountable Algorithms. Forthcoming, **Fordham Law Legal Studies Research Paper**, v. 165, n. 2765268. University of Pennsylvania Law Review, 2017. Disponível em: https://ssrn.com/abstract=2765268. Acesso em: 05 fev. 2024. p. 657-659 (Tradução nossa).

[155] *Ibidem*, p. 658 (Tradução nossa).

[156] EUROPEAN PARLIMENT RESEARCH SERVICE. A governance framework for algorithmic accountability and transparency. Scientific Foresight Unit, 2019. Disponível em: https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf. Acesso em: 10 out. 2023. (Tradução nossa).

(expressos pelos requisitos de imparcialidade e responsabilidade) e fatores como a razoabilidade dos resultados das decisões, facilitada pela transparência/explicabilidade algorítmica.

Nesse mesmo sentido, John Giannandrea, o cientista de AI da Google, destaca os riscos de sistemas inescrutáveis:

> É importante que sejamos transparentes sobre os dados de treinamento que estamos usando e estejamos procurando por vieses ocultos neles, caso contrário, estamos construindo sistemas tendenciosos [...] Se alguém está tentando vender para você um sistema de caixa preta para suporte a decisões médicas, e você não sabe como ele funciona ou quais dados foram usados para treiná-lo, então eu não confiaria nele.[157]

A área médica também vem enfrentando esses desafios. Para Daniel de Araujo Dourado e Fernando Mussa Abujamra Aith, os limites para se ter uma inteligência artificial explicável na saúde passaria por um *trade-off* entre a explicabilidade e a precisão dos sistemas de tomada de decisão pela Inteligência Artificial.

> Para que um sistema de IA seja explicável, na maioria das vezes é necessária uma redução das variáveis da solução a um conjunto pequeno o suficiente para que fique acessível ao entendimento humano. Isso pode inviabilizar o uso de alguns sistemas em problemas complexos. Alguns modelos de deep learning podem prever probabilidades de diagnósticos clínicos com precisão, mas serem humanamente incompreensíveis. Nesse sentido, um direito à explicação mais amplo, baseado na máxima transparência, pode ser incompatível com o uso de sistemas automatizados que busquem alta acurácia preditiva.[158]

É assim que para esses autores, por mais que se tenha explicações úteis para o desenvolvimento de modelo ou para auditorias, raramente essas explicações são informativas o suficiente com relação aos resultados específicos dados pelos sistemas de Inteligência Artificial.

Vale ressaltar, desde já, que inobstante a importância desses princípios, alguns autores já questionam a posição de que a transparência resolveria os problemas relativos à prestação de contas e *accountabillity* das decisões tomadas por sistemas de Inteligência Artificial:

---

[157] KNIGHT, Will. Forget Killer Robots – Bias Is the Real AI Danger. **MIT Technology Review**. 2017. Disponível em: https://www.technologyreview.com/s/608986/forget-killer-robotsbias-is-the-real-ai-danger. Acesso em: 28 jan. 2024. (tradução nossa).

[158] DOURADO, Daniel de Araújo; AITH, Fernando Mussa Abujamra. A regulação da inteligência artificial na saúde no Brasil começa com a Lei Geral de Proteção de Dados Pessoais. **Revista Saúde Pública**, [s.l.], v. 56, n. 80. 2022. Disponível em: https://www.scielo.br/j/rsp/a/k38jGvJdbQSYN4MpzGZpfXw/?format=pdf&lang=pt. Acesso em: 3 mar. 2024.

> A divulgação do código-fonte muitas vezes não é necessária (por causa de técnicas alternativas da ciência da computação) nem suficiente (devido às questões relacionadas à análise do código) para demonstrar a justiça de um processo. Além disso, a transparência pode ser indesejável, como quando revela informações privadas ou permite que fraudadores de impostos ou terroristas manipulem os sistemas que determinam auditorias ou triagens de segurança.[159]

Neste sentido, Isabela Ferrari[160] bem destaca o que ela chama da "falácia da transparência". Entendendo que os debates muitas vezes focam mais na acessibilidade do que na compreensibilidade e que a questão chave para uma prestação de contas efetiva estaria nessa segunda característica.

> Isso porque, diante da estrutura cada vez mais complexa dos algoritmos que empregam machine learning, a mera abertura do código-fonte, por si só, tende a não auxiliar a compreensão da forma como operam, já que o referido código só expõe o método de aprendizado de máquinas usado, e não a regra de decisão, que emerge automaticamente a partir dos dados específicos sob análise.[161]

E, por tudo quanto exposto, a resposta para a real possibilidade de *accountability*, não poderia ser encontrada apenas por meio de uma solução jurídica, mas sim também o desenho de políticas públicas e as ferramentas que as tecnologias e a ciência da computação podem prover conforme se verá mais adiante.

### 3.2.5 Prestação de contas e *accountability*

Intimamente ligado ao mencionado princípio da explicação é o da prestação de contas. Os princípios da transparência e explicação são, assim, ferramentas para a *accountability*, responsabilização, da Inteligência Artificial ao possibilitar expor a lógica da decisão e, assim, permitindo ao observador determinar a extensão em que um *input* particular ou a estrutura foi determinante ou influenciou um resultado de forma indevida. Para isso, a aplicação de testes em sistemas de Inteligência Artificial é essencial, apesar de não ser a solução técnica para todos os tipos de sistema.

---

159 KROLL, Joshua A. *et al*. Accountable Algorithms. Forthcoming, **Fordham Law Legal Studies Research Paper**, v. 165, n. 2765268. University of Pennsylvania Law Review, 2017. Disponível em: https://ssrn.com/abstract=2765268. Acesso em: 05 fev. 2024. p. 633 (Tradução nossa).

160 FERRARI, Isabela. Accountability de algoritmos: a falácia do acesso ao código e caminhos para uma explicabilidade efetiva. **Inteligência Artificial:** 3º Grupo de Pesquisa do ITS, ITS - Instituto de Tecnologia e Sociedade do Rio, 2018. Disponível em: https://itsrio.org/wpcontent/uploads/2019/03/Isabela-Ferrari.pdf. Acesso em: 10 jul. 2024.

161 *Ibidem* p. 12.

A *accountability* pode ser definida como um conjunto de mecanismos, práticas e atributos que somam a uma estrutura de governança que envolve o comprometimento com obrigações legais e éticas, políticas, procedimentos e mecanismos, explicando e demonstrando implementação ética para partes interessadas internas e externas e corrigindo qualquer falha em agir adequadamente[162].

Aqui percebe-se com clareza como todos os princípios apresentados são interligados. Não é possível se falar em prestação de contas e *accountability*, sem considerar a transparência, a explicabilidade, a prevenção de danos, dentre outros aspectos.

Assim, esses princípios envolvem o aspecto crucial da auditabilidade (algoritmos transparentes e passíveis de auditoria, com operações e decisões que devem ser compreensíveis e rastreáveis, permitindo que especialistas e partes interessadas examinem como as decisões foram alcançadas), mas não se resume a ele.

A minimização de impactos negativos das decisões algorítmicas também é um ponto chave. Isso envolve a consideração cuidadosa dos possíveis efeitos adversos das decisões e a implementação de medidas para mitigá-los. Além disso, há a comunicação de forma transparente e clara sobre esses impactos negativos para os indivíduos afetados e outras partes interessadas. Igualmente inclui a possibilidade de se recorrer e contestar as decisões.

Por esses princípios, deve haver mecanismos eficazes de recurso disponíveis para as pessoas afetadas pelas decisões dos algoritmos dos sistemas de IA. Isso significa que elas devem ter a capacidade de contestar decisões injustas ou prejudiciais e buscar correções ou compensações apropriadas.

Todos esses elementos são fundamentais para garantir que os sistemas de Inteligência Artificial sejam responsáveis e atuem de maneira ética e justa.

Importante contribuição ao tema foi feito pelo *European Parliamentary Research Service* no trabalho "*A Governance framework for algorithmic accountability and transparency*", trabalho que propõe várias medidas para promover a *accountability* em sistemas algorítmicos, incluindo a já mencionada transparência, a revisão de design, a verificação formal, bem como destaca a importância de padrões de processo e certificação como ISSO, IEC, ITU, IEEE[163].

---

[162] EUROPEAN PARLIMENT RESEARCH SERVICE. A governance framework for algorithmic accountability and transparency. Scientific Foresight Unit, 2019. Disponível em: https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf. Acesso em: 10 out. 2023. (Tradução nossa).

[163] *Ibidem* (Tradução nossa).

A revisão de design e revisão de código é assim recomendada juntamente com procedimentos de teste para garantir que os sistemas algorítmicos atendam às especificações e, por sua vez, a verificação formal é a sugestão do uso de técnicas para garantir que os sistemas algorítmicos atendam às especificações e funcionem conforme o esperado.

> Processos tradicionais de design de software incluem revisão de design, revisão de código e procedimentos de teste para garantir que sistemas algorítmicos atendam às especificações. Além disso, técnicas de verificação formal estão fazendo avanços significativos. A verificação formal foi demonstrada em artefatos de software importantes, é provável que essas técnicas se tornem parte da prática padrão de engenharia de software.[164]

O mencionado trabalho não tem a intenção de especificar e determinar quais ferramentas, técnicas e perspectivas futuras seriam melhores para manter os sistemas de tomada de decisão por meio de Inteligência Artificial responsáveis. Mas traz uma discussão sobre a importância de analisar sistemas com base em *inputs*, *outputs* e informações simples sobre os algoritmos utilizados, sem necessidade de acesso ao código-fonte subjacente e também menciona a necessidade de fornecer dados de treinamento ou um registro de decisões passadas para pesquisadores

Ademais, sinaliza que especialistas de uma ampla variedade de disciplinas precisarão ter flexibilidade para se adaptar a novos métodos de responsabilidade à medida que novas formas de tomada de decisão automatizada surgirem.

Não obstante, alguns caminhos são sugeridos. Assim, é sinalizado que a pesquisa e a auditoria realizadas nesses sistemas devem ser responsáveis perante o público e incluir um registro público de quais pesquisadores e especialistas têm acesso e com base em quais critérios.

As autoridades públicas devem garantir que as comunidades afetadas possam sugerir pesquisadores que elas sintam que representam os seus interesses e devem trabalhar com os pesquisadores para garantir que essas comunidades tenham voz na formulação das perguntas que são feitas e abordadas pela pesquisa e auditoria.

> É importante, para garantir a responsabilidade pública e um campo de pesquisa próspero, que os resultados e conclusões da pesquisa sejam publicados abertamente (mesmo que após um período de embargo) e sejam submetidos a padrões de escrutínio e revisão por pares dentro dos domínios de pesquisa apropriados.

---

[164] EUROPEAN PARLIMENT RESEARCH SERVICE. A governance framework for algorithmic accountability and transparency. Scientific Foresight Unit, 2019. Disponível em: https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf. Acesso em: 10 out. 2023. (Tradução nossa).

> Auditorias contínuas e acesso à pesquisa permitiriam que autoridades públicas, pesquisadores e comunidades afetadas trabalhassem juntas para desenvolver suas abordagens para testar e interrogar esses sistemas. Isso é especialmente importante dado que a pesquisa sobre responsabilidade algorítmica é recente e o desenvolvimento tecnológico avança rapidamente. Ainda não sabemos quais ferramentas, técnicas e perspectivas futuras poderiam manter os sistemas mais responsáveis.[165]

Vale olhar aqui esse princípio também sob a perspectiva do uso de dados pessoais, uma vez que a compreensão de quais dados são utilizados para alimentar um sistema de tomada de decisão de IA, determinando algum resultado específico, é igualmente um ponto chave para que se tenha prestação de contas e confiança na justiça e equidade de uma decisão.

É assim que o princípio de livre acesso, previsto no art. 6º, inciso IV da LGPD, prevendo a garantia, aos titulares, de uma consulta facilitada aos seus dados, incluindo a integralidade de seus dados pessoais, e o princípio da qualidade dos dados, previsto no art. 6º, inciso V, da LGPD, trazendo a garantia de exatidão, clareza, relevância e atualização dos dados, de acordo com a necessidade e para o cumprimento da finalidade de seu tratamento, são essenciais para a capacidade do sujeito de uma decisão automatizada para avaliar e identificar a correção desses dados que pode identificar resultados incorretos.

Embora os referidos princípios da LGPD por si só não sejam suficientes para fornecer uma explicação de um resultado, eles são, minimamente, importantes para determinar se um resultado individual é baseado em dados corretos ou incorretos.

Contudo, nem sempre os dados são possíveis de acessar e em alguns casos, inclusive, podem ser substituídos por novas entradas:

> Existem muitos casos, no entanto, onde o acesso aos dados que produziram um resultado pode não estar disponível. Os dados são frequentemente considerados um ativo valioso que as organizações relutam em compartilhar. O RGPD, por exemplo, não obriga o acesso a dados não pessoais, como dados estatísticos sobre grandes grupos populacionais, que podem ter desempenhado um papel importante em uma decisão.
>
> Além disso, a menos que esforços sejam implementados para garantir que os dados sejam retidos, por exemplo, para fins de auditoria de dados, eles podem ser sobrescritos por novas entradas. Um exemplo típico onde esforços deliberados são feitos para reter dados que de outra forma desapareceriam são os gravadores de dados de voo. A inclusão obrigatória de gravadores de dados em veículos autônomos, por exemplo, foi sugerida para ajudar futuros investigadores de acidentes a acessar dados de entrada que precederam acidentes de carros autônomos.[166]

---

[165] *Ibidem* (Tradução nossa).

[166] *Ibidem* (Tradução nossa).

As mencionadas técnicas como revisão de design e de código teriam igualmente pouca relevância direta para entender um resultado individual no processo de *accountability*. No entanto, divulgar sinopses de tais revisões pode fazer parte do processo de apresentação de informações significativas sobre a lógica envolvida, auxiliando, por exemplo a cumprir o Artigo 15 1(h) do RGPD.

Importante sinalizar que é apontado como um dos benefícios deste tipo de princípio o aumento da confiança e confiabilidade nos sistemas de tomada de decisão por meio de IA.

Indiscutível, diante do exposto, contudo, é a necessidade de se existir mecanismos que requeiram e tragam imposição de critérios de responsabilização, *accountabillity* e prestação de contas. Entretanto, é possível que existam também "efeitos colaterais" na aplicação destes mecanismos que, por mais que não se sobreponham aos benefícios da imposição destes critérios, merecem ser brevemente sinalizados.

Um deles é o eventual comprometimento do desempenho do sistema. Quando aplicados ao desenvolvimento de um sistema de Inteligência Artificial de tomada de decisão de algorítmico, essas imposições podem criar critérios de desempenho adicionais que modificam os objetivos da otimização do sistema.

> Esta é uma propriedade inerente que ocorre sempre que um sistema é otimizado para satisfazer múltiplos requisitos que não são completamente independentes entre si, resultando na necessidade de fazer compensações (por exemplo, otimizar a espessura do chassi de um veículo motorizado para simultaneamente maximizar os requisitos de resistência ao impacto e eficiência de combustível inerentemente leva a compensações entre eles). Esse tipo de problema é referido na literatura como otimização multiobjetivo (ou multi-objetiva).[167]

Outro aspecto pode ser apontado como o impacto na distribuição de custos, uma vez que implementar requisitos de transparência e prestação de contas durante o desenvolvimento do sistema provavelmente exigirá outros custos adicionais de desenvolvimento como esforços adicionais em testes/validação do sistema e potencialmente certificação.

Contudo, até mesmo neste ponto, o aumento do rigor durante o desenvolvimento do sistema com este tipo de obrigação pode resultar em uma maior confiabilidade, robustez ou até mesmo resiliência, o que pode. por outro lado, eventualmente reduzir os custos de manutenção e de outros riscos.

Ainda na esfera de compreensão das ferramentas para a *accountability*, algumas outras estratégias que vão além da auditoria/transparência para a compreensão das decisões

[167] *Ibidem* (Tradução nossa).

automatizadas são trazidas no trabalho "*Accountable Algorithms*", publicado pela University of Pennsylvania Law Review.[168]

Essas estratégias incluem uma transparência proativa (em oposição da dependência exclusiva em auditoria, e envolvendo a divulgação ativa de informações sobre como os algoritmos funcionam e quais critérios são utilizados em processos de tomada de decisão), o desenvolvimento de ferramentas técnicas específicas (como *zero-knowledge proofs*[169]) e a definição de propriedades certificáveis (identificar e definir propriedades específicas que os sistemas devem atender, como a exclusão de informações sensíveis).

Os autores dessa obra propõem uma abordagem diferente para definir a justiça em classificações realizadas por sistemas de Inteligência Artificial, que seria a de garantir que o resultado não revele se o sujeito pertence a um grupo protegido. A lógica é a de que se o resultado não fornece uma visão melhor dos atributos do indivíduo do que adivinhar sem informações, pode ser considerado justo. Por exemplo, se alguém sendo negado um empréstimo indica que é mais provável que viva em um bairro específico, isso sugere que a decisão é baseada em fatores além do risco de crédito objetivo, potencialmente levando à discriminação.

> Assim, a justiça pode ser vista como uma forma de requisito de ocultação de informações semelhante à privacidade. Se aceitarmos que uma decisão justa não nos permite inferir os atributos do sujeito da decisão, somos forçados a concluir que a justiça está protegendo a privacidade desses atributos. De fato, é frequentemente o caso de que as pessoas estão mais preocupadas que suas informações sejam usadas para tomar alguma decisão ou classificá-las de alguma maneira do que estão com o fato de as informações serem conhecidas ou compartilhadas. Essa preocupação se relaciona com a famosa concepção de privacidade como o "direito de ser deixado em paz", na qual, geralmente, as pessoas estão preocupadas com a ideia de que a divulgação interrompe seu gozo de uma "personalidade inviolável. [170]

Considera-se, assim, que os métodos para construir sistemas justos de análise de dados e classificação geralmente exigem acesso a informações de "status protegido" já desde a fase

---

[168] KROLL, Joshua A. *et al.* Accountable Algorithms. Forthcoming, **Fordham Law Legal Studies Research Paper**, v. 165, n. 2765268. University of Pennsylvania Law Review, 2017. Disponível em: https://ssrn.com/abstract=2765268. Acesso em: 05 fev. 2024.

[169] "Uma prova de conhecimento zero é uma ferramenta criptográfica que permite a um tomador de decisão, como parte de um compromisso criptográfico, provar que a política de decisão que foi realmente usada (ou a decisão particular alcançada em um determinado caso) possui uma certa propriedade, mas sem ter que revelar nem como essa propriedade é conhecida nem qual é a política de decisão de fato." KROLL, Joshua A. *et al.* Accountable Algorithms. Forthcoming, **Fordham Law Legal Studies Research Paper**, v. 165, n. 2765268. University of Pennsylvania Law Review, 2017. Disponível em: https://ssrn.com/abstract=2765268. Acesso em: 05 fev. 2024. p. 668 (Tradução nossa).

[170] KROLL, Joshua A. *et al.* Accountable Algorithms. Forthcoming, **Fordham Law Legal Studies Research Paper**, v. 165, n. 2765268. University of Pennsylvania Law Review, 2017. Disponível em: https://ssrn.com/abstract=2765268. Acesso em: 05 fev. 2024. p. 690 (Tradução nossa).

de *design* do algoritmo. No entanto, preocupações sobre possíveis abusos levaram a restrições no uso dessas informações, o que faria com que a implementação dessas ferramentas técnicas exigisse uma mudança de política com o uso de técnicas para abordar preocupações sobre o uso indevido de dados.

## 3.3 RISCOS E A GOVERNANÇA ALGORÍTMICA

Conforme discutido ao longo deste trabalho, os impactos das decisões algorítmicas, que são difíceis de prever, identificar ou medir, exigem a adoção de medidas adequadas para mitigar esses riscos de maneira proporcional à sua dimensão. Ou seja, em termos de governança, trabalhar mais os mecanismos de mitigação de risco de acordo com a dimensão e a proporcionalidade do impacto negativo parece fazer sentido. Evidentemente um sistema de Inteligência Artificial que recomenda uma música não levanta as mesmas preocupações éticas que os sistemas de Inteligência Artificial que propõem tratamentos médicos ou dão acesso a serviços essenciais.

É notório existe um abismo de interesses entre os desenvolvedores/patrocinadores das tomadas de decisão por meio de Inteligência Artificial e aqueles que são impactados por ela. Para Bruno Bioni e Maria Luciano,

> Dentre as razões para isso têm sido apontadas a falta de regulação, monopólios no setor de IA, estruturas de governança insuficientes dentro de empresas de tecnologia, assimetrias de poder entre empresas e usuários, a distância cultural entre os responsáveis por pesquisas em tecnologia e a diversidade das populações nas quais essa tecnologia é utilizada.[171]

O conceito de risco já vem sendo utilizado como ferramenta útil em análises econômicas, matemáticas, biológicas etc. e, a seguir, será mostrado como esse conceito pode ser utilizado igualmente como cerne para a abordagem do uso dos algoritmos.

E o recorte da "risquificação" impõe a reflexão do contexto político-econômico-social que envolve a gestão e análise de riscos. É ponto fundamental de análise que o processo de definição e regulação de qualquer risco é um exercício de poder, que carrega os interesses e concepções políticas, econômicas e sociais de quem toma a decisão.

---

[171] BIONI, Bruno; LUCIANO, Maria. O princípio da precaução na regulação de inteligência artificial: seriam as leis de proteção de dados o seu portal de entrada. **Inteligência Artificial e Direito.** São Paulo: Thomson Reuters Brasil, p. 207-231, 2019. Disponível em: https://brunobioni.com.br/home/wp-content/uploads/2019/09/Bioni-Luciano_O-PRINCI%CC%81PIO-DA-PRECAUC%CC%A7A%CC%83O-PARA-REGULAC%CC%A7A%CC%83O-DE-INTELIGE%CC%82NCIA-ARTIFICIAL-1.pdf. Acesso em: 10 set. 2023. p. 2.

Historicamente, essa abordagem baseada em riscos tem encontrado desenvolvimento científico especialmente no campo da segurança pública e das regulamentações internacionais voltadas à regulação de riscos à saúde humana coletiva. É assim que Marcus Navarro esclarece que o modelo do sistema regulador de riscos referentes à saúde dependeu de conjunturas políticas, econômicas e sociais de cada país. Para isso, ao estudar os diferentes modelos de regulação existentes, constatou que ao se observar os conflitos de interesse sobre a divisão e priorização dos riscos, "não era possível separar as análises técnicas sobre os riscos das decisões de quem deveria ser protegido, dos custos e das alternativas disponíveis"[172].

Constatou-se que os estudos ou as avaliações de riscos eram feitos, essencialmente, para subsidiar tomadas de decisão de quem possuía maior poder, e a ideia de risco que parecia ser concebida como a "probabilidade de ocorrência de um evento indesejado, calculado pelos especialistas e apresentado à sociedade como uma verdade absoluta e neutra"[173] passou a ser questionada. Essa compreensão muito se aproxima da "teoria construtivista", na qual o risco pode ser entendido como uma construção social - construção essa feita segundo os interesses dos grupos ou instituições[174].

Nessa perspectiva teórica, o risco, como parte integrante dos sistemas sociais, fundamenta a criação de normas sociais e permite que os sistemas cumpram uma de suas funções, que é proporcionar aos indivíduos uma compreensão dos eventos sociais inesperados - os perigos – e também dos efeitos indesejados - os riscos - de determinadas ações.

Considerando essas observações e as reflexões anteriores sobre os riscos jurídicos decorrentes da tomada de decisão por algoritmos, é possível traçar paralelos entre os riscos relacionados a diferentes direitos com relativa facilidade. Podemos notar semelhanças entre uma regulamentação de riscos centrada na preservação e mitigação de danos à saúde e à segurança pública e aquelas voltadas para diferentes direitos fundamentais, como a liberdade,

---

172 NAVARRO, Marcus Vinícius Teixeira. Conceito e controle de riscos à saúde. *In:* NAVARRO, Marcus Vinícius Teixeira. **Risco, radiodiagnóstico e vigilância sanitária**. Salvador: EDUFBA, 2009. p. 37-75.

173 *Ibidem*, p. 69.

174 Navarro (2009) (*Ibidem*, p. 40) traz como exemplo a luta dos trabalhadores de minas de carvão que buscavam melhorias nas suas condições de trabalho décadas atrás. Para isso, esses trabalhadores buscaram mostrar como a sua atividade era uma das mais arriscadas, utilizando o número de mortes que estava entre os mais altos da mineração. Contudo, na relação de poder e interesse envolvidos, os proprietários das mineradoras preferiam utilizar o indicador que não fazia referência ao número de mortes versus o número de trabalhadores, mas sim o número de mortes versus a tonelada produzida. Com essa opção (claramente política), a atividade dos trabalhadores de minas de carvão deixava de aparecer entre as de mais alto risco. Ao contrário, a baixa produtividade fazia com que ela passasse a ser de baixo risco. O autor conclui, assim, que um simples coeficiente de mortalidade não era uma medida objetiva e única, mas sim algo subjetivo e passível de muitas definições.

dignidade humana e privacidade. A proteção jurídica é uma escolha política que envolve a regulamentação e a priorização de riscos.

E dentro deste conjunto de riscos que atualmente vem se apresentando como de relevante observação quando o foco é a privacidade e o uso de dados pessoais temos a contribuição de Mike Ananny e Kate Crawford[175] sobre a insuficiência do princípio da transparência. Assim, temos o risco da manutenção de determinados *status quo* de problemas sociais com o uso de dados que levam a um aprofundamento da assimetria de poder, confirmando, refletindo e perdurando de forma acrítica ou até mesmo "inconsciente" problemas que deveriam estar sendo combatidos.

Outro risco relevante é a da exposição da privacidade e vida íntima de indivíduos e de grupos marginalizados. É comum a afirmação de que "os algoritmos são tão bons quanto os dados nos quais são treinados" e se os dados utilizados por eles contêm preconceitos, estereótipos ou representam desigualdades sociais, os algoritmos aprenderão e reproduzirão esses preconceitos marginalizando ainda mais determinados grupos pela falta de representatividade (não representam adequadamente a diversidade da população)[176] em razão do "feedback loop" (ciclo de *feedback* negativo, como por exemplo, um algoritmo de policiamento preditivo que envia mais patrulhas para bairros marginalizados pode levar a mais prisões nessas áreas, reforçando a percepção de que esses bairros são mais criminosos)[177] ou até mesmo pela desigualdade de acesso de grupos marginalizados a determinados recursos tecnológicos (refletindo em menos oportunidades de influenciar o desenvolvimento e a implementação de algoritmos)[178].

Ademais, outros riscos envolvidos neste processo são referentes a uma perpetuação da crença binária entre segredo e transparência, de que a transparência inviabilizaria a proteção ao segredo comercial ou do negócio, a inviabilidade da disponibilidade de toda a informação

---

[175] ANANNY, Mike; CRAWFORD, Kate. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. **New media & society**, v. 20, n. 3, [s.l.], 2018.

[176] BAROCAS, Solon; HARDT, Moritz; NARAYANAN, Arvind. **Fairness and Machine Learning:** Limitations and Opportunities. Massachusetts Institute of Technology: The MIT Press, 2023. Disponível em: https://fairmlbook.org/pdf/fairmlbook.pdf. Acesso em: 10 jun. 2024.

[177] O'NEIL, Cathy. **Algoritmos de destruição em massa**: como o big data aumenta a desigualdade e ameaça a democracia. 1. ed. Santo André, So: Editora Rua do Sabão, 2020.

[178] EUBANKS, Virginia. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. **Law Technology and Humans**, v. 1, n. 1. 2019. Disponível em: https://www.researchgate.net/publication/337578410_Virginia_Eubanks_2018_Automating_Inequality_How_High-Tech_Tools_Profile_Police_and_Punish_the_Poor_New_York_Picador_St_Martin%27s_Press. Acesso em: 20 jun. 2024.

dentro de contextos e especificidades históricas e as abordagens liberais que assumem que todos os indivíduos têm total capacidade de compreender e assimilar as informações fornecidas[179].

Assim, mesmo diante da dificuldade de se conseguiu enxergar e predizer as consequências que advirão do uso crescente de decisões automatizadas em sistemas de IA, é fundamental a consciência de que muitos direitos estarão em jogo, e de que os instrumentos jurídicos atuais ainda não serão capazes de oferecer uma efetiva proteção.

Este trabalho defende que a Governança Algorítmica surge, assim, como uma ferramenta útil na abordagem *ex ante* de atuação sobre esses riscos gerados pelas decisões de algoritmos. Métodos focados em minimizar os riscos e maximizar os benefícios do uso de decisões feitas por algoritmos, considerando os impactos e a probabilidade de acontecerem, podem estabelecer eficientes gradações de controle, possibilitando uma alocação de recursos regulatórios e privados mais precisa e proveitosa.

Para a cognição dos riscos envolvendo as decisões automatizadas tomadas por algoritmos em sistemas de IA, e considerando a tendência de que cada vez mais situações importantes do cotidiano passem a ser resolvidas dessa forma, é imprescindível reforçar mais uma vez a crescente dificuldade de se decodificar o resultado de algoritmos capacitados de autonomia e inteligência artificial.

Com o avanço tecnológico, "os seres humanos vão ficando cada vez menos capazes de compreender, explicar ou prever o funcionamento interno, os vieses e os eventuais problemas dos algoritmos"[180]. Essa dificuldade antecipatória contribui, também, para a complexidade de um processo de governança sobre as causas e os efeitos desse objeto.

Temos, aqui, conforme mencionado anteriormente, a "opacidade", como característica destes sistemas, motivado por questões técnicas – que exigiriam, por exemplo, novas forma de engenharia de programação para a sua redução – e questões não técnicas, como opções deliberadas de se utilizar algoritmos fechados por motivações concorrenciais, de preservação de propriedade intelectual ou outras estratégias negociais.

Na Governança Algorítmica, portanto, impõe-se a necessidade de se garantir uma maior transparência dos complexos racionais por trás das decisões automatizadas, bem como a

---

[179] BIONI, Bruno; LUCIANO, Maria. O princípio da precaução na regulação de inteligência artificial: seriam as leis de proteção de dados o seu portal de entrada. **Inteligência Artificial e Direito.** São Paulo: Thomson Reuters Brasil, p. 207-231, 2019. Disponível em: https://brunobioni.com.br/home/wp-content/uploads/2019/09/Bioni-Luciano_O-PRINCI%CC%81PIO-DA-PRECAUC%CC%A7A%CC%83O-PARA-REGULAC%CC%A7A%CC%83O-DE-INTELIGE%CC%82NCIA-ARTIFICIAL-1.pdf. Acesso em: 10 set. 2023. p. 2.

[180] DONEDA, Danilo; ALMEIDA, Virgílio A. F. O que é governança de algoritmos. *In:* BRUNO, Fernanda *et al*. **Tecnopolíticas da vigilância:** perspectivas da margem. São Paulo: Boitempo, 2018. p. 142.

necessidade de uma maior possibilidade de responsabilização e prestação de contas. Sem o conhecimento dos fatores que baseiam a decisão do algoritmo, é impossível saber se se está diante de decisões ilícitas, discriminatórias ou antiéticas.

Foi o caso das eleições nos Estados Unidos de 2016, no qual a falta de transparência em relação aos algoritmos utilizados nas campanhas e às bases de dados que os alimentam tornou-se um ponto crucial para determinar se houve violação dos direitos humanos e do processo democrático de eleições. Isso porque os algoritmos desempenham um papel cada vez mais significativo na formulação de estratégias de campanha, direcionamento de mensagens políticas e segmentação do eleitorado. No entanto, a opacidade em torno desses algoritmos levanta preocupações sobre a possibilidade de manipulação, discriminação e influência indevida nas eleições[181].

A exigência de transparência nos algoritmos utilizados durante as campanhas eleitorais visaria garantir que o processo eleitoral seja justo, transparente e democrático. Isso envolve não apenas divulgar os algoritmos em si, mas também as fontes de dados usadas para treiná-los e alimentá-los. Os eleitores deveriam ter o direito de entender como as decisões políticas são tomadas e como são direcionadas as mensagens políticas que recebem, especialmente quando essas decisões são influenciadas por algoritmos e tecnologias complexas.

Assim, a falta de transparência nos algoritmos eleitorais pode levar a diversas preocupações, como a manipulação dos resultados das eleições, a disseminação de desinformação direcionada e a exclusão de determinados grupos de eleitores. Além disso, a opacidade nos algoritmos pode dificultar a responsabilização das partes envolvidas em caso de violações dos direitos humanos ou do processo democrático.

Portanto, a transparência dos algoritmos usados na tomada de decisões é crucial para o interesse público. Isso permite a verificação de possíveis vieses ilegais e discriminatórios, a identificação de erros e incoerências e a realização de auditorias e fiscalizações.

Aqui, é primordial mencionar a profunda relação entre a Governança Algorítmica e a proteção de dados pessoais, visto que, conforme já mencionado, grande parte das decisões tomadas por algoritmos, em diversos setores e aplicações tecnológicas se baseiam em dados pessoais coletados, armazenados e tratados em gigantescos bandos de dados - o já mencionado fenômeno do *big data*.

---

[181] Diário de Notícias. 2018. Disponível em: https://www.dn.pt/mundo/como-a-cambridge-analytica-ajudou-na-eleicao-de-trump-9209379.html/. Acesso em: 10 set. 2023.

Neste ponto, as construções jurídicas em torno da necessidade de proteção dos dados pessoais estão intimamente relacionadas com as consequências que podem advir da utilização enviesada, distorcida e manipulada deles.

Assim, a ideia de uma regulamentação baseada em riscos também tem sido defendida como uma abordagem adequada para a mitigação de determinados riscos relacionados à proteção da privacidade dos titulares de dados pessoais. É o caso do mencionado Regulamento Geral de Proteção de Dados, regulamentação europeia que visa disciplinar o tratamento de dados pessoais dos cidadãos europeus e da Lei Geral de Proteção de Dados, que trazem nas suas disposições características voltadas para uma abordagem baseada em riscos.

A LGPD cria a Autoridade Nacional de Proteção de Dados (ANPD), responsável por zelar, implementar e fiscalizar o cumprimento da LGPD, e determina a adoção de abordagens de mensuração de riscos estabelecendo que a ANPD irá editar regulamentos e procedimentos "sobre relatórios de impacto à proteção de dados pessoais para os casos em que o tratamento representar alto risco à garantia dos princípios gerais de proteção de dados pessoais".

São estabelecidas pela LGPD também obrigações preventivas e de autorregulação, típicas de uma abordagem baseada em riscos e que busca criar governança. Os responsáveis pelo tratamento dos dados pessoais devem elaborar relatórios de impacto, que serão documentos que devem conter a descrição dos processos de tratamento de dados que potencialmente ponham em risco às liberdades civis e aos direitos humanos. Esses relatórios devem conter a "descrição dos tipos de dados coletados, a metodologia utilizada para a coleta e para a garantia da segurança das informações e a análise do controlador com relação a medidas, salvaguardas e mecanismos de mitigação de risco adotados".

Essas características da governança sobre o uso dos dados pessoais podem ser, quase que integralmente, aplicáveis à Governança Algorítmica. Questões como se poderia existir igualmente uma "autoridade nacional" capaz de estabelecer regras e padrões de conduta para a construção de algoritmos, se essa autoridade seria alguma já existente, como a própria ANPD, que tem declarado interesse em ser protagonista no processo de regulamentação da IA[182] ou se o mais eficiente seria uma regulamentação multidisciplinar com a articulação institucional multissetorial.

O mencionado protagonismo da ANPD, neste sentido, foi visto recentemente em razão de questionamento sobre a nova política de privacidade da empresa Meta, que autorizava o uso

[182] BRASIL. Diretora da ANPD defende protagonismo da Autoridade na regulamentação da IA. Ministério da Justiça e Segurança Pública. 2024. Disponível em: https://www.gov.br/anpd/pt-br/assuntos/noticias/diretora-da-anpd-defende-protagonismo-da-autoridade-na-regulamentacao-da-ia. Acesso em: 07 jul. 2024.

de dados pessoais publicados em suas plataformas para fins de treinamento de sistemas de IA. Assim, no dia 2 de julho de 2024, a ANPD publicou em sede de decisão cautelar, a determinação de que a Meta suspendesse o tratamento de dados pessoais para treinamento da sua IA, levantando debates sobre a relevância que uma autoridade que tem como foco a proteção do uso de dados pessoais pode impactar de sobremaneira o desenvolvimento de sistemas de IA.

Vale ressaltar que a decisão da ANPD[183] não fugiu ao escopo do atual papel da autoridade, tendo sido fundamentada na LGPD e guardando coerência com o histórico regulamentar da instituição. A decisão foi pautada em indícios de tratamento de dados pessoais pela Meta com base em hipótese legal inadequada (inadequação de legítimo interesse e ausência da obtenção de consentimento), falta de transparência, limitação aos direitos dos titulares (inclusive com o uso de *dark patterns* dificultando o exercício de direitos*)* e riscos para crianças e adolescentes.

Essa decisão já demonstra o quanto o uso da Inteligência Artificial não comporta pausas nem irá aguardar a existência de alguma regulamentação. As instituições atuais, como a ANPD e o próprio CNJ (que rejeitou pedido para barrar uso de inteligência artificial no Judiciário),[184] já estão decidindo a respeito do seu uso.

Assim, só a discussão sobre quem seria a autoridade adequada para regular o uso da Inteligência Artificial já seria suficiente para o desenvolvimento de um trabalho acadêmico autônomo, motivo pelo qual esta pesquisa não adentrou de forma mais profunda neste debate.

Analogias para se pensar na responsabilização dos atores envolvidos na criação desses riscos (engenheiros, programadores, empresas públicas e privadas, etc.) poderia existir com os agentes de tratamento de dados. Ademais, políticas e leis com princípios básicos e padrões de prevenção de riscos e a exigência de boas práticas com a criação de códigos, orientações e controles internos preventivos também poderiam ser pensadas.

Especificamente em relação a este último ponto, já há mobilizações de grandes empresas de base tecnológica favoráveis a esse caminho. Não obstante, a despeito de a criação de regras e padrões de condutas internas ser uma característica da adoção do modelo de "risquificação", alguns cuidados devem ser observados a fim de que não se verifique uma utilização fragmentada e, consequentemente, ineficiente desse modelo regulatório.

---

[183] BRASIL. Autoridade Nacional de Proteção de Dados. Processo nº 00261.004509/2024-36. Relatora: Miriam Wimmer. Conselho Diretor, ANPD, 2024. Disponível em: https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-determina-suspensao-cautelar-do-tratamento-de-dados-pessoais-para-treinamento-da-ia-da-meta/SEI_0130047_Voto_11.pdf. Acesso em: 07 jul. 2024.

[184] CONSULTOR JURÍDICO. CNJ rejeita pedido para barrar uso de inteligência artificial no judiciário. Consultor Jurídico. 2024. Disponível em: https://www.conjur.com.br/2024-jul-03/cnj-rejeita-pedido-para-barrar-uso-de-inteligencia-artificial-no-judiciario/. Acesso em: 07 jul. 2024.

Empresas como o Facebook[185] e o Google[186], por uma alegada preocupação com a governança dos seus sistemas de decisão automatizada, estão elaborando códigos e padrões de conduta para a governança dos algoritmos. Esses documentos certamente por mais que sejam relevantes para que exista uma expectativa de uso dos algoritmos, não podem ser considerados suficientes para garantir a real proteção da pessoa humana.

> A auto-regulação no nível do estabelecimento de padrões da indústria começou a tomar forma, mas ainda está em fase de desenvolvimento. Uma vez concluídos, os padrões da indústria podem fornecer um veículo útil para co-regulação. No nível da intervenção estatal, considera-se possíveis papéis para: medidas de informação, por exemplo, alfabetização algorítmica pública; incentivos por meio de financiamento e impostos, como investimento estratégico para aumentar a pesquisa em métodos algorítmicos que sejam transparentes e responsáveis, bem como apoio técnico/infraestrutural para o jornalismo investigativo tecnológico; medidas legislativas; e um possível papel para um órgão regulador.[187]

O European Parlimentary Research Service desde 2019 apresentou estudo criando algumas opções de governança para o uso de algoritmos, buscando superar as dificuldades técnicas e regulatórias destes sistemas:

> Com base em nossa revisão e análise da literatura atual sobre transparência e responsabilidade algorítmica, e os sucessos, falhas e desafios de diferentes estruturas de governança que foram aplicadas a desenvolvimentos tecnológicos (especialmente em TIC), propomos um conjunto de quatro opções de política, cada uma das quais aborda um aspecto diferente da transparência e responsabilidade algorítmica: 1. Elevação da consciência: educação, observadores e denunciantes. 2. Responsabilidade no uso de decisões algorítmicas no setor público. 3. Supervisão regulatória e responsabilidade legal no setor privado. 4. Dimensão global da Governança algorítmica.[188]

Esse estudo, publicizado pelo Parlamento Europeu, apresenta importantes e profundas reflexões, facilitando uma compreensão dessa questão multifacetada e complexa, e poderia ser utilizado como, com as devidas ressalvas, para discussões da agenda nacional brasileira sobre o tema.

---

185 FACEBOOK. Why Am I Seeing This? We Have an Answer for You. Meta. 2019. Disponível em: https://news room.fb.com/news/2019/03/why-am-i-seeing-this/. Acesso em: 05 jan. 2024.

186 GOOGLE. Artificial Intelligence at Google: Our Principles, Google AI. 2017. Disponível em: https://ai. google/principles/. Acesso em: 05 jan. 2024.

187 EUROPEAN PARLIMENT RESEARCH SERVICE. A governance framework for algorithmic accountability and transparency. Scientific Foresight Unit, 2019. Disponível em: https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf. Acesso em: 10 out. 2023. (Tradução nossa).

188 *Ibidem* (Tradução nossa).

# 4 CONTESTANDO UMA IA

O desenvolvimento de decisões por meio de sistemas de Inteligência Artificial vem ocorrendo em um território sem qualquer tipo de regulamentação brasileira específica mais efetiva até o momento. Muitos dos "avanços" e testes são, inclusive, possíveis pelas oportunidades trazidas pelo desconhecimento do potencial e consequências deste novo modelo de economia:

> As principais empresas de tecnologia foram respeitadas e tratadas como emissários do futuro; nada na experiência passada havia preparado as pessoas para essas novas práticas, havendo, portanto, escassez de barreiras para que se protegessem; os indivíduos passaram rapidamente a depender das novas ferramentas de informação e comunicação como recursos necessários na luta cada vez mais estressante, competitiva e estratificada para uma vida mais eficaz; as novas ferramentas, redes, aplicativos, plataformas e mídias tornaram-se requisito para a participação social.[189]

Assim, as pessoas e instituições começaram a confiar e utilizar amplamente, sem muito questionamento ou resistência, as novas tecnologias de comunicação e informação para lidar com os seus desafios diários. Essas ferramentas se tornaram indispensáveis, ou ao menos o seu uso é apresentado como indispensável, para a participação ativa da vida social, seja no trabalho, na educação, no entretenimento ou em outras áreas. Em essência, as novas tecnologias não são mais apenas conveniências, mas sim componentes essenciais para a integração e participação na sociedade contemporânea. Exemplo simples é a ferramenta do WhatsApp, que no Brasil conta com 147 milhões de contas do app.[190]

E quando voltamos o nosso olhar para o enfoque deste trabalho, as decisões automatizadas tomadas por meio de sistemas de Inteligência Artificial, percebemos um reforço a este padrão de confiança, falta de questionamento e ausência de resistência, mesmo sem garantias legais específicas quanto aos danos que o uso destas ferramentas poderia trazer.

Conforme demonstrado, grandes centros de risco decorrem das decisões baseadas em Inteligência Artificial, as quais são tomadas muitas vezes a partir de dados pessoais originados de informações frequentemente aleatórias, heterogêneas e, muitas vezes, não classificadas. Os algoritmos utilizados nesses processos demandam intervenções humanas mínimas - ou não

---

189 BRUNO, Fernanda *et al*. **Tecnopolíticas da vigilância:** perspectivas da margem. São Paulo: Boitempo, 2018. p. 58.

190 SINCH ENGAGE. WhatsApp no Brasil: saiba vários dados do app número 1 do país. Sinch engage. 2023. Disponível em: https://engage.sinch.com/pt-br/blog/whatsapp-no-brasil/#:~:text=Depois%20da%20India%2C%20o%20Brasil,o%20WhatsApp%20todos%20os%20dias. Acesso em: 04 jan. 2024.

demandam -, resultando na ausência de subjetividades que possam ser verificadas e triadas[191], e muito menos contestadas, contribuindo para a fragilidade jurídica dos indivíduos impactados por tais decisões.

É dentro deste contexto de vulnerabilidade individual que o direito de contestar as decisões tomadas por algoritmos se apresenta como uma ferramenta necessária. E, conforme abordado nos capítulos anteriores, a possibilidade de contestação e vias de recurso eficazes contra as decisões tomadas por sistemas de Inteligência Artificial são requisitos essenciais na dimensão processual da equidade, princípio para uma *Trustworthy* AI.

Antes de adentrarmos mais a fundo especificamente na contestação de decisões automatizadas, devemos refletir sobre o que é o objeto da contestação. Como observa Antoinette Rouvroy, as vezes é tentador ver as ameaças potenciais aos direitos e liberdades fundamentais envolvidos nos sistemas de tomada de decisão por meio de Inteligência Artificial apenas em termos de possíveis erros das máquinas causados ou pelo fato de que os dados processados são falsos ou incompletos ou pela inadequação do sistema de modelagem (ex. erros baseados em suposições).

No entanto, o conceito de erros de máquinas pressupõe que haja uma "verdade objetiva" equivalente aos fatos em si, que necessariamente teriam sido encontrados se não fosse pelo erro – que presume se existir mecanismos para serem detectáveis e corrigíveis.

Assim, sustenta muitas vezes essa ideia o argumento de que sempre haverá uma correspondência perfeita entre os fatos, - o que o pensamento jurídico já percebeu desde o seu nascimento ser uma utopia.

Esse racional também ignora que os fatos não têm valor em si mesmos:

> Não buscar mais entender as causas dos fenômenos e tentar, em vez disso, fazer previsões em uma base puramente estatística e indutiva, em outras palavras, uma que ignora totalmente as causas, equivale a não buscar mais entender e mudar as circunstâncias que cercam os fatos.[192]

Dessa forma, Rouvroy observa que o uso do raciocínio algorítmico pode tornar desnecessário questionar os motivos pelos quais ocorrem determinada situação. E, assim, situações como bancos de dados obtidos dos empregadores de uma determinada região que possui mais mulheres ou pessoas designadas como pertencentes a um grupo étnico específico

---

[191] ROUVROY, Antoniette: BERNS, Thomas. Governamentalidade algotítmica e perspecticas de emancipação: o díspar como condição de individuação pela relação. *In:* BRUNO, Fernanda *et al.* **Tecnopolíticas da vigilância**: perspectivas da margem. São Paulo: Boitempo, 2018.

[192] ROUVROY, Antoinette. **Of Data and Men**: Fundamental Rights and Freedoms in a World of Big Data. Council of europeu. Strasbourg, 2016. Disponível em: https://rm.coe.int/16806a6020. Acesso em: 03 jan. 2024. p. 32 (Tradução nossa).

entre os funcionários forçados a deixar a força de trabalho precocemente ou não sejam promovidos, a causa dessa situação/fato (que podem ser tendências discriminatórias na sociedade) será ocultada quando este procedimento for usado para desenvolver perfis de empregabilidade, os quais são concedidos o status de meros fatos objetivos.

> A dedução prática ou recomendação automatizada para os gestores de recursos humanos será a seguinte: "pessoas pertencentes a certos grupos étnicos e mulheres são estatisticamente menos bem-sucedidas profissionalmente", mas a discriminação, que nem sempre é necessariamente formalmente identificada como tal porque nem todos tomam medidas legais por discriminação ilegal, mas que é a circunstância subjacente a este "fato", não será mais percebida como um problema.[193]

Os sistemas de tomada de decisão por meio de Inteligência Artificial muitas vezes reivindicam uma forma de objetividade mecânica baseada na automação dos sistemas de processamento de dados, com desconsideração ao conhecimento das circunstâncias, do contexto e das causas dos fenômenos, e assim incorporam visões de mundo específicas sem perceber as “subjetividades” do modelo. E aqui podemos ter visões de mundo que toleram discriminação com base no uso de determinados dados pessoais.

Mas é claro que a opacidade dos processos algorítmicos e algumas proteções – como os segredos comerciais e industriais - tornam a discriminação muito difícil de provar e, assim, contestar.

Até mesmo do lado das instituições que utilizam sistemas de Inteligência Artificial para tomada de decisão, a alegação de uma discriminação e uma consequente contestação muitas vezes não é sequer cogitada. É comum que o uso do sistema tenha exatamente a intenção contraria, a intenção de usar um sistema automáticos para tornar suas próprias decisões mais objetivas.

> A discriminação indireta resultante do funcionamento de um sistema automatizado de recomendação não decorre tanto da pessoa que decide seguir a recomendação (na verdade, pode-se dizer que a vontade dessa pessoa de tornar suas decisões mais objetivas reflete um desejo de anular seus próprios preconceitos) quanto da existência prévia na sociedade de uma mentalidade discriminatória (um "apetite" por ou aceitação da discriminação) variando em escopo, mas refletida passivamente em conjuntos de dados e, portanto, adquirindo o status de um fato objetivo, apolítico, neutro e não problemático.[194]

---

[193] *Ibidem*, p. 32-33 (Tradução nossa).

[194] *Ibidem*, p. 33 (Tradução nossa).

Mas o foco de regulação não pode ser no campo da intenção das ações, mas nos resultados práticos, no mundo concreto que são por elas moldado.

## 4.1 A AUTODETERMINAÇÃO INFORMATIVA

Os impactos que podem decorrer do uso da Inteligência Artificial e do aprendizado de máquinas, a partir da aplicação de algoritmos nas tomadas de decisão, envolvendo resultados que vão além do direcionamento humano, são capazes de repercutir de forma perigosa sobre diversos direitos fundamentais, tradicionais ou emergentes.

A sociedade informacional veio a trazer alguns "novos" conceitos no que se refere aos direitos fundamentais, falando-se hoje, por exemplo, em digitalização dos direitos fundamentais[195]. Os direitos fundamentais passam pelo imperativo de serem tutelados também no que se refere aos novos riscos decorrentes do uso da tecnologia que vem afetando direitos como o da liberdade, igualdade, inviolabilidade do direito à vida, à privacidade, à segurança e à propriedade.

Este é precisamente o caso do direito à autodeterminação informativa, que apesar de derivado do princípio da dignidade da pessoa humana, do livre desenvolvimento da personalidade e, mais especificamente, da proteção à intimidade, é autônomo e com estes já não se confunde[196].

O conceito de autodeterminação informativa tem sua origem na decisão do Tribunal Constitucional Federal da Alemanha, em 1983, conhecida como BverfGE 65,1. A questão em destaque era a constitucionalidade de uma lei que ordenava o recenseamento geral da população, com a coleta de dados sobre profissão, moradia e local de trabalho para fins estatísticos.

O tribunal teve que analisar se essa lei violava o direito geral da personalidade, em relação à proteção do indivíduo contra o levantamento, armazenamento, uso e transmissão irrestritos de seus dados pessoais. A decisão se debruçou sobre os riscos para a personalidade que poderiam surgir do processamento eletrônico de dados pessoais da população alemã. Nessa decisão, foi forjado o conceito de autodeterminação informativa, estabelecendo princípios para

---

[195] PIAIA, Thami Covatti. A digitalização dos direitos fundamentais. **Revista de Direitos e Garantias Fundamentais**, [S. l.], v. 22, n. 2, p. 7–8, 2022. DOI: 10.18759/rdgf.v22i2.2079. Disponível em: https://sisbib.emnuvens.com.br/direitosegarantias/article/view/2079. Acesso em: 9 jan. 2024.

[196] COPEETI, Rafael; MIRANDA, Marcel Andreata de. Autodeterminação Informativa e Proteção de Dados: Uma Análise Crítica da Jurisprudência Brasileira. **Revista de Direito, Governança e Novas Tecnologias**. Minas Gerais, v. 1, n. 2, p. 28-48, jul./dez. 2015. Disponível em: https://indexlaw.org/index.php/revistadgnt/article/view/46. Acesso em: 10 jan. 2024.

proteger o indivíduo contra o levantamento, armazenamento, uso e transmissão de dados pessoais sem sua autorização, a menos que haja um interesse predominante da coletividade (não sendo, assim, um direito ilimitado).

Em resumo, a autodeterminação informativa pode ser vista como a faculdade que o indivíduo possui de determinar e controlar a utilização e tratamento que são exercidos sobre os seus dados pessoais. E a relevância deste direito à autodeterminação sobre a sua informação tem se evidenciado em razão da necessidade de se preservar o indivíduo frente aos novos perigos informáticos, especialmente os referentes à constante coleta de dados pessoais.

Neste conceito a possibilidade de determinar quais informações pessoais podem ser conhecidas por terceiros, a exigência de que a utilização dos dados pessoais siga a finalidade comunicada no momento da coleta, a publicidade dos bancos de dados, a exatidão dos dados, o livre acesso dos indivíduos aos dados armazenados para fiscalização e retificação, e a necessidade de medidas para evitar riscos de extravio, destruição, modificação, transmissão e acesso não autorizado.

Na mencionada decisão do Tribunal Alemã, foram ponderadas questões de como o processamento automático dos dados pessoais poderia representar uma ameaça ao poder do indivíduo de decidir por conta própria o se e o como ele desejaria eventualmente tornar públicos seus dados "no sentido de que o processamento de dados possibilitaria a elaboração de um "quadro completo da personalidade" por meio de "sistemas integrados sem que o interessado possa controlar o suficiente sua correção e aplicação"[197].

Ademais, em se tratando de um poder de escolha e controle do indivíduo, no fundo se está igualmente falando da preservação da autonomia existencial deste[198], na possibilidade de o indivíduo ter controle sobre como a sua vida e da sua vida e personalidade serem desenvolvidas sem mecanismos que mitiguem de forma obscura e não justificada a sua liberdade.

> A autonomia existencial, portanto, se identifica com a liberdade do sujeito em gerir sua vida, sua personalidade, de forma digna. É nesse ponto que se encontram questões delicadas como o uso ativo dos direitos da personalidade e as discussões sobre o direito à morte digna, eutanásia, aborto, pena de morte,

---

[197] MENDES, Laura Schertel. Habeas data e autodeterminação informativa. Os dois lados de uma mesma moeda. **Direitos Fundamentais & Justiça**, [s.l.], v. 39, 2018. p. 188.

[198] SANT´ANA, Maurício Requião. **Autonomia, incapacidade e transtorno mental**: propostas pela promoção da dignidade. 2015. Dissertação (Doutorado em Direito) – Faculdade de Direito, Universidade Federal da Bahia, Salvador, 2015. Disponível em: https://repositorio.ufba.br/bitstream/ri/17254/1/Tese%20Maur%C3%ADcio%20Requi%C3%A3o.pdf. Acesso em: 04 jan. 2024.

> manipulação de embriões, direitos pessoais de família, sexualidade e identidade de gênero.[199]

Para Maurício Requião, ao analisar o tema da privacidade e a sua relação com a promoção e garantia da autonomia do incapaz, o direito fundamental da privacidade é um elemento essencial quando se fala na garantia da autonomia do indivíduo. E não há como se preservar essa autonomia, que envolve poder impedir que terceiros interfiram na espera física, psíquica ou social[200], sem que o indivíduo tenha a possibilidade de ter o mínimo de controle sobre as suas informações e como elas são usadas pelos sistemas de Inteligência Artificial aos quais servem de alimento.

Neste mesmo sentido, Laura Schertel observa sobre o desenho da autodeterminação informativa pelo Tribunal Alemão:

> Assim, aumentaria a influência do Estado sobre o comportamento do indivíduo, que não mais seria capaz de tomar decisões livres em virtude "da pressão psíquica da participação pública". Uma sociedade, "na qual os cidadãos não mais são capazes de saber quem sabe o que sobre eles, quando e em que situação", seria contrária ao direito à autodeterminação informativa, o que prejudicaria tanto a personalidade quanto o bem comum de uma sociedade democrática.[201]

Essencial assim, que a possibilidade de contestar o uso dos dados pessoais por meio de sistemas de tomada de decisão por meio de Inteligência Artificial seja visto também sobre essa garantia da autodeterminação do indivíduo sobre as suas informações.

## 4.2 O DEVIDO PROCESSO LEGAL NAS RELAÇÕES PRIVADAS

Danielle Keats Citron e Frank Pasquale, ao tratar dos impactos que a prática de score de crédito social traz para os sujeitos submetidos a este tipo de decisão, sinalizam que o devido processo legal é essencial para aqueles estigmatizados por sistemas de pontuação "inteligentes" artificialmente e que a tradição processual americana de devido processo legal deve ser utilizada para informar salvaguardas básicas.

---

199 SANT´ANA, Maurício Requião. **Estatuto da pessoa com deficiência, incapacidades e interdição**. Salvador: Juspodivm, 2016. p, 31.

200 SANT´ANA, Maurício Requião. **Autonomia, incapacidade e transtorno mental**: propostas pela promoção da dignidade. 2015. Dissertação (Doutorado em Direito) – Faculdade de Direito, Universidade Federal da Bahia, Salvador, 2015. Disponível em: https://repositorio.ufba.br/bitstream/ri/17254/1/Tese%20Maur%C3%ADcio%20Requi%C3%A3o.pdf. Acesso em: 04 jan. 2024.

201 MENDES, Laura Schertel. Habeas data e autodeterminação informativa. Os dois lados de uma mesma moeda. **Direitos Fundamentais & Justiça**, [s.l.], v. 39, 2018. p. 188.

> Os reguladores devem ser capazes de testar sistemas de pontuação para garantir sua justiça e precisão. Indivíduos devem ter oportunidades significativas de contestar decisões adversas baseadas em pontuações que os categorizam erroneamente. Sem tais proteções em vigor, sistemas poderiam transformar dados tendenciosos e arbitrários em pontuações fortemente estigmatizantes.[202]

Assim, os referidos autores, ao se proporem a analisar o devido processo de decisões automatizadas de crédito social, observam que ao utilizarem *big data* para a classificação e avaliação de indivíduos, os algoritmos preditivos avaliam, com base nos dados pessoais apresentados ou tomados do indivíduo, se o sujeito possui um alto ou baixo riscos de crédito. Analisam, por exemplo, se o sujeito é funcionário desejável, um inquilino confiável, ou um cliente considerado valioso – ou, por outro lado, se o sujeito da análise é um caloteiro ou representa uma ameaças ou é irrelevante.

Com isso, oportunidades importantes para estes sujeitos estão em jogo e, infelizmente, "em uma área onde a regulamentação prevalece - crédito - a lei se concentra no histórico de crédito, não na derivação das pontuações a partir de dados"[203].

Neste sentido, os referidos autores retratam como a *Federal Trade Commission* (FTC), autarquia americana que cuida das práticas anticoncorrenciais, deve ter acesso aos sistemas de pontuação de crédito e outros sistemas de pontuação que prejudiquem injustamente os consumidores e devem poder testar estes sistemas quanto aos vieses, arbitrariedade e caracterizações injustas.

E, para isso, mais uma vez, a transparência entra em jogo, já que a FTC precisaria visualizar não apenas os conjuntos de dados explorados pelos sistemas de pontuação, mas também o código-fonte e as notas dos programadores descrevendo as variáveis, correlações e inferências incorporadas nos algoritmos dos sistemas de pontuação.

Idealmente, os sistemas de pontuação deveriam ser executados por meio de conjuntos de testes que executam cenários hipotéticos - esperados e inesperados. Os testes refletiriam a norma do desenvolvimento de algoritmo de tomada de decisão adequado e ajudariam a detectar o viés potencial dos programadores ou o viés emergente da evolução do sistema de IA. Mas neste contexto, os indivíduos objetos da decisão podem fazer pouco ou nada para se protegerem.

> Uma vez que a FTC avalie sistemas de pontuação de crédito para detectar "arbitrariedade por algoritmo" — como a presidente Ramirez astutamente coloca — ela deveria emitir uma Avaliação de Impacto na Privacidade e

---

[202] CITRON, Danielle Keats; PASQUALE, Frank A. The Scored Society: Due Process for Automated Predictions. **Washington Law Review**, n. 1, 2014. p. 2-27. Disponível em: https://digitalcommons.law.umaryland.edu/fac_pubs/1431/. Acesso em: 04 fev. 2024. p. 1 (Tradução nossa).

[203] *Ibidem*, p. 1.

> Liberdades Civis avaliando o impacto negativo e desproporcional de um sistema de pontuação em grupos protegidos, resultados arbitrários, má caracterizações e danos à privacidade. Nessas avaliações, a FTC poderia identificar medidas apropriadas de mitigação de riscos.[204]

Questão importante é quando migramos o debate da transparência frente a agências como a FTC e voltamos isso para o público afetado, o quanto este deve ter acesso aos conjuntos de dados e lógica dos sistemas de pontuação de crédito preditivos. Para Danielle Keats Citron e Frank Pasquale, cada titular de dados deve ter acesso a todos os dados relacionados a ele – o que vai em linha com o que é disposto na própria LGPD dentro dos direitos dos titulares.

Também aqui argumentos como a proteção ao segredo comercial tentam reforçar a concepção binária entre transparência e inovação. Contudo, há poucas evidências de que a incapacidade de manter esses sistemas em segredo diminuiria a inovação e também não há evidências adequadas para dar credibilidade a preocupações de que, uma vez que o sistema seja público, os indivíduos encontrarão maneiras de manipulá-lo.

> Embora a manipulação seja uma preocupação real em contextos online, onde, por exemplo, um otimizador de motores de busca poderia criar fazendas de links para manipular o Google ou outros algoritmos de classificação se os sinais se tornassem públicos, os sinais usados na avaliação de crédito são muito mais caros para fabricar. Além disso, a verdadeira base do sucesso comercial em indústrias impulsionadas por "grandes dados" provavelmente é a quantidade de dados relevantes coletados no agregado — algo não necessariamente revelado ou compartilhado por meio da divulgação pessoa a pessoa dos dados mantidos e dos algoritmos de pontuação usados.[205]

Evidentemente que pode existir legitimidade na argumentação sobre o segredo comercial e sigilos relativos aos bancos de dados, contudo ameaças à dignidade humana apresentadas por sistemas de pontuação secretos, automatizados e generalizados devem ter um peso maior na grande maioria dos casos – especialmente quando os outros interesses em jogo giram em torno apenas de interesses privados de empresas focadas em lucrar mais. E até para este argumento já existem soluções técnicas, como os *acordos criptográficos,* que serão apresentados mais à frente.

É assim que minimamente os indivíduos devem ter um aviso prévio e a chance de contestar pontuações preditivas que prejudiquem sua capacidade de obter crédito, emprego,

---

[204] CITRON, Danielle Keats; PASQUALE, Frank A. The Scored Society: Due Process for Automated Predictions. **Washington Law Review**, n. 1, 2014, p. 2-27. Disponível em: https://digitalcommons.law.umaryland.edu/fac_pubs/1431/. Acesso em: 04 fev. 2024. p. 25-26 (Tradução nossa).

[205] CITRON, Danielle Keats; PASQUALE, Frank A. The Scored Society: Due Process for Automated Predictions. **Washington Law Review**, n. 1, 2014, p. 2-27. Disponível em: https://digitalcommons.law.umaryland.edu/fac_pubs/1431/. Acesso em: 04 fev. 2024. p. 26 (Tradução nossa).

moradia ou outras oportunidades importantes. E, para isso, o devido processo legal se mostra como ferramenta adequada.

Para Margot Kaminski e Jennifer Urban[206], o racional do devido processo legal consiste comumente na obtenção de precisão dentro dos valores que norteiam o Estado Liberal, na teoria que enfatiza a importância do indivíduo afetado por determinada decisão de ter respostas e contestar.

A garantia da precisão, como motivo para a existência de um devido processo legal é uma visão instrumentalista. A função assim do devido processo legal neste prisma seria diminuir os riscos de decisões erradas. Contudo, comumente a precisão é vista apenas como mais um objetivo ou valor a ser equilibrado com tantos outros no momento de tomada de decisão, como por exemplo, a eficiência do sistema ou os custos.

> Alguns acadêmicos, portanto, consideraram a precisão como um objetivo de gestão sistemática, significando que, na medida em que o devido processo individual pode, no agregado, tornar um sistema decisório mais preciso, vale a pena protegê-lo. Se essa fosse a única razão para proteger o processo individual, no entanto, o processo individual não valeria a pena ser protegido na medida em que falhasse em tornar um sistema decisório mais preciso.[207]

Portanto, a precisão por si só não consegue explicar a percepção do que as pessoas têm sobre o que é ou não justo. Assim, uma segunda razão para a existência de um devido processo legal com direito de contestar é trazida pelos mencionados autores, os valores do Estado de Direito.

Estes valores preconizam que um sistema de tomada de decisão deve ser sempre justo, consistente, previsível e racional quando aplicado em diferentes indivíduos. Para prevenir arbitrariedades, identificar preconceitos e controlar de alguma maneira as decisões, as proteções processuais relacionadas a contestação trazem a obrigação do tomador de decisão de demonstrar um compromisso que seja examinável com um determinado resultado, apresentando as razões, os motivos para isso: "a contestação pode ajudar a eliminar as armadilhas da discrição decisória, ao mesmo tempo que deixa aos tomadores de decisão a capacidade de adaptar regras às circunstâncias individuais."[208]

---

[206] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024.

[207] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024. p. 1990-1991 (Tradução nossa).

[208] *Ibidem*, p. 1991 (Tradução nossa).

Ou seja, permitir que os indivíduos afetados pelas decisões tomadas por meio de sistemas de IA, ou mesmo decisões humanas que dependem significativamente de ferramentas de IA, contestem estas decisões, revela se um sistema decisório é injusto, inconsistente, arbitrário, imprevisível ou irracional. Contudo, o foco aqui de preocupação é a legitimidade do sistema de tomada de decisão, não necessariamente os indivíduos.

Dessa forma, é possível se argumentar que também sobre essa razão, que os direitos individuais de contestação são necessários apenas na medida em que revelam injustiça, arbitrariedade, imprevisibilidade e irracionalidade. E é aqui que entra a terceira razão, a da "teoria liberal"[209], teoria fundamentada nos indivíduos.

Na concepção liberal americana, três teorias poderiam ser utilizadas para justificar a necessidade de um devido processo individual, as utilitaristas, as lockeanas de direitos naturais e as kantianas. A primeira, utilitarista, pode ser desenvolvida na medida em que através da análise de custo-benefício, o devido processo individual produz aceitação ou até mesmo felicidade por parte dos indivíduos afetados, o que torna mais provável que o sistema decisório como um todo alcance os objetivos de bem-estar social (e isso evitar dor ou outras consequências negativas graves).[210]

Do ponto de vista das teorias de direitos naturais, o devido processo individual é meio de proteção de direitos naturais subjacentes. Se algum direito individual inerente ao indivíduo, como o direito à vida, à liberdade ou à propriedade, for ser restringido, isso deve ser feito por meio de um processo adequado. Aqui o devido processo ajudaria a mitigar erros e a evidenciar a aceitação individual, como o consentimento.

E, por fim, a teoria kantiana traz a perspectiva, por meio do imperativo categórico kantiano, que estabelece que os indivíduos devem ser tratados como um fim em si mesmos, e não como um meio para um fim. Assim, um direito de contestar decisões expressa esse valor, uma vez que usar categorias de aproximação para tomar decisões sobre os indivíduos os trata como objetos. Aqui teríamos um argumento dignatário para o devido processo individual.

> Uma abordagem Kantiana argumenta que os direitos de devido processo são necessários para respeitar a individualidade do ser. Decisões opacas ou arbitrárias interferem fundamentalmente nesse auto-respeito. Oferecer uma razão para uma decisão, por outro lado, mostra um sinal de respeito pelo

[209] MASHAW, Jerry L. Administrative Due Process: The Quest for a Dignitary Theory. **Boston University Law Review**, v. 61, n. 885, 1981. Disponível em: https://core.ac.uk/download/pdf/72827487.pdf. Acesso em: 04 fev. 2024.

[210] MASHAW, Jerry L. Administrative Due Process: The Quest for a Dignitary Theory. **Boston University Law Review**, v. 61, n. 885, 1981. Disponível em: https://core.ac.uk/download/pdf/72827487.pdf. Acesso em: 04 fev. 2024.

> sujeito da decisão. Da mesma forma, possibilitar a participação na legitimidade do sistema ao estabelecer um direito de contestar seus resultados. Quando uma decisão afeta certos direitos fundamentais, o processo é necessário tanto como um meio de promover a precisão no que diz respeito à privação de direitos quanto como um fim para possibilitar o auto-respeito.[211]

Este histórico legal, assim, serve como exemplo de como em diferentes sistemas legais o devido processo se desenvolveu. Contudo, voltando ao foco desta pesquisa, Danielle Keats Citron em 2008 já havia publicado a obra "*Technological Due Process*"[212], ressaltando a importância de equilibrar automação com a discrição humana. para evitar a perda do discernimento em prol da eficiência.

O conceito desenvolvido do "Devido Processo Tecnológico" é a proposta de um novo modelo de proteção dos direitos individuais em um ambiente administrativo cada vez mais automatizado. Esse modelo busca garantir a transparência, a prestação de contas e a equidade nas decisões automatizadas, sem abrir mão dos benefícios oferecidos pelos sistemas de decisão computadorizados. Ele visa proteger os interesses das pessoas em processos de adjudicação e formulação de políticas, buscando assegurar que as decisões sejam justas, consistentes e transparentes, mesmo em um cenário dominado pela tecnologia.

São reconhecidas as ameaças ao devido processo enfrentadas por indivíduos em situações de risco devido à automação e à tecnologia, especialmente considerando que estas decisões vêm sendo utilizadas para privar pessoas de direitos fundamentais, como o direito à vida, à propriedade ou à liberdade.

O que não faltam são exemplos. É o caso o seu uso na segurança policial preditiva[213], com algoritmos para identificar e direcionar indivíduos considerados "propensos ao crime", que traz preocupações sobre a criminalização de comunidades marginalizadas e a erosão da liberdade individual. Ou falhas em algoritmos de saúde, usados para diagnosticar doenças ou determinar a elegibilidade de um indivíduo para determinado tratamento. Estes sistemas podem apresentar falhas que impactam negativamente a saúde e o bem-estar das pessoas.[214]

---

[211] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024. p. 1994-1995.

[212] CITRON, Danielle Keats. Technological Due Process. **Whashigton University Law Review**, v. 85, n. 6, 2008. Disponível em: https://openscholarship.wustl.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1166&context=law_lawreview. Acesso em: 10 out. 2023.

[213] XAVIER, Paulo Ramón Suárez *et al*. Polícia preditiva e "negritude": modelos para a reprodução de um estado sem direitos. Direito. **Revista da Faculdade de Direito,** v. 6, n. 3, p. 99–128. Universidade de Brasília, 2022. Disponível em: https://periodicos.unb.br/index.php/revistadedireitounb/article/view/36383. Acesso em: 09 mar. 2024.

[214] DOURADO, Daniel de Araújo; AITH, Fernando Mussa Abujamra. A regulação da inteligência artificial na saúde no Brasil começa com a Lei Geral de Proteção de Dados Pessoais. **Revista Saúde Pública**, [s.l.], v. 56,

Aqui, dois pontos chaves do devido processo dever ser abordados no detalhe, a notificação e a oportunidade do sujeito afetado ser ouvido. A notificação adequada deve ser vista como um direito fundamental no devido processo, pois é ela que irá garantir que os sujeitos sejam informados sobre as ações pretendidas por uma agência / instituição, permitindo-lhes compreender e responder adequadamente a essas ações.

A falta de notificação adequada pode comprometer a capacidade dos indivíduos de se defenderem e participarem efetivamente do processo decisório, minando assim a equidade nas decisões automatizadas, como veremos mais afundo no direito à contestação. E sem transparência e conhecimento sobre as decisões, não há notificação adequada.

E a oportunidade de ser ouvido, por sua vez, é a oportunidade do sujeito, objeto da decisão, de apresentar suas perspectivas, suas evidências e os seus argumentos antes que uma decisão seja tomada de forma irreversível.

A oportunidade de ser ouvido é, assim, outro princípio fundamental do devido processo legal – aplicável ao devido processo em questões tecnológicas - que garante que as partes afetadas tenham a chance de apresentar suas perspectivas, evidências e argumentos, antes que uma decisão final seja tomada.

Esse princípio visa assegurar que os indivíduos tenham a oportunidade de participar ativamente do processo decisório, contribuindo com informações relevantes e defendendo seus interesses. E quando trazemos este princípio para a tomada de decisão por meio de sistemas de Inteligência Artificial, a oportunidade de ser ouvido assume uma importância ainda maior, uma vez que estas decisões estão sendo utilizada de forma a afetar cada vez mais significativamente os direitos e interesses dos titulares, sem a devida consideração de suas circunstâncias individuais.

Assim, é crucial que os sistemas automatizados incorporem mecanismos que permitam aos indivíduos afetados apresentarem suas visões, contestar decisões e ter acesso a processos de revisão justos e imparciais.

A oportunidade de ser ouvido não apenas fortalece a legitimidade e a justiça dos processos decisórios automatizados, mas também promove a transparência e a equidade nas decisões e possibilita mitigar possíveis vieses ou erros decorrentes da automação.

Por fim, este princípio garante o direito dos indivíduos de serem informados sobre decisões automatizadas que os afetam e de terem a chance de contestar esses resultados.

---

n. 80. 2022. Disponível em: https://www.scielo.br/j/rsp/a/k38jGvJdbQSYN4MpzGZpfXw/?format=pdf&lang=pt. Acesso em: 3 mar. 2024.

Contudo, alguns desafios devem ser ressaltados, como o da rapidez das decisões, uma vez que os sistemas automatizados podem tomar decisões rapidamente, sem oferecer tempo suficiente para as pessoas entenderem o que aconteceu e se defenderem. A falta de mecanismos de contestação ou clareza sobre eles é ponto a ser combatido. E se durante a tomada de decisão não há tempo hábil de contestação, que as informações e conhecimentos sejam oferecidos de forma prévia.

Vale observar, ademais, que ao tratar sobre técnicas de "desenviesamento" para a minimização da ocorrência de vieses cognitivos, Maurício Requião[215], citando Dierle Nunes, menciona que o contraditório e o devido processo legal funcionam como estratégias de "desenviesamento" para mitigar os vieses cognitivos nas decisões judiciais. Este pensamento poderia de forma análoga ser utilizado como argumento para o "desenviesamente" de decisões tomadas por meio de sistemas de Inteligência Artificial.

## 4.3 ARQUÉTIPOS DE CONTESTAÇÃO DA IA

O devido processo deve incluir além de uma notificação e uma oportunidade de ser ouvido, que isto ocorra frente a uma terceira parte neutra. Mas quando voltamos esse conceito para decisões por meio de Inteligência Artificial, especialmente em razão da velocidade e escala, bem como questões técnicas, um devido processo judicial deveria ser visto como última *ratio*, uma vez que este, no modelo de prestação jurisdicional que ocorre hoje em dia, representaria altos custos financeiros e temporais.

Dessa forma, este trabalho propõe um escalonamento deste processo de contestação, que deveria se iniciar dentro da organização responsável pela decisão (o indivíduo poderia solicitar uma revisão humana da decisão automatizada e maiores informações sobre a decisão), na ausência de solução satisfatória poder ter uma etapa prévia ao judiciário, com instituições / autoridades / autarquias setoriais específicas decidindo sobre o que se contesta e, em seu último degrau, em especial respeito ao princípio da inafastabilidade da jurisdição, que tem previsão no artigo 5º, inciso XXXV, da Constituição Federal vigente no Brasil, o Poder Judiciário poderia ser chamado a decidir sobre se houve ou não lesão ou ameaça a direito.

[215] REQUIÃO, Maurício. (no prelo). Inteligência artificial, vieses cognitivos e decisões judiciais.

E, para construir como estas camadas deveriam ser desenvolvidas, mas sem que se tenha a pretensão de criar um modelo único, a ideia de se trabalhar com arquétipos ou modelos de mecanismos de contestação, parece ser sedutora.

A distinção entre dois “modelos de mecanismos de contestação”, pode ser feita entre "regra de contestação" e entre *standards*, "padrão de contestação". Estes modelos oferecem vantagens ou desvantagens diferentes. As regras possuem a vantagem de serem mais precisas e oferecerem uma maior previsibilidade de direcionamento, enquanto padrões podem ter a vantagem de evitar situações de sub-inclusão ou excesso inclusivo. Além disso, padrões também podem estimular a deliberação moral, embora deixem margem para interpretações tendenciosas.

Uma regra de contestação, dessa forma, se assemelha a uma regra legal ao estabelecer detalhes específicos de maneira prévia, incluindo requisitos de notificação, prazos para reclamações e respostas, e procedimentos formais para contestações.

Por outro lado, um padrão de contestação simplesmente afirma a existência do direito de contestar, deixando os detalhes procedimentais para decisores futuros. Este oferece flexibilidade para se adaptar a circunstâncias específicas, permitindo que outros atores preencham lacunas nos procedimentos ao longo do tempo, o que pode ser valioso em um cenário de constante mutação.

Um padrão de contestação diria, por exemplo, que os indivíduos têm direito de contestar uma decisão automática tomada por meio de sistemas de Inteligência Artificial, mas não detalharia como esse direito seria exercido na prática. Se nessa previsão existissem determinações como a de que a contestação deve ser respondida por determinado meio ou formato ou dentro de um prazo específico, já se estaria diante de uma regra de contestação.

> Uma regra de contestação, ou na verdade, um conjunto de regras de contestação, poderia ditar não apenas a existência de um direito de contestar, mas seus detalhes granulares: se, quando e como o aviso deve ser concedido; como as decisões devem ser tomadas; por quem; e em que cronograma. Por exemplo, uma lei poderia declarar: "Os indivíduos devem ser notificados quanto a uma decisão adversa dentro de 5 dias úteis, usando o seguinte formato, e os desafios devem ser ouvidos perante um árbitro neutro dentro de 10 dias úteis, com os indivíduos tendo os seguintes direitos processuais. [216]

Já a regra de contestação tem a vantagem de uma maior clareza, o que facilita a compreensão sobre o seu direcionamento e, teoricamente, facilita a conformidade com ela,

---

[216] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024. p. 2006 (Tradução nossa).

inclusive com menores custos associados. Observa-se, contudo, que no que se refere a estes custos, essa vantagem não passa de uma suposição, pois apesar de fazer sentido o raciocínio de que a clareza auxilia na conformidade e, consequentemente, em custos com a não conformidade, é possível que a regra estabelece algo muito custoso a ser feito, caindo por terra referida vantagem.

A uniformidade de aplicação de uma regra de contestação também é maior do que a de um padrão de contestação. Com isso, haveria menor margem de manobra e o mesmo processo seria concedido para diferentes sujeitos. O ruído seria assim minimizado.

Outra característica chave para se pensar no design do modelo de contestação de decisão automática tomada por meio de sistemas de Inteligência Artificial é a influência que o direito substancial tem nele. Em um modelo de contestação dependente do direito substantivo, o devido processo estabeleceria não apenas como a contestação deve ocorrer (de forma procedimental e mecânica), mas também os critérios para uma contestação de decisão.

Um exemplo desse modelo seria "o sujeito tem o direito de contestar a decisão que tenha utilizado algum dado pessoal sensível seu de forma discriminatória". Assim, no Brasil, sempre que se estivesse diante de sistemas de Inteligência Artificial que tomam decisão automatizada e que fizeram uso de dados pessoais sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, a decisão poderia ser contestada.

A escolha por um design mais substantivo ou mais procedimental, assim, tem consequências práticas:

> O benefício de um direito à contestação que é mais focado no procedimento é que a lei substantiva pode ser sutil, altamente dependente de fatos e complicada. A expertise em lei substantiva pode, portanto, ser muito cara, o que pode impedir o acesso individual à justiça. Adicionalmente, como a teoria do devido processo discutida acima argumenta, o processo pode importar por si só. O processo por si só pode proporcionar transparência, revelar problemas na tomada de decisão, dar aos indivíduos agência em uma decisão e tornar a tomada de decisão responsável, mesmo que as normas substantivas subjacentes não sejam declaradas.[217]

Pode ser caro para um sujeito provar o uso discriminatório do seu dado pessoal sensível e com isso poder contestá-lo, como no exemplo da contestação com elemento substancial

[217] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024. (Tradução nossa).

trazida. E, de fato, a mera existência de um procedimento de contestação pode já trazer outros benefícios, como a possibilidade já mencionada de trazer maior transparência, revelar problemas na tomada de decisão e até mesmo tornar a tomada de decisão mais responsável.

No entanto, Margot Kaminski e Jennifer Urban[218] sinalizam que existem desvantagens em um direito de contestação que seja altamente procedimental e minimamente substantivo, como a arbitrariedade de se poder contestar qualquer decisão automatizada por meio de sistemas de Inteligência Artificial sem uma fundamentação substantiva. O que poderia inclusive trazer custos altos ao se ter um volume excessivamente (e sem fundamentação) alto de contestações. E, considerando a realidade de determinadas instituições, essa poderia ser até mesmo uma estratégia de empresas maiores para eliminar concorrentes menores, abusando da sua posição de poder econômico.

Aplicando estas distinções, o direito a contestação trazido pela RGPD, - que estabelece que os indivíduos têm o direito de não ficarem sujeitos a uma decisão baseada unicamente no processamento automatizado, incluindo a criação de perfis, se essa decisão produzir efeitos jurídicos sobre eles ou os afetar significativamente de maneira similar -, pode parecer procedimental, mas, dentro do contexto, tem componentes substantivos.

Para trazer alguns arquétipos de contestação extrajudicial para essa realidade de decisões em velocidade e escala, Margot Kaminski e Jennifer Urban[219] propõe a análise de alguns modelos de outros "sistemas".

O primeiro arquétipo é representado pelo direito de contestação previsto na RGPD da União Europeia que tem como foco criar padrões para a contestação com foco em procedimentos a serem seguidos, em vez de regras específicas a serem aplicadas. A legislação de proteção de dados pessoais europeia, assim, não fornece, por si só, bases substanciais para contestar decisões algorítmicas. É necessário olhar para outras áreas do direito substantivo ou em outra parte do RGPD, ex. direito que os titulares de dados têm sobre a transparência no uso dos seus dados pessoais e o aviso sobre os dados terem sido submetidos a uma decisão automatizada para se "completar" tal direito.

Outro direito que também possibilita de forma mais substancial o direito a contestação previsto na RGPD é o caso do direito de acesso, também previsto na LGPD, que garante ao titular o acesso facilitado sobre a forma de tratamento, o uso que é feito dos dados e a integralidade dos dados pessoais. Dados incorretos utilizados em sistemas de tomada de decisão

---

[218] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024. p. 2006.

[219] *Ibidem*.

de Inteligência Artificial poderiam assim serem identificado e, ao conjugar estes direitos com o de correção de dados incorretos, previsto na RGPD e na LGPD, o titular poderia exigir a correção da decisão.

> Vários outros direitos geralmente aplicáveis do GDPR, como o direito ao esquecimento, direito de objeção e direito de restringir o processamento, podem cada um ser entendido alternativamente como (1) complementos ao direito de contestação, (2) requisitos mínimos independentemente da força do direito de contestação, ou talvez (3) como modelos para o direito de contestação.[220]

É possível se argumentar inclusive, que haveria, de certa forma, uma relação deste direito com o princípio do ser humano como centro, apresentado anteriormente como um dos princípios para uma *trustworth* AI. Isso porque a RGPD trata da dignidade da pessoa humana e estabelece mecanismos para sempre haver, em alguma medida, um envolvimento humano nas tomadas de decisão com efeitos significativos.

O art. 22 da RGPD assim estabelece que:

> 1. O titular dos dados tem o direito de não ficar sujeito a nenhuma decisão tomada exclusivamente com base no tratamento automatizado, incluindo a definição de perfis, que produza efeitos na sua esfera jurídica ou que o afete significativamente de forma similar.
> 2. O n.o 1 não se aplica se a decisão:
> a) For necessária para a celebração ou a execução de um contrato entre o titular dos dados e um responsável pelo tratamento;
> b) For autorizada pelo direito da União ou do Estado-Membro a que o responsável pelo tratamento estiver sujeito, e na qual estejam igualmente previstas medidas adequadas para salvaguardar os direitos e liberdades e os legítimos interesses do titular dos dados; ou
> c) For baseada no consentimento explícito do titular dos dados.
> 3. Nos casos a que se referem o n.o 2, alíneas a) e c), o responsável pelo tratamento aplica medidas adequadas para salvaguardar os direitos e liberdades e legítimos interesses do titular dos dados, designadamente o direito de, pelo menos, obter intervenção humana por parte do responsável, manifestar o seu ponto de vista e contestar a decisão.
> 4. As decisões a que se refere o n.o 2 não se baseiam nas categorias especiais de dados pessoais a que se refere o artigo 9.o, n.o 1, a não ser que o n.o 2, alínea a) ou g), do mesmo artigo sejam aplicáveis e sejam aplicadas medidas adequadas para salvaguardar os direitos e liberdades e os legítimos interesses do titular.[221]

---

[220] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30, Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024. p. 2013 (Tradução nossa).

[221] REGULAMENTO GERAL DE PROTEÇÃO DE DADOS EUROPEU. Disponível em: https://gdpr-text.com/pt/read/article-22/. Acesso em: 04 fev. 2024.

Uma sugestão seria oferecer ao titular de dados, quando este recebe o resultado de uma decisão automatizadas por meio de sistema de Inteligência Artificial, também uma orientação sobre como é possível apelar da decisão, com prazos e ponto ou canal de contato. Neste modelo da RGPD, não há previsão de um terceiro externo neutro para a tomada de decisão, como mencionado como característica de um devido processo legal "ideal", mas ao menos obriga este devido processo pelo agente de tratamento de dados – o que consideramos por questões pragmáticas ser o primeiro passo.

Neste arquétipo, os Estados Membros da União Europeia possuem liberdade para criar suas próprias versões de direito de contestação – o que pode resultar em custos de conformidade elevados considerando a frequência na qual as mesmas empresas operam em diferentes países e em ruídos nas decisões a serem tomadas. Assim, a existência de alguma coordenação ou entidade supervisora, poderia ser um caminho interessante na mitigação deste tipo de ruído, até mesmo evitando um "planejamento territorial de centros de IA" na comunidade europeia.

Ademais, reforçando a distinção entre direito procedimental ou substantivo, o direito à contestação previsto na RGPD, ao se aproximar mais de um modelo procedimental, não deixa claro em que base alguém pode contestar a decisão.

> Como outras restrições à tomada de decisão por IA, o direito tem a intenção de proteger "os direitos e liberdades legítimos do sujeito dos dados". Pode-se argumentar, portanto, que o GDPR concede o direito de contestar não apenas decisões errôneas, mas decisões tendenciosas e discriminatórias, e a maioria das decisões baseadas em dados pessoais altamente sensíveis.[222]

Se estabelece assim, o risco do enfraquecimento da contestação, com agentes de tratamento de dados que estejam focados apenas em formalizar que a possibilidade de contestação foi oferecida, mas com total ineficácia na prática.

O segundo arquétipo analisado para a construção do direito a contestação é o do "*Notice-and-Takedown*" da *Digital Millennium Copyright Act* (DMCA) dos Estados Unidos e o direito de contestação no Reino Unido. Aqui o foco está em um modelo de contestação com ênfase nos procedimentos a serem seguidos, em vez de bases substanciais para contestar decisões.

No contexto da DMCA, o processo de "*Notice-and-Takedown*" permite que titulares de direitos autorais solicitem a remoção de conteúdo online que considerem infringir seus direitos. No entanto, os autores apontam que esse processo muitas vezes falha em cumprir seu propósito, com remoções questionáveis e poucos contra-avisos sendo apresentados. Isso levanta questões

---

[222] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024. p. 2015 (Tradução nossa).

sobre a eficácia do processo de contestação e a necessidade de aprimorá-lo para garantir que seja justo e equilibrado.

No Reino Unido, a implementação do direito de contestação sob o GDPR segue uma abordagem altamente procedimental para desafiar decisões tomadas por sistemas algorítmicos. Antes do GDPR, a Lei de Proteção de Dados do Reino Unido de 1998 já previa um processo para contestar decisões automatizadas, exigindo que as empresas notificassem os indivíduos sobre tais decisões e fornecessem a oportunidade de solicitar reconsideração ou uma nova decisão com envolvimento humano.

A Seção 512 da DMCA se tornou um modelo mundial para o devido processo legal individual aplicado ao meio online, contendo uma das versões mais adotadas de direito de contestação na era digital.

Com o objetivo de resolver disputas de direitos autorais online, a Seção 512 teve como premissas a redução de incertezas para os provedores de serviços online, bem como a redução dos custos para os titulares de direitos autorais que tinham obras suas sendo infringidas no âmbito online. Assim, estes titulares podem enviar avisos de remoção diretamente para os provedores. Se os provedores respondessem removendo o conteúdo, eles ficariam mais seguros sobre eventuais responsabilidades secundárias pelas infrações de direitos autoriais dos seus usuários.

Por outro lado, é possível que o usuário que teve o seu conteúdo removido conteste a remoção, enviando uma contranotificação para que o conteúdo seja restabelecido. Neste caso, o provedor deveria enviar a contranotificação para o sujeito que solicitou a remoção e este poderá optar por entrar com uma ação judicial por infração de direitos autorais ou deixar a disputa prosseguir de forma extrajudicial. Se após dez a quatorze dias nenhuma ação judicial for movida, o provedor deve substituir o material alvo.

A Seção 512 edita também vários elementos necessários para esse devido processo, como especificando quais informações devem ser enviadas pelo titular dos direitos autorais na contestação ou pelo usuário na contranotificação.

É, assim, possível que na prática os provedores de serviços online revisem a notificação para analisar se será ou não feita a remoção, analisando um aspecto substancial. Contudo, especialmente por questões procedimentais de escala, como a necessidade de gerar muitos avisos, os provedores, em sua maioria, removem de forma automatizada, desde logo, o conteúdo contestado.

Outros motivos sistêmicos que incentivam os provedores a agir dessa segunda maneira são: i) que o titular de direitos autorais enviando um aviso de remoção não precisa provar uma

violação de direitos autorais, basta eles afirmarem que acreditam que o uso direcionado não é autorizado; e ii) a força das ferramentas de proteção dos direitos autorais e os custos envolvidos com problemas desse tipo, tornando menos arriscado em termos de impacto da concretização do risco simplesmente olhar os requisitos procedimentais.

A Seção 512, contudo, não está livre de críticas:

> No entanto, a Seção 512 atraiu preocupações de devido processo desde sua concepção. Detentores de direitos autorais reclamaram que o processo de remoção é insuficientemente eficaz, caro demais e oneroso. OSPs argumentaram que o processo da seção 512 corre o risco de capturar usos legais juntamente com materiais que infringem direitos. O mecanismo de "contranotificação" foi adicionado tarde no processo legislativo em resposta a preocupações de processo. Alguns apontaram que o período de remoção obrigatório de dez a quatorze dias poderia silenciar discursos sensíveis ao tempo; outros observaram que isso dava ao remetente da notificação pouco tempo para entrar com uma ação judicial se uma contranotificação chegasse. Grupos da sociedade civil se preocuparam que os mecanismos de contestação da seção 512 e outras características de design são insuficientes para deter remoções abusivas ou equivocadas.[223]

Assim, os mecanismos de contestação da DMCA, portanto, aparentemente falharam em cumprir perfeitamente o seu propósito. Apesar da evidência de que remoções impróprias ou questionáveis são comuns, as contranotificações parecem ser raras e ações judiciais mais raras. A análise da Seção 512, portanto, traz lições ao ensinar sobre o que evitar, ou incluir, ao projetar o direito de contestação.

Ainda neste segundo arquétipo, se inclui a abordagem do Reino Unido altamente processual - antes do Brexit – para a contestação das decisões algorítmicas. A legislação anterior à RGDP estabelecia que a empresa deveria notificar os indivíduos sobre uma decisão automatizada assim que razoavelmente possível e fornecesse aos indivíduos um prazo de vinte e um dias (que atualmente é um mês) para que estes pudessem solicitar a reconsideração ou uma nova decisão com envolvimento humano. Mas não havia nada na lei que especificasse quais medidas deveriam ser tomadas de forma substancial no caso de um pedido de reconsideração.[224]

---

[223] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024. p. 2009 (Tradução nossa).

[224] Rights in relation to automated decision-taking. (1)An individual is entitled at any time, by notice in writing to any data controller, to require the data controller to ensure that no decision taken by or on behalf of the data controller which significantly affects that individual is based solely on the processing by automatic means of personal data in respect of which that individual is the data subject for the purpose of evaluating matters relating to him such as, for example, his performance at work, his creditworthiness, his reliability or his conduct.

(2) Where, in a case where no notice under subsection (1) has effect, a decision which significantly affects an individual is based solely on such processing as is mentioned in subsection (1) —

> A nova lei do Reino Unido implementando o RGPD estendeu os vinte e um dias para um mês. Como antes, uma empresa deve notificar um indivíduo de uma decisão automatizada por escrito "assim que razoavelmente prático"; o indivíduo tem um mês (em vez de vinte e um dias) para solicitar que uma empresa "(i) reconsidere a decisão, ou (ii) tome uma nova decisão que não seja baseada unicamente no processamento automatizado." A empresa então, ordinariamente, tem um mês para "considerar o pedido, incluindo qualquer informação fornecida pelo sujeito dos dados... cumprir com o pedido, e... por notificação escrita informar o sujeito dos dados de: (i) as etapas tomadas para cumprir com o pedido, e (ii) o resultado de cumprir com o pedido." Este processo pode ser alterado por regulamentação. Isso estabelece o que Gianclaudio Malgieri referiu como uma "explicação procedimentalizada" — os indivíduos recebem insights sobre o processo de contestação, o que por si só pode verificar o comportamento da empresa, ou até mesmo incentivar resultados pró-consumidor através da transparência.[225]

Portanto, este arquétipo destaca a importância de estabelecer procedimentos claros e acessíveis para a contestação de decisões baseadas em Inteligência Artificial, garantindo que os indivíduos tenham meios eficazes para contestar e questionar tais decisões, mesmo que a base substancial para contestação possa não ser tão clara.

---

(a) the data controller must as soon as reasonably practicable notify the individual that the decision was taken on that basis, and
(b) the individual is entitled, within twenty-one days of receiving that notification from the data controller, by notice in writing to require the data controller to reconsider the decision or to take a new decision otherwise than on that basis.
(3) The data controller must, within twenty-one days of receiving a notice under subsection (2)(b) ("the data subject notice") give the individual a written notice specifying the steps that he intends to take to comply with the data subject notice.
(4) A notice under subsection (1) does not have effect in relation to an exempt decision; and nothing in subsection (2) applies to an exempt decision.
(5) In subsection (4) "exempt decision" means any decision—
(a) in respect of which the condition in subsection (6) and the condition in subsection (7) are met, or
(b) which is made in such other circumstances as may be prescribed by the [F1Lord Chancellor] by order.
(6) The condition in this subsection is that the decision—
(a) is taken in the course of steps taken—
(i) for the purpose of considering whether to enter into a contract with the data subject,
(ii) with a view to entering into such a contract, or
(iii) in the course of performing such a contract, or
(b) is authorised or required by or under any enactment.
(7) The condition in this subsection is that either—
(a) the effect of the decision is to grant a request of the data subject, or
(b) steps have been taken to safeguard the legitimate interests of the data subject (for example, by allowing him to make representations).
(8) If a court is satisfied on the application of a data subject that a person taking a decision in respect of him ("the responsible person") has failed to comply with subsection (1) or (2)(b), the court may order the responsible person to reconsider the decision, or to take a new decision which is not based solely on such processing as is mentioned in subsection (1).
(9) An order under subsection (8) shall not affect the rights of any person other than the data subject and the responsible person.
UK Data Protection Act of 1998. Disponível em: https://www.legislation.gov.uk/ukpga/1998/29/section/12/2001-11-26#:~:text=(1)An%20individual%20is%20entitled,personal%20data%20in%20respect%20of.

[225] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024. p. 2021 (Tradução nossa).

Já o terceiro arquétipo trazido no estudo de Margot Kaminski e Jennifer Urban[226] é o do "direito ao esquecimento" e de contestação da Hungria e da Eslovênia sob o regime da RGPD.

Aqui o foco está em um modelo de contestação com ênfase em critérios substanciais para contestar decisões, em oposição aos com procedimentos específicos a serem seguidos como nos arquétipos anteriores.

O "Direito ao Esquecimento" surgiu da legislação europeia de proteção de dados antes do RGPD e foi fortalecido pelo caso Google Spain de 2014, no qual o Tribunal de Justiça da União Europeia determinou que os motores de busca devem responder a certas solicitações de indivíduos para remover dados pessoais dos resultados de busca. Esse direito permite que os indivíduos contestem a inclusão de seus dados pessoais em resultados de busca. Ele pode ser visto, assim, como um direito de contestação da inclusão dos seus dados pessoais nos resultados de mecanismos de busca.

Em circunstâncias específicas, como a de figuras públicas e nos casos em que o interesse do público geral em determinada informação supera as preocupações com a privacidade, as empresas podem manter informações nos resultados de busca. O balanceamento para a decisão de remoção ou não envolve, assim, mais do que os interesses econômicos da empresa operadora e da privacidade do indivíduo, e leva em conta também interesses sociais de ser informado. Não há, contudo, qualquer regra processual, apenas um padrão de contestação, deixando espaço para a interpretação sobre o que constitui interesse individual de privacidade ou interesse público de acesso ao conteúdo.

Assim, como resultado deste modelo - que pode ser dito como um modelo com foco substancial, as empresas seguiram no caminho de criarem sistemas de contestação privatizados próprios, como foi o caso do Google (sistemas esses influenciados pelas Diretiva de Comércio Eletrônico europeu).

No caso do Google, especificamente, a empresa possui um formulário[227] que permite ao indivíduo solicitar a exclusão de algum dado pessoal seu, e para o encaminhamento do processo é necessário fornecer documentos de verificação da titularidade.

> Primeiro, o Google estabeleceu um Conselho Consultivo que emitiu um relatório indicando os critérios substantivos que o motor de busca usaria na

[226] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024.

[227] Suporte do Google. Disponível em: https://support.google.com/websearch/contact/content_removal_form?visit_id=638450030873732525-3328134489&rd=1. Acesso em: 04 fev. 2024.

> avaliação de pedidos de remoção. Os reguladores, então, estabeleceram sua própria lista de critérios. Esse diálogo esclareceu em grande medida o padrão substantivo estabelecido pelo CJEU para algo mais semelhante a uma regra em natureza.[228]

Após a mencionada decisão, assim, a estratégia do Google envolveu contratar funcionários para processas as reinvindicações, ter advogados, engenheiros e gerentes de produto ou até mesmo especialistas externos. Em alguns casos havia até mesmo contato com o solicitante para obter mais informações.

A reivindicação poderia ser aceita e o conteúdo retirado ou rejeitada, hipótese na qual o indeferimento era comunicado juntamente com a informação sobre o direito do titular de apresentar uma reclamação perante uma autoridade nacional de proteção de dados.

Não obstante a teoria envolver a ponderação entre privacidade, interesses econômicos do mecanismo de busca e interesses públicos de acesso à informação, na prática, este arquétipo parece tender a favorecer a exclusão. E, caso não seja uma política da empresa a divulgação das informações sobre a tomada de decisão, o racional permanece obscuro e pouco contestável.

Outro exemplo deste terceiro arquétipo acontece na Hungria, país que proíbe a tomada de decisões automatizadas com base em dados sensíveis, a menos que previsto por lei. Ao se fundamentar na sensibilidade dos dados, assim, este direito de contestação é fundamentado em critérios substanciais.

Por fim, o quarto arquétipo é um modelo de regra de contestação com foco substancial. Aqui se tem como exemplo deste referencial o mecanismo de *chargeback* do *Fair Credit Billing Act* (FCBA) americano.

O FCBA oferece aos consumidores o direito de contestar cobranças indevidas em seus cartões de crédito por meio de um processo de *chargeback*, o que poderia ser equiparado a um processo de estorno. Para isso são estabelecidos quais são os erros de cobrança que podem servir como base substancial para o direito de contestar. São exemplos de cobranças contestáveis as não autorizadas ou incorretas ou por bens e serviços que não tenham sido entregues ou aceitos.

Mas essa lei vai além, pois também estabelece requisitos procedimentais detalhados – gratuitos, acessíveis e rápidos - para o processo de contestação, reduzindo a capacidade das empresas de cartão de crédito de rejeitar as solicitações de estorno. Com isso, o FCBA acaba

---

[228] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024. p. 2024 (Tradução nossa).

sendo mais bem sucedido do que o DMCA mencionado anteriormente na resolução de disputas. Nem sempre o comerciante é prejudicado e a confiança do consumidor aumenta.

> Estudiosos da resolução alternativa de disputas criticaram o processo do FCBA por vários motivos, incluindo a falta de danos, a limitada conscientização do consumidor sobre o processo e o método de julgamento à distância em vez de baseado em relacionamento. No entanto, defensores da proteção do consumidor veem, em grande parte, o processo de *chargeback* do FCBA como um sucesso. As empresas de cartão de crédito decidem a favor dos consumidores cerca de oitenta a noventa por cento das vezes. Talvez por causa dessa taxa de sucesso, os consumidores parecem ver o processo com satisfação e raramente entram com ações judiciais para desafiá-lo. [229]

Certamente apesar disso, o FCBA não é um modelo perfeito e muito menos pode não ser adequado para todos os tipos de disputas, que nem sempre são tão fáceis como a disputa de uma cobrança de cartão, como no caso de discriminações resultantes de *proxy*.

Estes arquétipos são interessantes pontos de partida para a reflexão do direito a contestação, mas existem muitos outros fatores subentendidos nestes modelos que impactam na maior ou menor eficácia do processo de contestação privatizado (sendo que consideramos que este deve ser apenas o primeiro passo).

O design desses processos, assim, de uma forma mais ampla passa pela consideração sobre as estruturas de incentivo (como foi exemplificado nos incentivos indiretos para os mecanismos de busca retirarem o conteúdo), a transparência ao longo do processo com foco em garantir um processo justo, a escolha de quem toma a decisão (um terceiro árbitro neutro, um outro algoritmo etc.), ou contexto regulatório (como é o caso do recorte deste trabalho na esfera do uso de dados pessoais, que já prevê no seu arcabouço regulamentar uma autoridade, mecanismos de governança, relatórios etc.).

Mas não é só, o próprio contexto sociocultural no qual o processo será implementado (o Brasil, por exemplo, exige a abordagem de questões de desigualdade de recursos, conhecimento, poder e informação, o analfabetismo, as raciais entre outras) e a legislação aplicável aos direitos conexos que preveem mais ou menos mecanismos de garantia de exercício de direitos e o acesso à justiça.

### 4.3.1. A construção de Arquétipos gerais

[229] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024. p. 2029 (Tradução nossa).

Durante muitos anos a construção destes arquétipos de contestação foram locais, setoriais, voltados para solucionar dores e riscos de atividades específicas, como os apresentados anteriormente. Contudo, a expansão do uso de sistemas de tomadas de decisão por meio de Inteligência Artificial, especialmente com a disseminação da Inteligência Artificial Generativa[230], que possibilitou a ampliação exponencial do campo de uso da Inteligência Artificial, a necessidade de uma articulação mais ampla sobre o tema se tornou necessária.

Assim o avanço tecnológico, a necessidade de proteger direitos fundamentais de uma forma ampla e a busca por alguma segurança jurídica em um cenário de incertezas quanto ao impacto da Inteligência Artificial na sociedade, foram combustíveis para que a União Europeia e também países como o Canadá e o Brasil começassem a estruturar regulamentações sobre o uso da Inteligência Artificial.

O *Artificial Intelligence Act* (AIA) europeu surge como uma resposta direta ao reconhecimento de que sistemas de Inteligência Artificial têm o potencial de transformar profundamente diversos aspectos da vida social, econômica e política da comunidade europeia. Dentre os principais objetivos declarados do AIA, estão a proteção de direitos fundamentais (como o da privacidade, da igualdade e da não discriminação) e o de garantir a segurança e confiabilidade destes sistemas (incluindo classificação baseada em graus de riscos). E no Brasil, o PL 2338/23 parece seguir linha semelhante.

Naturalmente, como esperado, dentro deste contexto, a problemática da possibilidade de contestação das decisões tomadas por meio de sistemas de Inteligência Artificial foi levantada. Contudo, no PL 2338/23, por exemplo, as disposições substanciais e procedimentais deste direito de contestação ainda dependerão de regulamentação para que este seja efetivamente observador.

O art. 3º estabelece que o desenvolvimento, a implementação e o uso de sistemas de Inteligência Artificial observarão a boa-fé e os princípios do devido processo legal, contestabilidade e do contraditório e o da rastreabilidade das decisões durante o ciclo de vida

---

230 "IA generativa refere-se a uma categoria de modelos e ferramentas de IA projetadas para criar novos conteúdos, como texto, imagens, vídeos, música ou código. A IA generativa usa uma variedade de técnicas – incluindo redes neurais e algoritmos de aprendizado profundo (Deep Learning) – para identificar padrões e gerar novos resultados.
IA generativa é um subcampo da IA que se concentra na criação de novos conteúdos, dados ou informações a partir de um conjunto de entradas existentes. Esses algoritmos de IA aprendem com os dados fornecidos e são capazes de gerar saídas semelhantes, mas não idênticas, com base no conhecimento adquirido durante o treinamento." Informação disponível em: https://blog.dsacademy.com.br/guia-completo-sobre-inteligencia-artificial-generativa/. Acesso em 10 ago. 2024.

de sistemas de Inteligência Artificial como meio de prestação de contas e atribuição de responsabilidades a uma pessoa natural ou jurídica.

Já o art. 5º dispões que as pessoas afetadas têm o direito de contestar decisões ou previsões de sistemas de Inteligência Artificial que produzam efeitos jurídicos ou que impactem de maneira significativa os interesses do afetado e tem "direito à determinação e à participação humana em decisões de sistemas de inteligência artificial, levando-se em conta o contexto e o estado da arte do desenvolvimento tecnológico".

Indo além, o PL 2338/23 traz seção específica sobre o direito de contestar decisões e de solicitar intervenção humana, dispondo que:

> Art. 9º A pessoa afetada por sistema de inteligência artificial terá o direito de contestar e de solicitar a revisão de decisões, recomendações ou previsões geradas por tal sistema que produzam efeitos jurídicos relevantes ou que impactem de maneira significativa seus interesses.
>
> § 1º Fica assegurado o direito de correção de dados incompletos, inexatos ou desatualizados utilizados por sistemas de inteligência artificial, SF/23833.90768-16 Página 7 de 33 Avulso do PL 2338/2023 assim como o direito de solicitar a anonimização, bloqueio ou eliminação de dados desnecessários, excessivos ou tratados em desconformidade com a legislação, nos termos do art. 18 da Lei nº 13.709, de 14 de agosto de 2018 e da legislação pertinente.
>
> § 2º O direito à contestação previsto no caput deste artigo abrange também decisões, recomendações ou previsões amparadas em inferências discriminatórias, irrazoáveis ou que atentem contra a boa-fé objetiva, assim compreendidas as inferências que: I – sejam fundadas em dados inadequados ou abusivos para as finalidades do tratamento; II – sejam baseadas em métodos imprecisos ou estatisticamente não confiáveis; ou III – não considerem de forma adequada a individualidade e as características pessoais dos indivíduos.
>
> Art. 10. Quando a decisão, previsão ou recomendação de sistema de inteligência artificial produzir efeitos jurídicos relevantes ou que impactem de maneira significativa os interesses da pessoa, inclusive por meio da geração de perfis e da realização de inferências, esta poderá solicitar a intervenção ou revisão humana.
>
> Parágrafo único. A intervenção ou revisão humana não será exigida caso a sua implementação seja comprovadamente impossível, hipótese na qual o responsável pela operação do sistema de inteligência artificial implementará medidas alternativas eficazes, a fim de assegurar a reanálise da decisão

> contestada, levando em consideração os argumentos suscitados pela pessoa afetada, assim como a reparação de eventuais danos gerados.
>
> Art. 11. Em cenários nos quais as decisões, previsões ou recomendações geradas por sistemas de inteligência artificial tenham um impacto irreversível ou de difícil reversão ou envolvam decisões que possam gerar riscos à vida ou à integridade física de indivíduos, haverá envolvimento humano significativo no processo decisório e determinação humana final.[231]

Apesar de ser extremamente relevante que o referido projeto de lei tenha contemplado tais disposições, assim como ocorre na previsão da LGPD - do direito de solicitar a revisão de uma decisão automatizada -, o caminho para a real concretização deste direito ainda parecer ser longo.

Termos como "impactem de maneira significativa seus interesses" em razão de serem vagos, deixam margem para interpretações subjetivas, com uma ausência de uma definição clara sobre o que constitui um impacto significativo, o que pode dificultar a aplicação efetiva do direito de contestação até que exista uma regulamentação que traga exemplos de forma mais concreta ou que se consolide alguma jurisprudência sobre o tema.

De igual maneira, a dispensa da intervenção humana quando "comprovadamente impossível" pode abrir indesejáveis brechas para que empresas ou operadores de sistemas de tomada de decisão por meio de Inteligência Artificial aleguem impossibilidades técnicas para evitar a revisão humana. Ou até mesmo que construam sistemas que já tornem essa intervenção impossível para evitar custos com eventual necessidade de intervenção humana. Isso enfraquece a garantia de que decisões automatizadas possam ser adequadamente revisadas, comprometendo a proteção dos direitos dos afetados.

Observa-se também que o artigo 10º permite que a pessoa afetada solicite a revisão humana da decisão, mas não detalha como essa revisão deve ser conduzida, quem são os responsáveis pela revisão, ou quais qualificações ou critérios os revisores devem atender. A ausência dessas especificações pode resultar em revisões superficiais ou ineficazes.

É certo que o desafio deste projeto de lei é gigantesco e o equilíbrio entre uma baixa normatividade e uma super normatividade é muito difícil de se alcançar. Mas, enquanto não há diretrizes claras, específicas do devido processo de contestação, a ser seguido por quem utiliza,

---

[231] BRASIL. SENADO FEDERAL. Projeto de lei n° 2338, de 2023. Dispõe sobre o uso da Inteligência Artificial. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9347622&ts=1720798347645&disposition=inline . Acesso em: 10 ago. 2024.

desenvolve ou de alguma forma oferece sistemas de tomada de decisão por meio de Inteligência Artificial, tanto prévias e posteriores, é difícil se visualizar efetiva proteção aos direitos individuais e a mitigação dos riscos tratados nesta pesquisa.

A adoção do princípio do "humano no circuito" também é recomendada na estruturação de um direito de contestar eficiente. Uma Inteligência Artificial desenvolvida para tomar decisões automatizadas, centrada em aspectos humanos, deveria incluir meios que possibilitem a um especialista contribuir para aperfeiçoar a elaboração e o desempenho de decisões da máquina.

> Essa contribuição pode ocorrer principalmente quando o know how do especialista é usado para melhorar a qualidade dos dados empregados, em uma atividade que chamamos 'humano no circuito'. Nesse caso, o modelo de aprendizado de máquina atua como um aprendiz capaz de ajudar o especialista, enquanto também aprende observando as decisões do ser humano, adotando-as como exemplos de treinamento adicionais.[232]

As funções do "homem no circuito" podem envolver atividades de supervisão (monitoramento do desempenho do sistema de Inteligência Artificial e garantia de que ele esteja funcionando de acordo com as expectativas e com os princípios éticos); de intervenção quando necessário para interromper o processo de tomada de decisão automatizada, especialmente em situações complexas ou de alto risco; de interpretação para ajudar a explicar as decisões tomadas pelo sistema de Inteligência Artificial, tanto para os indivíduos afetados quanto para a sociedade em geral e até mesmo de responsabilidade, assumindo a responsabilidade pelas decisões tomadas pelo sistema de Inteligência Artificial, eventualmente em conjunto com os desenvolvedores e outros stakeholders.

Idealmente este conceito é desenhado para trazer benefícios de maior justiça e equidade, para reduzir o risco de viés e discriminação nas decisões automatizadas, maior transparência ao aumentar a compreensão de como as decisões automatizadas são tomadas, permitindo que os indivíduos afetados compreendam o processo e seus impactos e traga uma maior confiança do público em sistemas de Inteligência Artificial, ao garantir que exista mecanismos de controle e *accountability*.

Contudo, os desafios do "homem no circuito" também não podem ser ignorados. O custo de implementação deste modelo pode ser caro, especialmente em sistemas mais complexos e poderia se argumentar que haveria uma redução de eficiência e velocidade da decisão. Assim,

---

[232] Disponível em: CHAVES, Alaor. (org.) **Ciência para prosperidade:** sustentável e socialmente justa. Belo horizonte: EMBRAPII, 2022. 160 p. Disponível em: https://homepages.dcc.ufmg.br/~nivio/papers/Ciencia-para-prosperidade.pdf.

este tipo de modelo tem sido visto atualmente como mais defendido em processos cujos riscos são altos em razão do impacto da decisão. Seria o caso de revisão por juízes de decisões de um algoritmo ou por um médico no uso de sistema de Inteligência Artificial para diagnosticar doenças.

Uma interessante "evolução" deste modelo é o do "*Community-in-the-loop*"[233], que considerando a necessidade de uma criação pluralistas de valor na Inteligência Artificial, destaca a importância da participação da comunidade e do envolvimento de diversas partes interessadas no processo de tomada de decisão relacionado à Inteligência Artificial.

Aqui as implicações éticas das decisões de Inteligência Artificial seriam vistas como conflitos entre os valores de diversas partes interessadas e, assim, uma proposta de solução destes conflitos poderia ser por meio de uma ética de ordem deliberativa, na qual as partes interessadas de uma comunidade deliberam sobre os custos e benefícios e concordam com regras para *trade-offs* aceitáveis quando os sistemas de Inteligência Artificial são empregados.[234]

## 4.4 CRIANDO AS BASES PARA O DIREITO DE CONTESTAR A DECISÃO

Por tudo quanto exposto neste trabalho, falar em direito de contestar uma decisão tomada por um sistema de Inteligência Artificial ou mesmo decisões humanas que dependem significativamente de ferramentas de Inteligência Artificial, demanda uma abordagem multifocal, o que envolve pensar em criação de confiança, em devido processo legal e informacional, realidade contextual, legitimidade, incentivos, o humano envolvido no processo e a interação dele com a máquina, o arcabouço regulatório dos direitos substanciais correlatos etc. E, por tudo isso, como dito, é ilusório o objetivo de uma prescrição, solução, única.

Não obstante, é possível se falar em melhores práticas, em lições aprendidas, riscos evitáveis e mitigação desses riscos. É igualmente possível se pensar em soluções articuladas na esfera jurídica, técnica e política.

---

[233] WANG, Ye *et al*. Communityin-the-loop: Creating Artificial Process Intelligence for Co-production of City Service. **Proceedings of the ACM on Human Computer Interaction,** v. 6, n. 285. 2022. Disponível em: https://doi.org/10.1145/3555176. Acesso em: 05 fev. 2024.

[234] WANG, Ye *et al*. Communityin-the-loop: Creating Artificial Process Intelligence for Co-production of City Service. **Proceedings of the ACM on Human Computer Interaction,** v. 6, n. 285. 2022. Disponível em: https://doi.org/10.1145/3555176. Acesso em: 05 fev. 2024.

Pelos exemplos trazidos ao longo da pesquisa, um foco substantivo do direito de contestar uma decisão tomada por sistemas de Inteligência Artificial parece ser indicado quando pensamos nos reflexos de questões de privacidade, precisão, discriminação, vieses, direitos fundamentais e até mesmo os valores que norteiam um Estado de Direito. Ter essa base substantiva clara é o que irá garantir no contexto a adequação para a aplicação, criando maior efetividade ao direito de contestar. Uma simples previsão legal do direito de indivíduos, sujeitos a uma decisão automatizada por meio de sistemas de Inteligência Artificial, de contestar essa decisão é absolutamente ineficaz.

O mencionado PL 2338/23, por exemplo, trouxe um foco substantivo ao estabelecer critérios como a produção de efeitos jurídicos relevantes ou que impactem de maneira significativa os seus interesses do indivíduo (inobstante a vagueza do critério).

Mas, a pouca prescrição procedimental, acaba deixando para futuros atores, ou até mesmo para a autorregulação, o como, o caminho para a efetiva aplicação. O que já tem se mostrado uma opção perigosa.

Parece haver ainda uma ausência de desenhos legais ou regulatórios sobre o próprio desenvolvimento destes sistemas. O máximo que se tem observado são algumas escolhas sobre riscos que seriam inadmissíveis. De forma especulativa, o motivo para tal fato pode ser o próprio desconhecimento sobre o funcionamento da Inteligência Artificial. Quando pouco se sabe sobre o funcionamento, regular o impacto, com foco nos riscos conhecidos, pode ser um caminho mais comum.

Remédio para isso poderia ser a conjunção de abordagens como a do *privacy by design*, desenvolvida pela Ann Cavoukian, que aplicada ao direito de contestar sistemas de tomada de decisão por meio de Inteligência Artificial, criaria o conceito de uma tecnologia contestável *by design*. Essa aplicação seria fundamental para a efetivação desse direito de contestar.

A abordagem da Ann Cavoukian, de maneira bastante sintética, tem como foco garantir que a privacidade será incorporada em todas as etapas de um sistema, desde a sua concepção até o fim da sua vida útil. E a privacidade aqui aparece não apenas com o seu recorte legal, mas também ético.

Sendo desenvolvida em 2009, a metodologia do *privacy by design,* foi apresentada pela sua criadora pela primeira vez ao público na 31ª Conferência Internacional de Comissários de Proteção de Dados e Privacidade em 2009, o que fez com que em 2010 ela fosse adotada como um parâmetro internacional. Foram desenvolvidos sete princípios (que inclusive estão atualmente fundamentando a ISO 31700:2023 de *Consumer protection – Privacy by design for consumer goods and services*).

Os princípios são: i) Proativo, não reativo; ii) Privacidade como padrão; iii) Privacidade incorporada ao design; iv) Funcionalidade Total; v) Segurança de ponta a ponta; vi) Visibilidade e Transparência; vii) Respeito ao usuário – centrada no usuário.

Assim, em um exercício de utilização desses guias ao direito de contestar uma decisão tomada por Inteligência Artificial em um contexto de eventual necessidade de proteção de dados pessoais, poderíamos ver, por exemplo, o princípio da transparência aplicado quando, antes da tomada de decisão, é explicado claramente os critérios e os dados utilizados para se chegar a um determinado resultado. Nessa metodologia todo o processo de uso do dado deve ser feito de acordo com os objetivos declarados, passíveis de verificação e com componentes visíveis aos envolvidos no processo.

Não ocorreria aqui uma atuação posterior, não seria possibilitado ao indivíduo, caso ele considerasse que a decisão não foi correta, questionar e contestar a decisão, para apenas aí ter conhecimento sobre quais foram os critérios adotados e correlações feitas. Propomos, assim, transparência informacional prévia, sistemas criados para garantir efetivamente a autodeterminação, sem espaço (como regra que poderia comportar justificadamente alguma exceção) para obscuridades que possam ter impacto negativo em relação aos direitos fundamentais.

Seria o caso do uso de técnicas de Explainable artificial intelligence (XAI)[235], tanto de construção de modelos de "caixas branca" ou por meio de sistemas que permitem a aplicação de técnicas de pós-explicação. A XAI é usada para descrever a Inteligência Artificial, seu impacto esperado e até potenciais vieses, são um conjunto de técnicas que ajudam a caracterizar a precisão do modelo, a justiça, a transparência e os resultados na tomada de decisão[236].

Exemplos de XAI seriam os *Local Interpretable Model-Agnostic Explanations* (LIME)[237], que explicam a previsão de classificação pelo algoritmo de *machine learning*, ou o *Deep Learning Important FeaTures* (DeepLIFT)[238] que compara a ativação de cada "neurônio" (dentro da ideia de redes neurais) ao seu "neurônio" de referência e mostra uma ligação rastreável entre cada "neurônio" ativado, além de mostrar dependências entre eles.

---

[235] ZHU, Jichen *et al.* Explainable AI for designers: A human-centered perspective on mixed-initiative co-creation. **IEEE Conference on Computational Intelligence and Games (CIG)**. Netherlands, 2018. Disponível em: https://ieeexplore.ieee.org/abstract/document/8490433. Acesso em: 10 jul. 2024.

[236] IBM. *What is explainable AI*? Disponível em: https://www.ibm.com/topics/explainable-ai#:~:text=Explainable%20artificial%20intelligence%20(XAI)%20is,expected%20impact%20and%20potential%20biases. Acesso em: 10 jul. 2024.

[237] *Ibidem*.

[238] *Ibidem*.

É importante destacas que a XAI pode ir além de uma forma de entender a causa de uma decisão (conceito de *interpretability* utilizada em soluções tecnológicas), ela pode examinar o como a Inteligência Artificial chegou ou chegará ao resultado.

O princípio da funcionalidade total, por sua vez, estabelece que os diferentes interesses devem ser vistos não como excludentes, ou soma zero, mas sim por meio de uma soma positiva, na qual não são feitas compensações desnecessárias. E, quando aplico ao contexto desta pesquisa, ele serve como um guia para se evitar dicotomias como privacidade versos segurança ou explicabilidade versos segredos comerciais ou até mesmo eficiência e velocidade versos contestação e revisão.

O ser humano é criativo o bastante para criar soluções tecnológicas que operem neste modelo e, nos tempos atuais, as máquinas também podem ser "criativas" para desenvolver soluções a partir destes parâmetros. Tais modelos deveriam ser incentivados, ou até obrigatórios, a depender do âmbito de aplicação, independente de eventuais custos (financeiros, de tempo, de recurso ou de pessoal) que eles possam representar para os desenvolvedores. Os interesses público e social devem ser prioritários frente ao interesse de criar tecnologias para o lucro.

Um exemplo da aplicação deste princípio seria o uso dos *acordos criptográficos*, que Isabela Ferrari apresenta como os "equivalentes digitais a um documento selado por uma terceira parte, ou à manutenção de um documento em local seguro"[239].

Estes acordos assegurariam que o programa não foi alterado nem revelado (afastando os já mencionados argumentos do receio de violação aos segredos industriais ou comerciais ou evitando que os sujeitos impactados possam enganar e manipular o sistema) ao manter por sigilo o programa por determinado tempo, ocultando os critérios da tomada de decisão de forma provisória.

> Assim, passado certo tempo, os acordos criptográficos dão certeza sobre os critérios utilizados, e a partir daí pode-se seguir análise sobre a legitimidade de sua operação pretérita. A certeza de que haverá disclosure futuro tem o efeito de refrear a tendência a usar critérios inadequados, discriminatórios, etc.[240]

---

[239] FERRARI, Isabela. Accountability de algoritmos: a falácia do acesso ao código e caminhos para uma explicabilidade efetiva. Inteligência Artificial: 3º Grupo de Pesquisa do ITS, ITS - Instituto de Tecnologia e Sociedade do Rio, 2018. Disponível em: https://itsrio.org/wpcontent/uploads/2019/03/Isabela-Ferrari.pdf. Acesso em: 10 jul. 2024. p. 14.

[240] *Ibidem*. p. 15.

Fazendo uma conjunção dos já mencionados princípios do *by design*, soluções como as dos *acordos criptográficos* seriam exceções, quando realmente a explicação e transparência prévias representassem o risco a um bem maior a ser tutelado.

A privacidade incorporada ao *design* também é um princípio que pode auxiliar neste processo de construção ao estabelecer que a privacidade deve ser uma parte integrante do sistema, sem diminuir a sua funcionalidade e que ela é essencial, não devendo ser adicionada como um complemento depois, estando incorporada a toda a arquitetura do sistema.

De maneira analógica, a auditabilidade incorporada ao *design* poderia criar a obrigação de desenvolvimento e comercialização apenas de sistemas que tivessem sido desenvolvidos com a capacidade de registrar, monitorar e revisar suas decisões e processos internos, de maneira que seja possível entender, rastrear e, se necessário, contestar essas decisões. Um *design* tendo a vista a auditabilidade algorítmica e um poder de contestação *by desing*.

Seria o caso de uso da ferramenta *Shapley Additive Explanations* (SHAP)[241], utilizada para se explicar, de forma visual gráfica e intuitiva, os critérios utilizados para se chegar nas decisões tomadas pelos sistemas de Inteligência Artificial. Ou da Inteligência Artificial Semântica[242], que se concentra na compreensão e interpretação do significado (da semântica) das informações e dados utilizados pelo sistema e utilizam por exemplo, *knowledge graphs,* ou gráficos de conhecimento, para organizar informações de maneira que as relações entre os dados sejam claras e acessíveis, permitindo uma compreensão mais rica e interconectada dos dados[243].

O ponto de vista procedimental, o controle do usuário, ser centrado nele, também garantiria que o titular tivesse sempre acesso a critérios claros sobre como poder apresentar uma contestação – formulário online, canal do encarregado pelo tratamento de dados, um terceiro independente, como se opera a notificação, a justificativa etc. e que seja a ele permitido decidir sobre o uso dos seus dados, podendo manifestar sua discordância sobre determinada decisão a qualquer momento.

---

241 Mais informações sobre a ferramenta disponível em: https://shap.readthedocs.io/en/latest/ e https://bixtecnologia.com.br/como-eu-gostaria-que-alguem-me-explicasse-shap-values/#:~:text=A%20sigla%20significa%20SHapley%20Additive,com%20uma%20matem%C3%A1tica%20bem%20robusta. Acesso em: 07 jul. 2024.

242 FERRARI, Isabela. Accountability de algoritmos: a falácia do acesso ao código e caminhos para uma explicabilidade efetiva. Inteligência Artificial: 3º Grupo de Pesquisa do ITS, ITS - Instituto de Tecnologia e Sociedade do Rio, 2018. Disponível em: https://itsrio.org/wpcontent/uploads/2019/03/Isabela-Ferrari.pdf. Acesso em: 10 jul. 2024.

243 CHAUDHRI, Vinay K.; CHITTAR, Naren; GENESERETH, Michael. **An introduction to knowledge graphs**. The Stanford AI Lab Blog. Disponível em: https://ai.stanford.edu/blog/introduction-to-knowledge-graphs/. Acesso em: 13 jul. 2024.

O sujeito, com base neste princípio, teria direito a um processo que o capacitasse para desafiar as decisões automatizadas tomadas por sistemas de Inteligência Artificial.

Um foco meramente procedimental de direito de contestação claramente destoa de toda a lógica de governança das legislações de proteção de dados pessoais e do que temos visto atualmente no PL 2338/23. Assim, a opção pela adoção deste tipo de modelo seria certamente uma opção meramente formal e ineficaz na proteção do direito fundamental à privacidade e à proteção dos dados pessoais. É necessário, assim, ir além.

Para Margot Kaminski e Jennifer Urban[244] o direito de contestar deve ser construído por meio de uma abordagem híbrida, que incorpore no seu *design* elementos de diversos arquétipos com ferramentas regulatórias. Um modelo que combine regras procedimentais com regras de substanciais. A busca deve ser por um modelo que equilibre flexibilidade e clareza dentro do processo de contestação.

Sem que exista regras procedimentais claras, um titular de dados, sujeito da decisão inconformado, teria dificuldade de saber se e como pode contestar, e a controladora que foi responsável pela tomada de decisão sobre o tratamento dos dados não teria prazo para responder ou poderia criar procedimentos próprios com exigências que impossibilitassem na prática o direito de contestação. E, por outro lado, sem a justificativa, a compreensão sobre a decisão, seria difícil saber o que se contesta de forma fundamentada.

Mas um modelo apenas de "padrão de contestação", ao invés de regras claras e substantivas, poderia representar uma falta de clareza procedimental que desempodera o titular dentro de um processo de contestação.

> Um direito de contestação que, como a implementação do Reino Unido, ilustra uma regra de contestação com foco procedimental enfrenta problemas semelhantes. Ele também corre o risco de se tornar excessivamente procedimental em detrimento da substância, comprometendo valores de devido processo como precisão e estado de direito. No entanto, ao fornecer regras procedimentais claras e prazos, potencialmente coloca a contestação ao alcance de mais indivíduos, dando a mais pessoas um senso de agência ao reduzir os custos de informação de contestar decisões.[245]

---

[244] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024. p. 2006 (Tradução nossa).

[245] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024. p. 2032 (Tradução nossa).

A conjunção de procedimentos com cronogramas e processos detalhados com quais direitos e danos substantivos podem justificar um processo de contestação parece ser um caminho adequado.

Existe também uma discussão sobre se o objeto do direito de contestação de decisões tomadas por meio de sistemas de Inteligência Artificial deve ser apenas para decisões tomadas exclusivamente por meio de Inteligência Artificial ou se decisões hibridas também deveriam ser abrangidas:

> Como observado, existe um debate político ativo sobre se apenas decisões "exclusivamente" algorítmicas devem ser reguladas ou se as regulamentações devem ser aplicadas de forma mais ampla para cobrir decisões humanas facilitadas por máquinas. Enquanto o GDPR cobre apenas decisões automatizadas "exclusivamente" (embora a orientação tenha interpretado isso para incluir pelo menos a aprovação automática por humanos), a proposta do Escritório do Comissário de Privacidade do Canadá sugere eliminar o qualificador para cobrir o uso de IA mais amplamente. A legislação proposta nos Estados Unidos igualmente se aplicaria à IA que ajuda a tomar decisões impactantes. Como a regulamentação poderia ser facilmente evitada usando um humano para aprovar automaticamente o que é essencialmente um processo de IA, e por causa das preocupações sobre a competência humana para questionar ferramentas de IA, essa definição mais ampla é preferível.[246]

Optar por apenas garantir este direito para decisões tomadas exclusivamente por sistemas de Inteligência Artificial parece ser um caminho desenhado para a inefetividade, dificultando o atingimento do real objetivo da contestação – evitar injustiças – e facilitando para quem não pretende respeitar os direitos de os indivíduos afetados simplesmente envolver algum ser humano no caminho para escapar de qualquer necessidade legal.

### 4.4.1. Estruturas para a contestação

Ao se propor uma estrutura ou estruturas para a concretização do direito de contestação de decisões tomadas por meio de sistemas de Inteligência Artificial, não se pode deixar de lado as características destes sistemas, como trazido no capítulo dois deste trabalho. Contudo, não obstante os desafios que estes elementos trazem, é necessário que remédios concretos sejam ventilados e debatidos.

Conforme já antecipado, esta pesquisa sustenta a posição um devido processo judicial deveria ser visto como última *ratio*, uma vez que este, no modelo de prestação jurisdicional que ocorre hoje em dia, representaria altos custos financeiros e temporais.

---

[246] *Ibidem*, p. 2046 (Tradução nossa).

Dessa forma, poderia haver um escalonamento de estruturas para este processo de contestação. Este se iniciaria dentro da organização responsável pela decisão (o indivíduo poderia solicitar uma revisão humana da decisão automatizada e maiores informações sobre a decisão).

Fator fundamental aqui para este direito são os elementos já analisados do devido processo legal com a participação ativa do indivíduo. Sem ele, o risco de o direito a contestação ser fictício é grande. O processo de notificação, justificativa, exposição dos motivos decisórios e em quais dados a decisão se baseia e a oportunidade de ser ouvido ou de apresentar esclarecimentos ou argumentos contrários são peças-chave.

> Indivíduos não podem ter certeza de que a tomada de decisão está sendo aplicada de forma não arbitrária se eles não conseguirem entender a lógica de um sistema de tomada de decisão. E é improvável que os indivíduos se sintam respeitados por um direito de contestação que não forneça uma janela suficiente para a tomada de decisão — por meio de notificação, evidência e fornecimento de razões — para tornar possíveis desafios significativos.[247]

Olhando sob este prisma para a LGPD, já é possível se ver uma influência da RGPD com o estabelecimento de alguns elementos para esse devido processo legal. O art. 20 estabelece que o "controlador deverá fornecer, sempre que solicitadas, informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada, observados os segredos comercial e industrial".

Contudo, não é difícil se deduzir que este artigo não vem sendo muito aplicado. Empiricamente pode-se reconhecer que ele, por si só, não tem empoderado o titular do dado pessoal frente ao combate do uso discriminatório ou ilegal dos seus dados pessoais.

A previsão do caput do art. 20, de que o titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, não parece ser suficiente para que o indivíduo conheça o como exercer este direito. Como ocorrerá essa decisão? Por qual via? Quem revisa será a própria empresa, um terceiro árbitro neutro, um outra pessoa ou máquina?

Na RGPD, no seu art. 22, há um direito de intervenção humana, bem como um direito de o titular expressar o seu ponto de vista sobre a decisão automatizada. Mas também aqui, para que estes direitos sejam efetivos, o titular deve ter elementos para concretizar essa sua

---

[247] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024. p. 2035-2036 (Tradução nossa).

participação significativa. O processo vai envolver custos? Qual a duração? Regras procedimentais fazem falta.

Um procedimento desenvolvido para que o titular possa questionar diretamente a empresa/órgão/instituição controladora do tratamento sempre será indicado, especialmente por questões de eficiência e custo. Mas quando o processo interno não solucionar o questionamento de forma satisfatória? Um caminho poderia ser a obrigatoriedade de prazos, formatos e detalhamentos específicos para o oferecimento das respostas aos indivíduos e que a ausência deste cumprimento já fosse capaz de gerar automaticamente o envolvimento de um terceiro "imparcial" para se formar a triangulação e o devido processo legal.

Alguns exemplos interessantes de aplicação são a criação de legislações que forneçam listas abertas de exemplos ou processos de certificação ou códigos de conduta apoiados por supervisão regulatória setorial, a previsão de orientações de *soft law*[248] ou de participação externa como a exigência de consultas das partes afetadas ou especialistas que tragam definições mais específicas.

Caso interessante é da abordagem francesa aplicável ao setor público, que com o objetivo de garantir essa intervenção mais significativa, combina justificativa com supervisão humana e exige uma "compreensibilidade da decisão"[249]. Os funcionários públicos que utilizam algoritmos devem poder exercer um controle sobre eles e compreender e explicar o seu funcionamento para os sujeitos afetados.

Outro fator importante de se levar em consideração quando da construção de um direito de contestação, mesmo nesta etapa inicial "privada", é a possibilidade de existência de um tomador de decisão legítimo, um árbitro minimamente "neutro". Dentro de um devido processo legal, a existência de um terceiro neutro gera uma expectativa de decisões mais justas em razão da suposta limitação da discricionariedade e parcialidade.

Contudo, evidentemente, nem sempre essa figura ajuda na necessidade de se ter eficiência ou baixos custos no exercício desse direito. A velocidade e a escala são grandes preocupações – e muitas vezes a necessidade que motivou o uso de uma Inteligência Artificial para uma tomada de decisão. Assim, criar processos que possam ser lentos ou custosos é

---

[248] "*Soft law*" pode ser conceituado como instrumentos e regras não vinculantes no direito - internacional ou nacional - que têm efeitos práticos ou normativos sem, no entanto, possuir o caráter coercitivo das leis tradicionais (também chamadas de "*hard law*"). A *soft law* inclui declarações, princípios, diretrizes, códigos de conduta, recomendações e outras formas de compromissos que, apesar de não serem juridicamente obrigatórios, podem influenciar o comportamento, moldar práticas normativas e preencher lacunas legais.

[249] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024. p. 2038.

exatamente o que essas empresas privadas estão buscando evitar e não medirão esforços nos seus *lobbys*.

O atual modelo brasileiro de proteção de dados pessoais, no art. 20 da LGPD, não estabelece a necessidade de a decisão automatizada ser revisada por um humano (diferentemente da RGPD), o que significa, por uma interpretação estria, ser legal apenas máquinas revisando as decisões das máquinas. Isto mudaria caso o PL 2338/23 seja aprovado, uma vez que neste, no já mencionado art. 10º, quando a decisão, previsão ou a recomendação de sistema de inteligência artificial vier a produzir efeitos jurídicos relevantes ou que impactem de maneira significativa os interesses da pessoa sujeito da decisão, esta poderá solicitar a intervenção ou revisão humana.

Um modelo de revisão da decisão por uma outra Inteligência Artificial pode, de fato ser um processo menos custoso financeiramente de forma inicial, mas é bem provável que situações como vieses sejam difíceis de serem identificadas por uma outra Inteligência Artificial. O "humano no circuito" parece ser um caminho mais sustentável para estas situações como se verá mais adiante.

E, mesmo nas situações em que existe recurso e interesses suficientes para bancar este árbitro, a complexidade para a escolha desse árbitro neutro legítimo é maior do que simplesmente colocar um terceiro para intermediar a disputa. Por exemplo, no processo de exercício do direito de notificação de autores que estão diante de supostos plágios do DMCA, as plataformas *online* foram designadas como esse terceiro neutro que deveria intermediar as alegações de infração de direitos autorais.

Contudo, como já se sinalizado, estas plataformas não são propriamente árbitros neutros, uma vez que elas possuem interesses próprios de não atrair para si responsabilidades desnecessárias e limitar os seus custos com estes processos de disputas. O mesmo se pode dizer das plataformas de redes sociais, que criaram processos de moderação de conteúdo com modelos de disputas que são geridas por elas mesmas. Assim, quão neutro é o *Oversight Board* do Facebook para derrubar um conteúdo *fake* que viralizou e está fazendo todo mundo movimentar a plataforma?

Este *Oversight Board* foi criado para ser uma entidade com governança independente do Facebook para tomar decisões sobre a moderação de conteúdos da plataforma e com autoridade para revisar os casos específicos de conteúdo que foram removidos ou retirados pela empresa. O órgão teria a responsabilidade de ser uma dupla revisão dos conteúdos devendo equilibrar questões como liberdade de expressão, transparência e responsabilidade – seguindo os padrões da comunidade do Facebook, ou seja, padrões criados por eles mesmos.

Sim, teoricamente os membros deste Board são contratados para atuarem com independência, sendo especialistas multidisciplinares para analisarem as questões sob diferentes pontos de vista: jurídico, jornalístico, ético etc. Mas será que existe realmente esta independência? O Facebook teria um poder de veto sobre estas decisões? Será que as reuniões deveriam ser abertas e públicas? As respostas disponíveis não são tão claras ou seguras quanto se gostaria.

O Relatório *Corporate Accountability Index* 2020 da organização não governamental (ONG) *Ranking Digital Rights*[250], por exemplo, ao avaliar as vinte e seis plataformas digitais e de telecomunicação mais poderosas do mundo em suas políticas e práticas que afetam os direitos de liberdade de expressão e privacidade, classificou o Board do Facebook como "moderadamente transparente", ficando em quinto lugar no ranking.

Antes desse relatório, diversos fatos haviam acontecido que forçaram o Facebook a se movimentar e anunciar uma série de novas regras de moderação de conteúdo em resposta à disseminação de notícias falsas sobre as eleições presidenciais dos Estados Unidos e a pandemia da COVID-19, a proliferação de discursos de ódio e incitação à violência e até mesmo a multa que recebeu da FTC dos EUA por violações de privacidade no valor de cinco bilhões de dólares.

Assim, apesar da existência do Board e da formalização de um devido processo legal para revisão das decisões automatizadas de remoção de conteúdo, o relatório concluiu que: i) a empresa falhou na diligência devida aos direitos humanos, sem evidência de que produz avaliações de impacto sistemáticas da aplicação das suas políticas e termos de uso na implementação do seu algoritmo; e ii) que a transparência oferecida era limitada no que se refere às restrições de conteúdo e contas, bem como havia pouca informação sobre como o algoritmo utilizado nas apelações e sistemas de classificação[251].

Outras críticas a este modelo é a restrição da "jurisdição", que seria limitada demais a tipos específicos de conteúdo controverso ou prejudicial, ou a lentidão do processo, que mina a eficácia das decisões ou se a composição do Board possui adequadamente a diversidade de perspectivas e peculiaridades globais – o que impactaria na justiça das decisões.

E como poderia ser possível equilibrar tudo isso e ter um tomador de decisão que produzisse resultados mais legítimos? Poderia ser por meio de determinados incentivos?

---

[250] Relatório disponível em: https://rankingdigitalrights.org/index2020/companies/Facebook. Acesso em: 01 mar. 2024.

[251] Relatório disponível em: https://rankingdigitalrights.org/index2020/companies/Facebook. Acesso em: 01 mar. 2024.

Estes incentivos, por exemplo, teriam que promover a objetividade, a integridade e a independência dos árbitros. Formas de mitigar estes riscos, especialmente de conflito de interesses, poderiam ser a compensação justa e transparente destes terceiros, independência institucional (ex. nomeados por um órgão independente, sem vínculos com as partes interessadas no resultado ou o anonimato do terceiro) ou formas de proteção contra realizações e pressões externas.

Certamente a publicização das decisões e dos processos de deliberação também ajudariam a manter a transparência e auditabilidade do procedimento.

Assim, um modelo híbrido, público e privado pode ser uma solução interessante neste procedimento, bem como a existência de diversidade em termos de gênero, etnia, formação profissional, contexto social entre estes terceiros.

Margot Kaminski e Jennifer Urban[252] sugerem uma abordagem de que os reguladores considerem exigir que as decisões de revisão sejam tomadas ou até mesmo supervisionadas por um oficial independente dentro de uma empresa, ou exigir que as empresas relatem aos reguladores específicos ou forneçam proteções para denunciantes que desejam relatar sobre sistemas de contestação.

Ademais, uma outra maneira de limitar a discrição do tomador de decisões seria por meio de incentivos que levem até mesmo tomadores de decisões não neutros a serem atraídos pela busca da precisão e da legitimidade[253]. Ex. estabelecendo riscos de responsabilização para os próprios tomadores de decisão.

Importante que se sinalize, contudo, que propostas como a da mencionada, "*Accountable Algorithms*"[254] têm se afastado da ideia de direito individual ao devido processo no contexto da tomada de decisões por meio de sistemas de Inteligência Artificial. Lá são levantadas preocupações sobre os possíveis danos da transparência, que poderiam minar o aspecto de aviso prévio dos direitos individuais de devido processo.

Além deles, alguns estudiosos colocam os direitos individuais como ineficazes devido a questões de capacidade individual e acesso à justiça, observado que o devido processo geralmente se aplica mais precisamente em casos de ação estatal. E autores como Aziz Huq chegam a defender que direitos de devido processo não devem se aplicar mesmo à ação estatal:

---

[252] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024. p. 2032.

[253] *Ibidem*.

[254] KROLL, Joshua A. *et al.* Accountable Algorithms. Forthcoming, **Fordham Law Legal Studies Research Paper**, v. 165, n. 2765268. University of Pennsylvania Law Review, 2017. Disponível em: https://ssrn.com/abstract=2765268. Acesso em: 05 fev. 2024.

> Aziz Huq foi mais longe, argumentando que os direitos de devido processo não deveriam se aplicar nem mesmo à ação estatal — ou melhor, que em vez de desafios individualizados à tomada de decisão algorítmica, os indivíduos deveriam ser capazes de desafiar se o algoritmo é sistemicamente "bem calibrado".
>
> Há exceções limitadas à tendência. Rory Van Loo, por exemplo, pede um sistema de apelações complexo para ser aplicado à tomada de decisão de plataformas, envolvendo precedentes transparentes estabelecidos por árbitros humanos neutros. No contexto de julgamento criminal, Andrea Roth pede um direito de confrontação em relação a evidências criadas por máquinas, incluindo ferramentas que vão desde testes em tribunal até o interrogatório cruzado de programadores responsáveis.[255]

Estes estudos, assim, não vão em linha com um direito geral e amplo de contestar as decisões tomadas por meio de sistemas de Inteligência Artificial, envolvendo outros caminhos a partir de eventuais falhas ou dificuldades encontradas no processo de contestação das decisões com base em um direito específico de proteção individual.

Contudo, não obstante este trabalho defender como necessário um direito geral e amplo de contestar, por mais que ineficiências e falhas existam, estes mencionados estudos auxiliam na identificação e compreensão destas falhas aumentando as possibilidades de diminuição da probabilidade de concretização dos riscos.

Vale observar que atualmente muitas empresas de tecnologia por pressão externa parecem ter seguido apenas um padrão de contestação, e outras até criaram regras próprias, mas dentro de um contexto de autorregularão. Na prática o que se tem visto é a ausência no cumprimento das suas próprias regras de contestação.

A existência de regras gerais - e não criadas pelas próprias empresas - poderia empoderar o indivíduo. Caso a regra determinasse uma resposta em até cinco dias e a empresa não respondesse, a ausência de resposta já seria um descumprimento legal que legitimaria o indivíduo à acionar as autoridades fiscalizadoras ou o próprio poder judiciário. E a inclusão deste terceiro imparcial no processo já seria suficiente para incomodar e pressionar uma mudança no comportamento das reguladas.

Passo seguinte, caso o sujeito considere que o processo privado de solicitação de revisão e contestação de alguma decisão que o afetou não tenha sido satisfatório, o segundo degrau seria levar a questão para uma segunda estrutura, uma esfera administrativa pública de contestação da decisão.

---

[255] KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. p. 21-30. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024. p. 1986 (Tradução nossa).

Esta esfera deveria ser apta não apenas para a análise material do que se contesta, mas igualmente a verificação de se o procedimento privado, a autorregulação do "tomador da decisão" foi devida – o que pressupõe a existência de regras mínimas procedimentais que deveriam ser seguidas para garantia de uma segurança jurídica e alinhamento das expectativas.

Para que essa esfera administrativa pública funcione, é fundamental que ela supere as mencionadas críticas feitas aos riscos da autorregulação, se mostrando apta a "sanar" os mencionados desafios que uma contestação feita aos próprios desenvolvedores, vendedores, fornecedores ou utilizadores de sistemas de tomada de decisão por meio de Inteligência Artificial possuem.

Assim uma autoridade pública poderia oferecer um fórum mais neutro e imparcial para a revisão de decisões automatizadas, reduzindo o risco de conflitos de interesse que podem ocorrer quando a revisão é feita pela própria organização, bem como poderia ser garantida a observância de um processo formal e estruturado, aumentando a transparência e a confiança no sistema.

Ademais, esta estrutura poderia ser composta por especialistas em Inteligência Artificial, direito e outras disciplinas relevantes, que poderiam fornecer uma análise mais profunda e técnica das decisões contestadas. Isso asseguraria que os aspectos complexos dos sistemas de Inteligência Artificial fossem devidamente considerados de forma equilibrada, aumentando até mesmo a legitimidade das decisões tomadas por esses sistemas, já que as decisões finais não dependeriam exclusivamente da entidade privada envolvida, mas também de um órgão independente. A presença de uma estrutura administrativa pública aumenta também a transparência do processo de revisão, permitindo que as decisões e os processos sejam auditados e monitorados publicamente.

Essa segunda estrutura poderia também atuar como uma camada de proteção adicional contra abusos e práticas discriminatórias, com roupagem fiscalizatória e regulatória, assegurando que as decisões automatizadas estejam em conformidade com a lei.

Por fim, ao proporcionar uma via administrativa pública para contestação, evita-se que os indivíduos afetados precisem recorrer diretamente ao sistema judiciário, que pode ser custoso e demorado. Isso torna o processo de contestação mais acessível e eficiente.

Contudo, para que todos esses elementos acima sejam concretizados, o caminho mais indicado parece o de "autoridades" públicas e não apenas uma estrutura centralizada. A Inteligência Artificial é multifacetada, podendo ser aplicada e afetar diversos setores com características muito diferentes um dos outros. Alguns setores possuem alta concentração de agentes de Inteligência Artificial, ou sistemas pulverizados com muitos *players*, alguns setores

possuem usos que representam alto risco no uso da Inteligência Artificial outros riscos baixos etc. Sem desconsiderar que alguns setores já possuem até mesmo estruturas sólidas de autoridades regulatórias.

A revisão sobre uma decisão automatizada deve, necessariamente, respeitar as peculiaridades e contextos regulatórios setoriais. Diferentes setores têm características, necessidades e desafios específicos. Agentes regulatórios setoriais desenvolvem expertise e conhecimento sobre as particularidades do setor que regulam, tornando-se mais capazes de lidar com as complexidades e nuances das decisões automatizadas dentro de seus respectivos domínios.

Quanto aos processos de contestação, reguladores setoriais podem adaptar suas abordagens e critérios de revisão às particularidades do setor, criando processos que são mais adequados e relevantes para os tipos de decisões que estão sendo contestadas. Bem como cada setor entenderia e poderia aplicar de forma mais apropriada as regulamentações setoriais. Isso pode levar a resultados mais precisos, eficazes e justos.

Ademais, um ambiente mais democrático poderia ser fomentado com uma maior representatividade das partes interessadas, como empresas, consumidores e especialistas do setor, no processo de regulamentação e contestação. E estruturas setoriais podem promover mais facilmente a inovação regulatória, com cada setor desenvolvendo e testando novas abordagens regulatórias que podem ser adaptadas e adotadas por outros setores conforme necessário.

A proposta do PL 2338/23 de criação de um Sistema Nacional de Regulação e Governança de Inteligência Artificial (SIA) parece poder complementar a ideia trazida acima, permitindo que diferentes autoridades e instituições setoriais possam atuar, mas garantindo que exista alguma uniformidade se supervisão sem sobrecarga significativa.

Essa seria uma estrutura híbrida que valorizaria o trabalho de regulação e de fomento que já feito pelas agências reguladoras setoriais existentes, mas também traria uma supervisão que é igualmente benéfica para garantir que exista um mínimo de coordenação entre os reguladores setoriais garantindo uma agenda uniforme no país sobre o tema. Neste sentido, o Governo lançou em julho de 2024 o Novo Plano Brasileiro de Inteligência Artificial (PBIA) 2024-2028, o que demonstra uma ação no sentido de um tratamento nacional sobre o tema. Observa-se, contudo, que questões sobre a contestabilidade das tecnologias não foram contempladas pelo plano.

Importante observar que a atuação dos agentes reguladores já é uma realidade independente da regulação geral da Inteligência Artificial proposta pelo mencionado PL. A

ANPD, por exemplo, proferiu em julho de 2024 decisão emblemática determinando a suspensão cautelar do tratamento de dados pessoais para treinamento da Inteligência Artificial da Meta. A decisão teve como foco o uso de dados pessoais e guardou coerência com as decisões e posicionamentos posteriores da ANPD no que diz respeito ao uso dos dados pessoais.[256]

O CNJ, por sua vez, também já possui atuação na regulamentação do uso da Inteligência Artificial dentro do Poder Judiciário. É o caso da Resolução 332/2020[257] que dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Judiciário.

Não obstante mencionada resolução não tratar da competência ou de procedimentos a administrativos a serem feitos no âmbito do CNJ para a contestação de decisões automatizadas, o Capítulo 9º já traz regras de qualidade e governança, inclusive preocupação com o uso dos dados pessoais:

> Art. 9o Qualquer modelo de Inteligência Artificial que venha a ser adotado pelos órgãos do Poder Judiciário deverá observar as regras de governança de dados aplicáveis aos seus próprios sistemas computacionais, as Resoluções e as Recomendações do Conselho Nacional de Justiça, a Lei no 13.709/2018, e o segredo de justiça.[258]

Um caminho poderia ser a determinação de que estes agentes setoriais desenvolvessem uma regulamentação e mecanismos robustos – procedimentais e materiais – para poder atuar como esta instância administrativa apta a decidir sobre eventual contestação de decisão de sistema de tomada de decisão automatizada.

Por fim, em respeito ao princípio da inafastabilidade da jurisdição, que tem previsão no artigo 5º, inciso XXXV, da Constituição Federal vigente no Brasil, o Poder Judiciário seria a terceira estrutura, o último degrau, que poderia ser chamado a decidir sobre se houve ou não lesão ou ameaça a direito na decisão contestada.

---

[256] AGÊNCIA NACIONAL DE PROTEÇÃO DE DADOS (ANPD). *ANPD determina suspensão cautelar do tratamento de dados pessoais para treinamento da IA da Meta*. 2023. Disponível em: https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-determina-suspensao-cautelar-do-tratamento-de-dados-pessoais-para-treinamento-da-ia-da-meta. Acesso em: 30 ago. 2024.

[257] CONSELHO NACIONAL DE JUSTIÇA (CNJ). *Resolução nº 332, de 21 de agosto de 2020*. Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências. 2020. Disponível em: https://atos.cnj.jus.br/files/original191707202008255f4563b35f8e8.pdf. Acesso em: 15 ago. 2024.

[258] CONSELHO NACIONAL DE JUSTIÇA (CNJ). *Resolução nº 332, de 21 de agosto de 2020*. Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências. 2020. Disponível em: https://atos.cnj.jus.br/files/original191707202008255f4563b35f8e8.pdf. Acesso em: 15 ago. 2024.

Aqui, poderia ser proposta uma via judicial específica de contestação das decisões a depender do impacto, das questões técnicas ou até mesmo do setor.

Contudo, a intervenção judicial em decisões automatizadas e tecnicamente complexas exige uma especialização cada vez maior dos julgadores, considerando não apenas conhecimento sobre o impacto social e econômico dessas decisões, mas também as nuances técnicas que as envolvem.

De uma forma ainda mais ampla, para enfrentar esses novos desafios, é essencial investir na capacitação contínua de julgadores, advogados e peritos. Estes devem estar equipados com conhecimentos básicos sobre Inteligência Artificial, *machine learning*, e *big data*, além de serem capacitados a entender o funcionamento dos algoritmos. Só assim seria possível se cogitar uma análise crítica e informada das evidências e argumentos apresentados e que as partes estivessem aptas a dialogar com peritos técnicos e entender as complexidades subjacentes às evidências apresentadas.

# 5 CONCLUSÃO

A integração de tecnologias avançadas com alta conectividade, resultando em uma transformação digital disruptiva, trouxe a necessidade de repensar as relações interpessoais e o desenvolvimento humano em interação com a tecnologia. E a combinação delas com o grande volume de dados pessoais coletados diariamente, muitas vezes sem o conhecimento ou consentimento dos titulares, introduziu uma nova lógica social e econômica. Isso impõe desafios complexos que o Direito precisa enfrentar. Para entender esse fenômeno, é crucial compreender o funcionamento dos algoritmos e a evolução do seu uso e o seu principal insumo, os dados.

Os dados pessoais passaram a ser cada vez mais extraídos de forma quase ubíqua, por meio de sensores inteligentes, dispositivos conectados à internet, os dispositivos vestíveis ou *wearables*, carros automatizados, nanopartículas que patrulham o corpo, dispositivos inteligentes para o monitoramento doméstico, câmeras e tecnologias de geolocalização, etc.

Com isso a coleta de dados pessoais se tornou algo praticamente permanente e o seu uso para alimentar sistemas de tomada de decisão se mostrou algo valioso para diversos modelos de negócios e para a construção da nova lógica de acumulação, a do capitalismo de vigilância. A privacidade e a proteção em relação ao uso dos dados pessoais surgem como uma das grandes preocupações das últimas décadas uma vez que se percebe o quanto que a ausência destas pode trazes consequências sérias no âmbito individual e social.

Apesar dos impressionantes avanços tecnológicos, o progresso científico nem sempre é acompanhado de progresso social e mecanismos de tomada de decisão por meio de sistemas de Inteligência Artificial podem gerar impactos sociais profundos, que podem ser antiéticos e perigosos, colocando em risco diversos direitos humanos fundamentais.

Sistemas algorítmicos, não são meramente técnicos ou neutros. Eles incorporam e refletem escolhas sociais e políticas que podem ter impactos significativos na vida das pessoas. Isso inclui desde decisões sobre quem recebe determinados serviços, até a forma como os recursos e oportunidades são distribuídos.

Até mesmo sistemas criados com a intenção de facilitar a vida dos usuários e com objetivos éticos podem, inadvertidamente, tomar decisões indesejadas, preconceituosas e discriminatórias. Uma das fontes de problemas é quando informações enviesadas, distorcidas ou privadas, sem os devidos filtros, são usadas para alimentar essas decisões. Isso se agrava

quando a decisão é automatizada e há falta de controle sobre os parâmetros aplicados, comprometendo a qualidade e os resultados das decisões.

Com o avanço tecnológico sendo inevitável e desejável, é crucial que esse progresso se concentre no bem-estar humano. É necessário pensar cuidadosamente no controle dos impactos sociais, econômicos e políticos da Inteligência Artificial, desde sua concepção até o uso final, para garantir que os direitos fundamentais sejam respeitados e priorizados.

Assim, podemos ter como consequência vieses algorítmicos discriminatórios e a decisão automatizada pode tanto sistematizar quanto ocultar a discriminação. Devido à dificuldade de prever os efeitos de uma regra complexa derivada de um sistema que se utiliza de *machine learning*, reguladores e os próprios indivíduos impactados podem ser incapazes de perceber os efeitos discriminatórios ou atentatórios contra direitos de alguma regra decisória.

O *design* de sistemas de decisão automatizada pode apresentar falhas e tendências que levam à identificação de vieses algorítmicos. Muitas das aplicações matemáticas que fomentam a economia dos dados são baseadas em escolhas feitas por humanos, muitas vezes boas intenções, mas que acabam resultando em preconceitos, equívocos e vieses nos sistemas de *software*.

Estes vieses podem surgir nos resultados produzidos por algoritmos devido a vários fatores, pela seleção de dados (quando os conjuntos de dados usados para treinar os algoritmos não são representativos da população ou do fenômeno que estão tentando modelar), de viés de programação (o próprio algoritmo exibe padrões discriminatórios devido à maneira como foi projetado ou aos dados usados para treiná-lo), viés de implementação (quando há falhas na implementação prática do algoritmo, como erros de codificação ou configuração incorreta, resultando em resultados injustos ou imprecisos) ou até viés de retroalimentação (quando os resultados produzidos pelo algoritmo são usados para tomar decisões subsequentes, criando um ciclo de retroalimentação que reforça e amplifica quaisquer vieses presentes nos dados ou no próprio algoritmo).

A opacidade, por sua vez, na qual os detalhes internos do funcionamento do sistema não são visíveis, pode mascarar manipulações, a perda de liberdade e da privacidade, ofensas a dignidade e a autonomia humana.

E mesmo quando se percebe, ou há indícios de que o resultado de uma decisão está ofendendo algum direito, as dificuldades atualmente enfrentadas pelos indivíduos ao tentarem contestar decisões automatizadas não são jurídicas nem tecnicamente endereçadas, mesmo com a existência da LGPD, que traz formalmente o direito de revisão de decisões automatizadas. A opacidade e a complexidade técnica dessas decisões que muitas vezes limitam a existência de

um “direito de defesa” efetivo muitas vezes têm mecanismos tecnológicos para serem mitigados, faltando “apenas” mecanismos legais ou interesse político.

O problema central que este trabalho se propôs a explorar, assim foi o poder de contestar decisões tomadas por meio de Inteligência Artificial à luz da proteção do uso dos dados pessoais em sistemas que deveriam ser construídos para serem confiáveis. À medida que algoritmos e sistemas automatizados se tornam mais presentes nos processos decisórios e coletam cada vez mais dados pessoais, a capacidade dos indivíduos de questionar, revisar e, se necessário, reverter essas decisões se torna essencial. É crucial entender como as estruturas legais e regulamentares podem ser moldadas para proteger os direitos fundamentais contra decisões automatizadas injustas, enviesadas ou prejudiciais.

O poder de contestações também se entrelaça com a necessidade de transparência e a possibilidade de influências externas, como interesses econômicos, que podem afetar as decisões automatizadas.

As ferramentas de Inteligência Artificial devem operar de forma ética e justa, respeitando os direitos fundamentais e sendo transparente e contestável.

Evidentemente que decisões tomadas unicamente por humanos também podem ser ilícitas, antiéticas ou discriminatórias – com vieses e ruídos - e, eventualmente, testes feitos com base em resultados de tomada de decisão automatizada podem até apresentar números melhores nestes quesitos do que decisões humanas.

Contudo, nas decisões tomadas por meio de sistemas de Inteligência Artificial parece haver uma “dispersão da responsabilidade”, a decisão e o consequente resultado parecem não ser de responsabilidade de ninguém e não poder, assim, ser questionado para ninguém. Com isso a capacidade dos indivíduos de trazer algum questionamento, revisar e, se necessário, reverter as decisões ou o impacto destas decisões se torna um novo desafio que merece especial atenção. E a reversibilidade de determinadas tecnologias pode ser impossível.

Ademais, existe o fundado receio de que este tipo de processo de tomada de decisão leva o humano a dar como se fosse certa a sua validade e, com isso, reduzem as suas responsabilidades por exemplo de investigar, analisar e determinas questões envolvidas[259]. Haveria um viés contra a impugnação da decisão.

A construção dos sistemas de Inteligência Artificial deve assim seguir em direção a uma "Trustworthy AI", ou seja, uma Inteligência Artificial confiável. Esse conceito se refere à

[259] LEE, Kai-Fu. **Inteligência artificial:** como os robôs estão mudando o mundo, a forma como amamos, nos relacionamos, trabalhamos e vivemos. 1 ed. Rio de Janeiro: Globo Livros, 2019.

construção de uma Inteligência Artificial centrada na pessoa e no bem-estar dos indivíduos, respeitando os direitos fundamentais, regulamentações e princípios éticos, e orientada para um fim justo e ético. Ou seja, uma Inteligência Artificial Sólida, social e tecnicamente (menos propensa a cometer erros graves ou a tomar decisões injustas), Legal (que respeite toda a legislação e regulamentação aplicáveis) e Ética (devendo observar assim não apenas a lei mas também princípios e valores éticos).

Para definir uma Inteligência Artificial de confiança, componentes essenciais, os princípios éticos, devem estar presentes no desenvolvimento e uso da Inteligência Artificial. Os sistemas de tomadas de decisão por meio de Inteligência Artificial devem ser desenvolvidos e utilizados com o ser humano como foco principal, respeitando a autonomia dos indivíduos e garantindo que as tecnologias atuem como facilitadoras, e não como controladoras das decisões humanas.

Deve-se assegurar que eles não causem danos aos indivíduos, prevenindo riscos e implementando medidas de segurança para evitar consequências negativas e devem operar de forma justa, evitando discriminações e preconceitos, e garantindo que todos os indivíduos sejam tratados de maneira equitativa.

Para a real confiança os sistemas de Inteligência Artificial também devem ser transparentes em seu funcionamento, permitindo que os usuários compreendam como as decisões são tomadas. Além disso, deve ser possível explicar as decisões de forma clara e compreensível. E para isso é crucial que haja mecanismos para responsabilizar os desenvolvedores e operadores de sistemas de Inteligência Artificial, garantindo que eles possam ser responsabilizados por decisões prejudiciais ou injustas.

Uma Trustworthy AI está intimamente ligada à possibilidade de contestar as decisões tomadas por esses sistemas. A governança algorítmica é essencial para garantir que os riscos associados ao uso de sistemas de tomada de decisão automatizadas sejam minimizados e que os sistemas operem de maneira ética.

Mas essa capacidade de contestar as decisões automatizadas deve ser real e não apenas utópica ou falaciosa. Essa capacidade é crucial para a confiança nos sistemas, uma vez que permite que os indivíduos tenham recursos para questionar e, se necessário, reverter decisões que possam ser injustas ou prejudiciais. E a relevância dessa capacidade não é apenas repressiva, após a tomada de decisão, mas sim preventiva, uma vez que as soluções são aplicadas desde a concepção da tecnologia, mitigando os mencionados riscos.

Garantir que as pessoas possam controlar e contestar o uso de suas informações é crucial para proteger seus direitos fundamentais. O direito de contestação passa, assim, pela garantia

da autodeterminação informativa (ao controle que os indivíduos devem ter sobre seus próprios dados e como esses dados são utilizados).

Assim, a análise do devido processo legal na contestação de decisões automatizadas é central. O devido processo legal implica que qualquer decisão tomada por sistemas de Inteligência Artificial que afete os direitos dos indivíduos deve ser passível de revisão e contestação. Isso inclui a necessidade de mecanismos transparentes e acessíveis para que os indivíduos possam entender como as decisões foram tomadas e, se necessário, recorrer contra elas.

Neste sentido, foram identificados e discutidos diversos arquétipos de contestação de decisões de Inteligência Artificial. Esses arquétipos variam desde simples pedidos de explicação até complexos processos de revisão judicial. A diversidade desses arquétipos demonstra a necessidade de diferentes níveis de contestação, dependendo da complexidade e do impacto das decisões automatizadas.

Diante de toda a pesquisa apresentada, a ausência de um direito de contestação das decisões tomadas por meio de sistemas de Inteligência Artificial - com ou sem algum envolvimento humano - parece não ser um caminho prudente. As dificuldades operacionais e práticas para garantir a existência e aplicação deste direito não podem ser vistas como barreiras intransponíveis incapacitantes.

Contudo, tecnicamente, mecanismos de aprendizado de máquina ou *machine learning*, por exemplo, com a criação de algoritmos capazes de aprender automaticamente a partir de dados, substituindo a necessidade de programar códigos específicos para cada tarefa, permitindo que as máquinas desempenhem funções e gerem soluções de forma inteligente sem serem explicitamente programadas para isso, podem ser impossíveis de contestar tecnicamente se não já forem desenvolvidos para serem passíveis de contestação.

Assim, o desenvolvimento de um processo para contestar decisões tomadas por sistemas de Inteligência Artificial deve abordar tanto os aspectos técnicos, políticos quanto os legais. E, quanto a este último, com atenção aos aspectos procedimentais e os substantivos, garantindo que o processo seja relevante, capacite os indivíduos e promova os valores que se busca alcançar em um Estado de Direito. Ao incluir diretrizes claras, regras significativas e a possibilidade de envolvimento das partes interessadas, reguladores podem melhorar a eficácia e a legitimidade dos mecanismos de contestação.

Neste trabalho, assim, se defende que para o uso de sistemas de tomada de decisão automatizada que tenha impacto relevante sobre a vida das pessoas, o seu desenvolvimento deve envolver a construção de um mapa auditável dos seus impactos em todas as etapas do

sistema. Os indivíduos impactados pela decisão precisam ser empoderados para que não tenham os seus dados e as suas vidas cada vez mais manipulados por interesses externos. E, para isso, parecem ser necessárias regras de contestação e não apenas a construção de padrões de contestação.

Como já trazido, a regra de contestação tem a vantagem de uma maior clareza, o que facilita a compreensão sobre o seu direcionamento e, teoricamente, facilita a conformidade com ela, inclusive com menores custos associados. O desenvolvimento e o uso destes sistemas assim, teriam uma "trilha" sobre como deveriam ser.

E, como para se ter uma regra de contestação efetiva, a possibilidade de contestação deve ser construída bem antes do surgimento da necessidade da contestação, esta "trilha" deve permear todo o procedimento, inspirado no princípio da privacidade incorporada do *privacy by design* desenvolvido pela Ann Cavoukian.

É assim que a mencionada "opacidade" técnica, por exemplo, poderia ser mitigada com novas forma de engenharia de programação para a sua redução ou o uso de ferramentas como as mencionadas *Shapley Additive Explanations* (SHAP), a Inteligência Artificial Semântica, a *Explainable* AI (XAI), o *Local Interpretable Model-Agnostic Explanations* (LIME) ou o *Deep Learning Important FeaTures* (DeepLIFT). Explicitações inteligíveis dos complexos racionais por detrás das decisões automatizadas devem ser premissas.

Se um sistema foi construído de forma a ser uma "caixa preta", sem a mínima possibilidade de entendimento do racional do uso dos dados pessoais dos sujeitos nas decisões automatizadas e, consequentemente, sem a real possibilidade de contestação dos seus resultados, talvez este sistema não deveria estar sendo utilizado para alguma função que impacte a vida das pessoas de forma significativa.

Sem o conhecimento dos fatores que baseiam a decisão do algoritmo, é impossível saber se se está diante de decisões ilícitas, discriminatórias ou antiéticas.

É encarada com naturalidade a necessidade de uma simples caixa de suco conter um rótulo com a descrição de todos os ingredientes utilizados na sua fabricação. E por qual motivo não seria encarada com a mesma naturalidade a necessidade das ferramentas tecnológicas que utilizamos também nos oferecerem transparência sobre os "insumos". De ambos os lados, inclusive, podemos ter um produto que ponha em risco a saúde do consumidor.

Não por outro motivo Ana Frazão defende ser necessário o reconhecimento de que "sistemas algorítmicos que impactam na vida das pessoas envolvem escolhas sociais e

políticas"[260] e que o *design* destes sistemas deve se equiparar para fins de cuidados e estratégia à formulação de políticas sociais. A solução, ou ao menos a mitigação, para estes desafios, assim, é multidisciplinar e conjunta.

Neste sentido, um sistema de tomada de decisão deve ser sempre justo, consistente, previsível e racional quando aplicado em diferentes indivíduos. E para prevenir arbitrariedades, identificar preconceitos ou até mesmo possibilitar o controle de alguma maneira das decisões por eles tomadas, as proteções processuais relacionadas a contestação trazem a obrigação do tomador de decisão de demonstrar um compromisso que seja examinável com um determinado resultado, apresentando as razões, os motivos para isso.

Dessa forma, este trabalho conclui que para a real existência de um direito de contestação, são necessárias regulamentações no sentido de estabelecer a obrigatoriedade de opções de soluções tecnológicas que permitiriam tecnicamente a auditabilidade e contestação da decisão. E estas opções, especialmente considerando a complexidade e autonomia das decisões tomadas por meio de *machine learning* e redes neurais, conforme trazido, precisam ser pensadas já *by desing*. Focar no "acesso ao código" é insuficiente e irreal para um efetivo "controle" e entendimento sobre as saídas de um sistema de Inteligência Artificial. Ou seja, defende-se a existência de desenhos legais ou regulatórios sobre o próprio desenvolvimento destes sistemas, indo além de escolhas sobre riscos que seriam inadmissíveis.

É inegável que atualmente o impacto das decisões de sistemas de Inteligência Artificial de empresas privadas é muitas vezes maior e afeta muito mais pessoas do que muitas decisões de órgãos governamentais. E, por mais que o direito de contestação tenha como foco uma ação *ex post*, ou seja, posterior a eventual dano que tenha sido causado a um indivíduo ou a uma coletividade, a simples existência dele, combinado com o modelo *by design*, já é capaz de fazer a construção e uso da tecnologia ser mais responsável.

Ademais, o uso destas tecnologias deve estar sempre amparado por processos já previamente estabelecidos de possibilidade de contestação das decisões. E o desenvolvimento desses processos envolve uma análise mais abrangente das estruturas de incentivo (como os mencionados incentivos indiretos que levam os mecanismos de busca a removerem conteúdo), a transparência durante todo o processo, a determinação de quem será responsável pela tomada de decisões e o contexto regulatório.

---

[260] FRAZÃO, Ana. Discriminação algorítmica: ciência dos dados como ação política. JOTA, 2021. Disponível em: https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/discriminacao-algoritmica-ciencia-dos-dados-como-acao-politica-21072021?amp. Acesso em: 06 jul. 2024.

Considerando que a estruturação de um sistema para contestar decisões tomadas por Inteligência Artificial deve considerar as características específicas desses sistemas e os desafios que apresentam, é sugerido que um processo escalonado de contestação inicie internamente, dentro da organização responsável pela decisão, permitindo ao indivíduo ou coletividade afetada solicitar uma revisão humana.

Para garantir que o direito à contestação seja efetivo, é fundamental a transparência sobre o processo decisório, divulgação de canais, prazos para respostas, requisitos de informação, o fornecimento de justificativas, e a possibilidade de participação ativa do indivíduo afetado.

Além do processo interno, é sugerida a criação de uma esfera administrativa pública para a contestação, que ofereça um ambiente mais neutro e imparcial para a revisão das decisões automatizadas. Essa estrutura poderia envolver estruturais setoriais já existentes, com especialistas em diversas áreas, proporcionando uma análise mais técnica e equilibrada das decisões contestadas. A existência de autoridades públicas para revisão aumentaria a confiança nos sistemas, além de proporcionar uma camada adicional de proteção contra práticas discriminatórias e abusos.

Tecnicamente deve ser exigido um registro detalhado do uso dos dados pessoais nas decisões, com todos os dados de entrada, os parâmetros do modelo e os passos do processo de tomada de decisão. Estas informações devem ser documentadas e compreensíveis, permitindo que especialistas e os sujeitos afetados entendam a lógica por trás das decisões (isso inclui informações sobre a arquitetura do modelo, os métodos de treinamento e os critérios de avaliação).

E, finalmente, como último recurso, o Poder Judiciário poderia intervir para avaliar se houve lesão ou ameaça a direitos na decisão contestada. No entanto, a complexidade das decisões automatizadas requer uma crescente especialização dos julgadores, que devem ser capacitados para lidar com as nuances técnicas envolvidas. Esse modelo híbrido, que integra processos internos, administrativos e judiciais, visa criar um sistema de contestação mais eficiente, acessível e justo, adaptado às peculiaridades de diferentes setores e contextos regulatórios.

**REFERÊNCIAS**

AGÊNCIA NACIONAL DE PROTEÇÃO DE DADOS (ANPD). ANPD determina suspensão cautelar do tratamento de dados pessoais para treinamento da IA da Meta. 2023. Disponível em: https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-determina-suspensao-cautelar-do-tratamento-de-dados-pessoais-para-treinamento-da-ia-da-meta. Acesso em: 12 ago. 2024.

ALBIANI, Christine. **Responsabilidade Civil e Inteligência artificial:** Quem responde pelos danos causados por robôs inteligentes? Instituto de Tecnologia e Sociedade do Rio. Rio de Janeiro, 2019. Disponível em: https://itsrio.org/wp-content/uploads/2019/03/Christine-Albiani.pdf. Acesso em: 13 out. 2023.

ALMEIDA NETO, Amaro Alves de. Dano existencial e tutela da dignidade da pessoa humana. **Revista de Direito Privado**, [s.l.], v. 6, n. 24, p. 21-53, out./dez. 2005.

ALVES, Ariane. Uso de algoritmos em análise de currículo pode gerar seleção enviesada. Exame. 2018. Disponível em: https://exame.com/tecnologia/uso-de-algoritmos-em-analise-de-curriculo-pode-gerar-selecao-enviesada/. Acesso em: 13 jun. 2022.

ANANNY, Mike; CRAWFORD, Kate. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. **New media & society**, [s.l.], v. 20, n. 3, 2018.

AUERBACH, David. The Programs That Become the Programmers. Slate. 2015. Disponível em: https://slate.com/technology/2015/09/pedro-domingos-master-algorithm-how-machine-learning-is-reshaping-how-we-live.html. Acesso em: 21 jun. 2020.

AUGUSTO, Victor. Quais são os limites éticos da Inteligência Artificial? UPlab. 2019. Disponível em: http://uplab.cc/future-sight/quais-sao-os-limites-eticos-da-inteligencia-artificial. Acesso em: 13 jun. 2022.

BABO, Gustavo Schainberg S. Discriminação algorítmica: origens, conceitos e perspectivas regulatórias (parte 1). DTIBR. Belo Horizonte, 2020. Disponível em: https://www.dtibr.com/post/discrimina%C3%A7%C3%A3o-algor%C3%ADtmica-origens-conceitos-e-perspectivas-regulat%C3%B3rias-parte-1. Acesso em: 10 set. 2023.

BAHIA, Saulo José Casali. Digital justice, robot judges and new challenges and perspectives for the justice: issues posed by AI in the Brazilian court system. **European Review of Public Law**, v. 36, n. 1, 2024.

BARBOSA, Mafalda Miranda. Inteligência Artificial, e-persons e direito: desafios e perspectivas. **Revista Jurídica Luso-Brasileira**, [s.l.], v. 3, n. 6, p. 1475-1503, 2017. Disponível em: http://www.cidp.pt/revistas/rjlb/2017/6/2017_06_1475_1503.pdf. Acesso em: 10 set. 2023.

BAROCAS, Solon; HARDT, Moritz; NARAYANAN, Arvind. **Fairness and Machine Learning:** Limitations and Opportunities. Massachusetts Institute of Technology: The MIT Press, 2023. Disponível em: https://fairmlbook.org/pdf/fairmlbook.pdf. Acesso em: 10 jun. 2024.

BERTAGNOLLI, Danielle; RIZZOTO, Felipe; TONIAL, Maira Angélica Dal Contre. As relações de trabalho e a automação industrial: reflexões sobre os aspectos históricos, econômicos, conceituais e sociais. **Revista Justiça do Direito**, [s.l.], v. 24, n. 1, 2011. Disponível em: http://seer.upf.br/index.php/rjd/article/view/2149. Acesso em: 08 jun. 2022.

BIONI, Bruno; LUCIANO, Maria. O princípio da precaução na regulação de inteligência artificial: seriam as leis de proteção de dados o seu portal de entrada. **Inteligência Artificial e Direito.** São Paulo: Thomson Reuters Brasil, p. 207-231, 2019. Disponível em: https://brunobioni.com.br/home/wp-content/uploads/2019/09/Bioni-Luciano_O-PRINCI%CC%81PIO-DA-PRECAUC%CC%A7A%CC%83O-PARA-REGULAC%CC%A7A%CC%83O-DE-INTELIGE%CC%82NCIA-ARTIFICIAL-1.pdf. Acesso em: 10 set. 2023.

BRASIL. **Decreto nº 9.854, de 25 de junho de 2019**. Institui o Plano Nacional de Internet das Coisas e dispõe sobre a Câmara de Gestão e Acompanhamento do Desenvolvimento de Sistemas de Comunicação Máquina a Máquina e Internet das Coisas. Diário oficial da União, Brasília, 2019.

BRASIL. Diretora da ANPD defende protagonismo da Autoridade na regulamentação da IA. Ministério da Justiça e Segurança Pública. 2024. disponível em: https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-determina-suspensao-cautelar-do-tratamento-de-dados-pessoais-para-treinamento-da-ia-da-meta/SEI_0130047_Voto_11.pdf. Acesso em: 07 jul. 2024.

BUCCO, Rafael. Para organizações civil, algoritmos não devem ser sigiloso. Tele.síntese. 2017. Disponível em: https://www.telesintese.com.br/para-organizacoes-civis-algoritmos-nao-devem-ser-sigilosos/. Acesso em: 10 set. 2023.

CARDOSO, Letycia. Uso de algoritmos em processo seletivo de emprego pode prejudicar candidatos. Extra. 2020. Disponível em: https://extra.globo.com/economia/emprego/uso-de-algoritmos-em-processo-seletivo-de-emprego-pode-prejudicar-candidatos-24643581.html. Acesso em: 13 jun. 2022.

CARELLI, Rodrigo de Lacerda. O caso Uber e o controle por programação: de carona para o século XIX. *In:* LEME, Ana Carolina Paes; RODRIGUES, Bruno Alves; CHAVES JÚNIOR, José Eduardo Resende (Coord.). **Tecnologias disruptivas e a exploração do trabalho humano:** a intermediação de mão de obra a partir das plataformas eletrônicas e seus efeitos jurídicos e sociais. São Paulo: LTr, 2017.

CARPANEZZI, Leonardo *et al*. História e evolução da mecanização. **Revista Cientifica Eletrônica Agronomia,** Garça, v. 1, n. 25, p. 45-51, 2018. Disponível em: http://faef.revista.inf.br/imagens_arquivos/arquivos_destaque/CxbNYOvf8fSKep0_2018-1-25-14-45-46.pdf. Acesso em: 13 jun. 2022.

CASTELLS, Manuel. **A sociedade em rede**. v. 1. ed. 2. São Paulo: Paz e Terra, 1999.

CHAVES, Alaor. (org.) **Ciência para prosperidade**: sustentável e socialmente justa. Belo horizonte: EMBRAPII, 2022.

CITRON, Danielle Keats. Technological Due Process. **Whashigton University Law Review**, v. 85, n. 6, 2008. Disponível em:

https://openscholarship.wustl.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1166&context=law_lawreview. Acesso em: 10 out. 2023.

CITRON, Danielle Keats; PASQUALE, Frank A. The Scored Society: Due Process for Automated Predictions. **Washington Law Review**, n. 1, 2014, p. 2-27. Disponível em: https://digitalcommons.law.umaryland.edu/fac_pubs/1431/. Acesso em: 04 fev. 2024.

CÓDIGO DE ÉTICA DA ASSOCIAÇÃO DE MÁQUINAS DE COMPUTAÇÃO. Association for Computing Machinery. 2018. Disponível em: https://www.acm.org/code-of-ethics. Acesso em: 04 jan. 2024.

COMITÉ ECONÓMICO E SOCIAL EUROPEU. Parecer INT/806. Inteligência artificial – Impacto no mercado único (digital), na produção, no consumo, no emprego e na sociedade (parecer de iniciativa). Relatora: Catelijne Muller, 2017. Disponível em: https://webapi2016.eesc.europa.eu/v1/documents/EESC-2016-05369-00-00-AC-TRA-PT.docx/content. Acesso em: 20 jan. 2021.

CONSELHO NACIONAL DE JUSTIÇA (CNJ). *Resolução nº 332, de 21 de agosto de 2020*. Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências. 2020. Disponível em: https://atos.cnj.jus.br/files/original191707202008255f4563b35f8e8.pdf. Acesso em: 15 ago. 2024.

CONSTANTIOU, Ioanna; KALLINIKOS, Jannis. New games, new rules: big data and the changing context of strategy. **Journal of Information Technology**, v. 30, n. 1. [S.l.] 2015. Disponível em: https://eprints.lse.ac.uk/63017/1/Kallinikos_New%20Games%20New%20Rules.pdf . Acesso em 10 out. 2023.

COPEETI, Rafael; MIRANDA, Marcel Andreata de. Autodeterminação Informativa e Proteção de Dados: Uma Análise Crítica da Jurisprudência Brasileira. **Revista de Direito, Governança e Novas Tecnologias**. Minas Gerais, v. 1, n. 2, p. 28-48, jul./dez. 2015. Disponível em: https://indexlaw.org/index.php/revistadgnt/article/view/46. Acesso em: 10 jan. 2024.

CRAWFORD, Kate. Artificial Intelligence's White Guy Problem. The New York Times. 2016. Disponível em: https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html. Acesso em: 17 out. 2023.

CUMMINGS, Mary. **Supervising automation**: humans on the loop. Aero-Astro Magazine Highlight: MIT Department of Aeronautics and Astronautics. 2008. Disponível em: http://web.mit.edu/aeroastro/news/magazine/aeroastro5/cummings.html. Acesso em: 02 fev. 2024.

DE MAURO, Andrea; GRECO, Marco; GRIMALDI, Michele. A formal definition of Big Data based on its essential features. **Library Review**, v. 65 n. 3, p. 122-135. Disponível em: https://doi.org/10.1108/LR-06-2015-0061. Acesso em: 10 jul. 2024.

DIAKOPOULOS, Nicholas. Algorithm accountability: journalistic investigation of computational power structures. **Digital Journalism**, v. 3, n. 3, 2015.

DIVINO, Sthefano Bruno Santos; MAGALHÃES, Rodrigo Almeida. Inteligência Artificial e Direito Empresarial: Mecanismos de Governança Digital para Implementação e Confiabilidade. **Revista dos Tribunais Online**, [s.l.], v. 1021, p. 191-212, 2020. Disponível em: https://www.thomsonreuters.com.br/content/dam/ewp-m/documents/brazil/pt/pdf/other/rt-1021-inteligencia-artificial-e-direito-empresarial-mecanismos-de-governanca-digital-para-implementacao.pdf. Acesso em: 04 fev. 2024.

DOMINGOS, Pedro. **O Algoritmo Mestre.** Novatec: São Paulo, 2017.

DONEDA, Danilo; ALMEIDA, Virgílio A. F. O que é governança de algoritmos. *In:* BRUNO, Fernanda *et al*. **Tecnopolíticas da vigilância:** perspectivas da margem. São Paulo: Boitempo, 2018.

DOURADO, Daniel de Araújo; AITH, Fernando Mussa Abujamra. A regulação da inteligência artificial na saúde no Brasil começa com a Lei Geral de Proteção de Dados Pessoais. **Revista Saúde Pública**, [s.l.], v. 56, n. 80. 2022. Disponível em: https://www.scielo.br/j/rsp/a/k38jGvJdbQSYN4MpzGZpfXw/?format=pdf&lang=pt. Acesso em: 3 mar. 2024.

ELIAS, Paulo Sá. Algoritmos, inteligência artificial e o direito. Consultor Jurídico, 2017. Disponível em: https://www.conjur.com.br/dl/algoritmos-inteligencia-artificial.pdf. Acesso em: 04 fev. 2021.

EUBANKS, Virginia. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. **Law Technology and Humans**, v. 1, n. 1. 2019. Disponível em: https://www.researchgate.net/publication/337578410_Virginia_Eubanks_2018_Automating_Inequality_How_High-Tech_Tools_Profile_Police_and_Punish_the_Poor_New_York_Picador_St_Martin%27s_Press. Acesso em: 20 jun. 2024.

EUROPEAN ECONOMIC AND SOCIAL COMMITTEE. Artificial Intelligence: Europe needs to take a human-in-command approach, says EESC. Press Release, n. 27, 2017. Disponível em: https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fwww.eesc.europa.eu%2Fsites%2Fdefault%2Ffiles%2Fresources%2Fdocs%2Fcp-27-artificial-intelligence.docx&wdOrigin=BROWSELINK. Acesso em: 04 jan. 2024.

EUROPEAN PARLIMENT RESEARCH SERVICE. A governance framework for algorithmic accountability and transparency. Scientific Foresight Unit, 2019. Disponível em: https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf. Acesso em: 10 out. 2023.

FABRÈGUE, Brian F. G.; BOGONI, Andréa. Privacy and Security Concerns in the Smart City. **Smart Cities**, v. 6, p. 586–613, 2023. Disponível em: https://doi.org/10.3390/smartcities6010027. Acesso em: 20 jun. 2024.

FACEBOOK. Why Am I Seeing This? We Have an Answer for You. Meta. 2019. Disponível em: https://news room.fb.com/news/2019/03/why-am-i-seeing-this/. Acesso em: 05 jan. 2024.

FERRARI, Isabela. Accountability de algoritmos: a falácia do acesso ao código e caminhos para uma explicabilidade efetiva. **Inteligência Artificial:** 3º Grupo de Pesquisa do ITS, ITS - Instituto de Tecnologia e Sociedade do Rio, 2018. Disponível em: https://itsrio.org/wpcontent/uploads/2019/03/Isabela-Ferrari.pdf. Acesso em: 10 jul. 2024.

FERREIRA, Francisco de Paula. Implicações sociais da automação. **Revista de Administração de Empresas**. São Paulo, v. 4, n. 13, p. 45-61.1964. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-75901964000400002&lng=en&nrm=iso. Acesso em: 04 fev. 2024.

FONSECA FILHO, Cléuzio. **História da computação**: O Caminho do Pensamento e da Tecnologia. Porto Alegre: EDIPUCRS, 2007. Disponível em: http://www.pucrs.br/edipucrs/online/historiadacomputacao.pdf. Acesso em: 04 fev. 2024.

FONTOURA, Paula Renata. Alan Turing, o pai da computação. Invivo – Museu da vida. 2021. Disponível em: https://www.invivo.fiocruz.br/historia/alan-turing-o-pai-da-computacao/#:~:text=Considerado%20o%20pai%20da%20computa%C3%A7%C3%A3o,ideia%20do%20que%20era%20isso. Acesso em: 04 fev. 2024.

FRAZÃO, Ana. Algoritmos e inteligência artificial. JOTA. 2018. Disponível em: https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/algoritmos-e-inteligencia-artificial-15052018. Acesso em: 18 jan. 2022.

FRAZÃO, Ana. Dados, estatísticas e algoritmos. JOTA. 2017. Disponível em: https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/dados-estatisticas-e-algoritmos-28062017. Acesso em: 18 jan. 2022.

FRAZÃO, Ana. Discriminação algorítmica. JOTA. 2021. Disponível em: https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/discriminacao-algoritmica-3-30062021. Acesso em 18 jan. 2022.

FRAZÃO, Ana. Discriminação algorítmica: ciência dos dados como ação política. JOTA. 2021. Disponível em: https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/discriminacao-algoritmica-ciencia-dos-dados-como-acao-politica-21072021?amp. Acesso em: 06 jul. 2024.

FRAZÃO, Ana; MULHOLLAND, Caitlin. Inteligência Artificial e Direito: Ética, Regulação e Responsabilidade. **Revista dos Tribunais**. São Paulo, 2020.

GEORGE, Gerard; HAAS, Martine R.; PENTLAND, Alex. Big data and management. **Academy of Management Journal**, v. 57, n. 2, p. 321-326. 2014. Disponível em: https://ink.library.smu.edu.sg/lkcsb_research/4621 Acesso em: 05 abr. 2021.

GOMES, Pedro César Tebaldi. Ética e Inteligência Artificial: viés em machine learning. Data Geeks. 2019. Disponível em: https://www.datageeks.com.br/etica-e-inteligencia-artificial/. Acesso em: 10 jan. 2024.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. Deep Learning**.** Cambridge: MIT Press, 2016. Disponível em:

http://imlab.postech.ac.kr/dkim/class/csed514_2019s/DeepLearningBook.pdf. Acesso em: 10 jul. 2024.

GOOGLE. Artificial Intelligence at Google: Our Principles, Google AI. 2017. Disponível em: https://ai. google/principles/. Acesso em: 05 jan. 2024.

GRGIĆ-HLAČA, Nina *et al.* **Human Perceptions of Fairness in Algorithmic Decision Making:** A Case Study of Criminal Risk Prediction. *In:* WWW 2018: The 2018 Web Conference, 2018. p. 903-912. Disponível em: https://doi.org/10.1145/3178876.3186138. Acesso em: 03 fev. 2024.

HARARI, Yuval Noah. **21 lições para o século 21**. São Paulo: Companhia das Letras, 2018.

HARARI, Yuval Noah. **Homo Deus:** Uma breve história do amanhã. 3ª reimpressão, São Paulo: Companhia das Letras, 2016.

HAWKING, Stephen. Discurso na Web Summit, Lisboa. O observador. 2017. Disponível em: https://observador.pt/2018/03/14/e-um-lugar-excitante-para-estar-e-voces-sao-os-pioneiros-tudo-o-que-stephen-hawking-disse-a-web-summit-palavra-por-palavra/. Acesso em: 10 set. 2023.

IBM Developer Blog. Wha tis big data? More than volume, velocity and Variety. 2017. Disponível em: https://developer.ibm.com/blogs/what-is-big-data-more-than-volume-velocity-and-variety. Acesso em: 10 set. 2023.

INTERNATIONAL TELECOMUNICATION UNION -T. New ITU standards define the Internet of Things and provide the blueprints for its development. [S.l.]: ITU, 2012. Disponível em: https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=11559&lang=en. Acesso em: 03 jan. 2024.

JUNQUEIRA, Thiago. Discriminação: o desafio da inteligência artificial em processos seletivos. Veja Negócios. 2020. Disponível em: https://veja.abril.com.br/economia/discriminacao-o-desafio-da-inteligencia-artificial-em-processos-seletivos/. Acesso em: 10 out. 2023.

KAHNEMAN, Daniel; SUNSTEIN, Cass; SIBONY, Olivier. **Ruído:** Uma falha no julgamento humano. São Paulo: Objetiva, 2021.

KAMINSKI, Margot E.; URBAN, Jennifer M. The Right to Contest AI. **Columbia Law Review**, v. 121, n. 7, 2021. Disponível em: https://ssrn.com/abstract=3965041. Acesso em: 13 jan. 2024.

KAUFMAN, Dora. O protagonismo dos algoritmos da Inteligência Artificial: observações sobre a sociedade de dados. Teccogs: **Revista Digital de Tecnologias Cognitivas**, São Paulo, n. 17, p. 44-58, jan-jun. 2018. Disponível em: https://revistas.pucsp.br/index.php/teccogs/article/view/48589/32069. Acesso em: 05 set. 2022.

KLEINBERG, Jon. The Mathematics of Algorithm Design. Cornell University, Ithaca, 2008. Disponível em: https://www.cs.cornell.edu/home/kleinber/pcm.pdf. Acesso em: 10 set. 2023.

KNIGHT, Will. Forget Killer Robots – Bias Is the Real AI Danger. **MIT Technology Review**. 2017. Disponível em: https://www.technologyreview.com/s/608986/forget-killer-robotsbias-is-the-real-ai-danger. Acesso em: 28 jan. 2024.

KOCHENDERFER, Mykel J.; WHEELER, Tim A.; WRAY, Kyle H. **Algorithms for decision making**. Cambridge: Massachusetts Institute of Technology, 2022. Disponível em: https://algorithmsbook.com/files/dm.pdf. Acesso em: 04 jan. 2024.

KROLL, Joshua A. *et al.* Accountable Algorithms. Forthcoming, **Fordham Law Legal Studies Research Paper**, v. 165, n. 2765268. University of Pennsylvania Law Review, 2017. Disponível em: https://ssrn.com/abstract=2765268. Acesso em: 05 fev. 2024.

LEE, Kai-Fu. **Inteligência artificial:** como os robôs estão mudando o mundo, a forma como amamos, nos relacionamos, trabalhamos e vivemos. 1 ed. Rio de Janeiro: Globo Livros, 2019.

LIPOVETSKY, Gilles. **O império do efêmero:** a moda e seu destino nas sociedades modernas. Tradução Maria Lucia Machado. 7 ed. São Paulo: Companhia das Letras, 2004.

LOPES, Alexandra Krastins; MORAES, Thiago Guimarães; PEREIRA, José Renato Laranjeira. A (ausência da) intervenção humana na revisão de decisões automatizadas. JOTA. 2019. Disponível em: https://www.jota.info/opiniao-e-analise/artigos/a-ausencia-da-intervencao-humana-na-revisao-de-decisoes-automatizadas-13102019. Acesso em: 10 set. 2023.

LOPES, Giovana Figueiredo Peluso. LGPD e Revisão de Decisões Automatizadas. DTIBR. 2019. Disponível em: https://www.dtibr.com/post/lgpd-e-revis%C3%A3o-de-decis%C3%B5es-automatizadas. Acesso em: 10 out. 2023.

LUZ, Gilberto Barbosa; KUIAWINSKI, Darci Luíz. Mecanização, Autonomação e Automação – Uma Revisão Conceitual e Crítica. **XIII Simpósio de Engenharia de Produção** – SIMPEP. Bauru, 2006. Disponível em: https://simpep.feb.unesp.br/anais/anais_13/artigos/1210.pdf. Acesso em: 26 jun. 2020.

MAGRANI, Eduardo. **Entre dados e robôs:** ética e privacidade na era da hiperconectividade. 2 ed. - Porto Alegre: Arquipélago Editorial, 2019.

MAGRANI, Eduardo. New perspectives on ethics and the laws of artificial intelligence. **Internet Policy Review**, v. 8, n. 3. 2019. Disponível em: https://doi.org/10.14763/2019.3.1420. Acesso em: 10 set. 2023.

MAIA, Deborah Vieira de Alencar. **Automação Industrial e Robótica**. Programa de Pós-graduação em Engenharia Elétrica – Universidade federal do Rio Grande do Norte. Natal, 2008. Disponível em: http://professor.pucgoias.edu.br/SiteDocente/admin/arquivosUpload/17829/material/ARTIGO08.pdf. Acesso em: 10 jan. 2024.

MARRAFON, Marco Aurélio, MEDON, Filipe. Importância da revisão humana das decisões automatizadas na Lei Geral de Proteção de Dados. Consultor Jurídico. 2019. Disponível em:

https://www.conjur.com.br/2019-set-09/constituicao-poder-importancia-revisao-humana-decisoes-automatizadas-lgpd. Acesso em: 10 set. 2023.

MARTINEZ, Luciano; MALTEZ, Mariana. O direito fundamental à proteção em face da automação. **Revista Nova Hileia,** [s.l.], v. 2, n. 2, jan.-jun., 2017. Disponível em: http://periodicos.uea.edu.br/index.php/novahileia/article/view/1240/784. Acesso em: 10 set. 2023.

MARTINS, Pedro Bastos Lobo. **A regulação do profiling na lei geral de proteção de dados**: o livre desenvolvimento da personalidade em face da governamentalidade algorítmica. 2021. Dissertação (Mestrado em Direito) – Universidade Federal de Minas Gerais. Belo Horizonte, 2021. Disponível em: https://repositorio.ufmg.br/bitstream/1843/43900/4/Pedro%20Martins%20-%20Disserta%C3%A7%C3%A3o%20-%20A%20REGULA%C3%87%C3%83O%20DO%20PROFILING%20NA%20LEI%20GERAL%20DE%20PROTE%C3%87%C3%83O%20DE%20DADOS%20o%20livre%20desenvolvimento%20da%20personalidade%20em%20face%20da%20governamentalidade%20algor%C3%ADtmica.pdf. Acesso em: 20 ago. 2022.

MASHAW, Jerry L. Administrative Due Process: The Quest for a Dignitary Theory. **Boston University Law Review**, v. 61, n. 885, 1981. Disponível em: https://core.ac.uk/download/pdf/72827487.pdf. Acesso em: 04 fev. 2024.

MATSUURA, Sérgio. Como nova Lei de Proteção de Dados fortalece a ditadura dos algoritmos. O Globo – Época. 2019. Disponível em: https://epoca.globo.com/sociedade/como-nova-lei-de-protecao-de-dados-fortalece-ditadura-dos-algoritmos-23802395. Acesso em: 17 out. 2020.

MEDINA, Marco; FERTING, Cristina. **Algoritmos e Programação:** Teoria e Prática. São Paulo: Novatec Editora, 2006.

MEDON, Filipe. Tendências para a responsabilidade civil da Inteligência Artificial na Europa: a participação humana ressaltada. Migalhas. 2020. Disponível em: https://www.migalhas.com.br/coluna/migalhas-de-responsabilidade-civil/335801/tendencias-para-a-responsabilidade-civil-da-inteligencia-artificial-na-europa--a-participacao-humana-ressaltada. Acesso em: 20 set. 2020.

MENÁRGUEZ, Ana Torres. Os privilegiados são analisados por pessoas; as massas, por máquinas. El País. 2018. Disponível em: https://brasil.elpais.com/brasil/2018/11/12/tecnologia/1542018368_035000.html. Acesso em: 10 set. 2023.

MENDES, Laura Schertel. Habeas data e autodeterminação informativa. Os dois lados de uma mesma moeda. **Direitos Fundamentais & Justiça**, [s.l.], v. 39, 2018.

MIRANDA, Paula. O que acontece nos bastidores da Inteligência Artificial. Asksuite. 2019. Disponível em: https://asksuite.com/br/blog/como-funciona-inteligencia-artificial/. Acesso em: 08 mar. 2020.

MORAES, Everton. Automação não é Mecanização. Sala da automação. 2013. Disponível em: http://saladaautomacao.com.br/automacao-nao-e-mecanizacao/. Acesso em: 15 jun. 2022.

NAVARRO, Marcus Vinícius Teixeira. Conceito e controle de riscos à saúde. *In:* NAVARRO, Marcus Vinícius Teixeira. **Risco, radiodiagnóstico e vigilância sanitária**. Salvador: EDUFBA, 2009.

NIKLAS, Jędrzej. SZTANDAR-SZTANDERSKA, Karolina. SZYMIELEWICZ, Katarzyna. Profiling the unemployed in poland: social and political implications of algorithmic decision making. Fundacja Panoptykon. Warsaw, 2015. Disponível em: https://panoptykon.org/sites/default/files/leadimage-biblioteka/panoptykon_profiling_report_final.pdf. Acesso em: 10 set. 2023.

NYBO, Erik Fontenele. **O Poder dos Algoritmos**. São Paulo: Enlaw, 2019.

O'NEIL, Cathy. **Algoritmos de destruição em massa**: como o big data aumenta a desigualdade e ameaça a democracia. 1 ed. São Paulo: Editora Rua do Sabão, 2020.

OECD LEGAL INSTRUMENTS. Recommendation of the Council on Artificial Intelligence. 2019. Disponível em https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Acesso em: 13 jan. 2024.

OTTERLO, Martijn van. A machine learning view on profiling. *In:* HILDEBRANDT, Mireille; VRIES, Katja. **Privacy, Due process and the Computational Turn:** The Philosophers of Law meet Philosophers of Technology. London: Routledge, 2013. Disponível em: https://www.taylorfrancis.com/chapters/edit/10.4324/9780203427644-4/machine-learning-view-profiling-martijn-van-otterlo. Acesso em: 20 ago. 2022.

PARRA, Henrique Zoqui Martins. Experiências com tecnoativistas: resistências na política do dividual? *In:* BRUNO, Fernanda *et al*. **Tecnopolíticas da vigilância**: perspectivas da margem. São Paulo: Boitempo, 2018.

PASQUALE, Frank. **The Black Box Society:** The Secret Algorithms That Control Money and Information. Cambridge, Harvard University Press, 2015.

PÉREZ LUÑO, Antonio-Enrique. **Los derechos fundamentales**. Imprenta: Madrid, Tecnos, 1988.

PIAIA, Thami Covatti. A digitalização dos direitos fundamentais. **Revista de Direitos e Garantias Fundamentais**, [S. l.], v. 22, n. 2, p. 7–8, 2022. DOI: 10.18759/rdgf.v22i2.2079. Disponível em: https://sisbib.emnuvens.com.br/direitosegarantias/article/view/2079. Acesso em: 9 jan. 2024.

RELATÓRIO DA COMISSÃO EUROPEIA. Orientações éticas para uma IA de confiança. European Commission. 2019. Disponível em: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. Acesso em: 13 jan. 2024.

REQUIÃO, Maurício; COSTA, Diego Carneiro. Discriminação algorítmica: ações afirmativas como estratégia de combate. **Civilistica.com**, Rio de Janeiro, v. 11, n. 3, p. 1-24, 2022.

Disponível em: https://civilistica.emnuvens.com.br/redc/article/view/804. Acesso em: 07 jul. 2024.

REQUIÃO, Maurício. (no prelo). Inteligência artificial, vieses cognitivos e decisões judiciais.

REVISTA SUPERINTERESSANTE. Alan Turing O pai das ciências da computação e precursor da inteligência artificial. 2018. Disponível em: https://super.abril.com.br/historia/alan-turing/. Acesso em: 9 jan. 2024.

RIJIMENAM, Mark van. Why We Should Be Careful When Developing AI. Medium. 2019. Disponível em: https://medium.com/dataseries/why-we-should-be-careful-when-developing-ai-8c866914fd8b. Acesso em: 10 set. 2023.

ROSETTI, Adroaldo Guimarães; MORALES, Aran Bey Tcholakian. O papel da tecnologia da informação na gestão do conhecimento. **Ciência da Informação**, Brasília, v. 36, n. 1, p. 124-135, jan./abr., 2007. Disponível em: http://www.scielo.br/pdf/ci/v36n1/a09v36n1.pdf. Acesso em: 10 set. 2023.

ROUVROY, Antoinette. **Of Data and Men**: Fundamental Rights and Freedoms in a World of Big Data. Council of europeu. Strasbourg, 2016. Disponível em: https://rm.coe.int/16806a6020. Acesso em: 03 jan. 2024.

ROUVROY, Antoniette: BERNS, Thomas. Governamentalidade algotítmica e perspecticas de emancipação: o díspar como condição de individuação pela relação. *In:* BRUNO, Fernanda *et al.* **Tecnopolíticas da vigilância**: perspectivas da margem. São Paulo: Boitempo, 2018.

RUBIO, Isabel. Precisamos da inteligência artificial para sobreviver como espécie. El País. 2018. Disponível em: https://brasil.elpais.com/brasil/2018/11/12/tecnologia/1542038734_872245.html. Acesso em: 10 jun. 2022.

RUEL, Renata. Fifa apresenta tecnologia semi-automatizada para impedimento. ESPN. 2020. Disponível em: http://www.espn.com.br/blogs/renataruel/765432_fifa-apresenta-tecnologia-semi-automatizada-para-impedimento. Acesso em: 10 jun. 2022.

SALESFORCE BRASIL. Machine Learning e Deep Learning: aprenda as diferenças. Sales Force. 2018. Disponível em: https://www.salesforce.com/br/blog/2018/4/Machine-Learning-e-Deep-Learning-aprenda-as-diferencas.html. Acesso em: 08 jun. 2022.

SANT´ANA, Maurício Requião. **Autonomia, incapacidade e transtorno mental**: propostas pela promoção da dignidade. 2015. Dissertação (Doutorado em Direito) – Faculdade de Direito, Universidade Federal da Bahia, Salvador, 2015. Disponível em: https://repositorio.ufba.br/bitstream/ri/17254/1/Tese%20Maur%C3%ADcio%20Requi%C3%A3o.pdf. Acesso em: 04 jan. 2024.

SANT´ANA, Maurício Requião. **Estatuto da pessoa com deficiência, incapacidades e interdição**. Salvador: Juspodivm, 2016.

SANTOS, Boaventura de Sousa. Os tribunais e as novas tecnologias de comunicação e de informação. **Sociologias**, Porto Alegre, v. 7, n. 13, jan./jun., 2005, p. 82-109. Disponível em:

http://www.boaventuradesousasantos.pt/media/Tribunais%20e%20novas%20tecnologias_Sociologias_2005(1).pdf. Acesso em: 10 out. 2023.

SANTOS, Johann Ortnau Cirio E. **Responsabilidade civil e Inteligência Artificial**: Uma análise da resolução sobre disposições de direito civil e robótica da União Europeia. 2018. Trabalho de Conclusão de Curso (Graduação em Direito) - Universidade Federal do Rio Grande do Sul. Porto Alegre, 2018. Disponível em: https://lume.ufrgs.br/handle/10183/192797. Acesso em: 08 jun. 2022.

SARLET, Ingo Wolfgan. As aproximações e tensões existentes entre os Direitos Humanos e Fundamentais. Consultor Jurídico. 2015. Disponível em: https://www.conjur.com.br/2015-jan-23/direitos-fundamentais-aproximacoes-tensoes-existentes-entre-direitos-humanos-fundamentais. Acesso em: 23 ago. 2020.

SARLET, Ingo Wolfgang. **A eficácia dos direitos fundamentais**: uma teoria geral dos direitos fundamentais na perspectiva constitucional. 12 ed. rev. atual. e ampl. Porto Alegre: Livraria do Advogado Editora, 2015.

SARLET, Ingo Wolfgang. **A eficácia dos direitos fundamentais**: uma teoria geral dos direitos fundamentais na perspectiva constitucional. 6 ed. Porto Alegre: Livraria do Advogado Editora, 2006.

SARLET, Ingo Wolfgang. **Dignidade da Pessoa Humana e Direitos Fundamentais na Constituição Federal de 1988**. 10 ed. rev. atual. e ampl. Porto Alegre: Livraria do Advogado Editora, 2015.

SARLET, Wolfgang Ingo; MARINONI, Luiz Guilherme; MITIDIERO, Daniel. **Curso de direito constitucional**. 7 ed. – São Paulo: Saraiva Educação, 2018.

SCHWARTZ, Paul. Data Processing and Government Administration: The Failure of the American Legal Response to the Computer. **Hastings Law Journal**, v. 43, n. 5, 1992. p. 1341. Disponível em: https://repository.uclawsf.edu/cgi/viewcontent.cgi?article=3086&context=hastings_law_journal. Acesso em: 10 abr. 2022.

SILBERG, Jake; MANYIKA, James. Como lidar com vieses na inteligência artificial (e nos seres humanos). **Mckinsey & Company**. 2019. Disponível em: https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans/pt-br. Acesso em: 08 jun. 2023.

SILVA, Maurício Prado. Inteligência Artificial: defina limites e reduza o viés. Convergencia Digital. 2020. Disponível em: https://www.convergenciadigital.com.br/cgi/cgilua.exe/sys/start.htm?UserActiveTemplate=site&from%5Finfo%5Findex=51&infoid=54847&sid=15. Acesso em: 08 jun. 2023.

SILVEIRA, Cristiano Bertulucci. Jidoka: Automatização com um toque humano. Citisystems. 2013. Disponível em: https://www.citisystems.com.br/jidoka/. Acesso em: 10 out. 2023.

SOUZA, Carlos Affonso. O debate sobre personalidade jurídica para robôs. JOTA. 2017. Disponível em: https://www.jota.info/opiniao-e-analise/artigos/o-debate-sobre-personalidade-juridica-para-robos-10102017. Acesso em: 10 out. 2023.

SOUZA, Gustavo Henrique Costa. **Análise da relação entre a transparência da inteligência artificial e a tomada de decisões gerenciais**. 2023. Tese (Doutorado em Ciências Contábeis) – Universidade Federal de Pernambuco, Recife, 2023. Disponível em: https://repositorio.ufpe.br/bitstream/123456789/51312/1/TESE%20Gustavo%20Henrique%20Costa%20Souza.pdf. Acesso em: 15 set. 2023.

SPICE, Byron. Questioning the Fairness of Targeting Ads Online. Carnegie Mellon University. 2015. Disponível em: https://www.cmu.edu/news/stories/archives/2015/july/online-ads-research.html. Acesso em: 10 jun. 2022.

STEEL, Katie; STEFÁNSSON, H. Orri. **Decision Theory.** The Stanford Encyclopedia of Philosophy. 2015. Revisado em 2020. Disponível em: https://plato.stanford.edu/entries/decision-theory/. Acesso em: 04 fev. 2024.

STEFANINI. **Machine Learning × Deep Learning:** entenda a diferença. Stefanini. 2019. Disponível em: https://stefanini.com/pt-br/trends/artigos/machine-learning-vs-deep-learning. Acesso em. 10 set. 2023.

STEFANO, Valerio de. "Negotiating the algorithm": Automation, artificial intelligence and labour protection. International Labour office. Employment Policy Department. Working Paper n. 246, Geneva, 2018. Disponível em: https://www.ilo.org/wcmsp5/groups/public/---ed_emp/---emp_policy/documents/publication/wcms_634157.pdf. Acesso em: 10 out. 2023.

TAURION, Cezar. Inteligência artificial: até os algoritmos têm "preconceito". Neofeed. 2020. Disponível em: https://neofeed.com.br/blog/home/inteligencia-artificial-ate-os-algoritmos-tem-preconceito/. Acesso em: 10 out. 2023.

TAVARES, Osny. O uso cotidiano do algoritmo. Gazeta do povo. 2011. Disponível em: https://www.gazetadopovo.com.br/vida-e-cidadania/o-uso-cotidiano-do-algoritmo-4x3n9sw4bkhoam6fzqcp27mfi/. Acesso em: 10 out. 2023.

TEFFÉ, Chiara Spadaccini de; MEDON, Filipe. Responsabilidade civil e regulação de novas tecnologias: questões acerca da utilização de Inteligência Artificial na tomada de decisões empresariais. **Revista Estudos Institucionais**, [s.l.], v. 6, n. 1, p. 301-333, jan./abr. 2020. Disponível em: https://estudosinstitucionais.com/REI/article/view/383/493. Acesso em: 10 out. 2023.

TEPEDINO, Gustavo. SILVA, Rodrigo da Guia. Desafios da Inteligência Artificial em matéria de responsabilidade civil. **Revista Brasileira de Direito Civil** – RBDCivil | Belo Horizonte, v. 21, p. 61-86, jul./set. 2019. Disponível em: https://rbdcivil.ibdcivil.org.br/rbdc/article/download/465/308. Acesso em: 10 out. 2023.

UNIÃO EUROPEIA. Resolução do Parlamento Europeu, processo 2015/2103(INL), de 16 de fevereiro de 2017, com recomendações à Comissão de Direito Civil sobre Robótica, Estrasburgo, 2017.

VALOR ECONÔMICO. Somos experimenta ferramenta que mede atenção do aluno. Valor Econômico, 1 out. 2020. Disponível em: https://valor.globo.com/empresas/noticia/2020/10/01/somos-experimenta-ferramenta-que-mede-atencao-do-aluno.ghtml. Acesso em: 12 ago. 2024.

VERBEEK, Peter. **Moralizing Technology**: Understanding and Designing the Morality of Things, Chicago - London, The University of Chicago Press. 2011. Disponível em: https://philpapers.org/rec/VERMTU. Acesso em: 04 fev. 2024.

VIEIRA, Leonardo Marques. A Problemática da Inteligência Artificial e dos Vieses Algorítmicos: Caso Compas. **Brazilian Technology Symposium**, v. 1. 2019. Disponível em: https://www.lcv.fee.unicamp.br/images/BTSym-19/Papers/090.pdf. Acesso em: 10 out. 2023.

WAKEFIELD, Jane. Inteligência artificial: máquinas que pensam devem surgir 'até 2050'. G1. 2015. Disponível em: http://g1.globo.com/tecnologia/noticia/2015/09/inteligencia-artificial-maquinas-que-pensam-devem-surgir-ate-2050.html. Acesso em: 10 out. 2023.

WANG, Ge. **Humans in the Loop:** The Design of Interactive AI Systems. Human-centered Artificial Intelligence. Stanford University. 2019. Disponível em: https://hai.stanford.edu/news/humans-loop-design-interactive-ai-systems Acesso em: 04 fev. 2024.

WANG, Ye *et al*. Communityin-the-loop: Creating Artificial Process Intelligence for Co-production of City Service. **Proceedings of the ACM on Human Computer Interaction,** v. 6, n. 285. 2022. Disponível em: https://doi.org/10.1145/3555176. Acesso em: 05 fev. 2024.

WEISER, Mark. The computer for the 21st Century. Scientific American. p. 94-10. 1991. Disponível em: https://ics.uci.edu/~corps/phaseii/Weiser-Computer21stCentury-SciAm.pdf. Acesso em: 10 set. 2023.

WEISER, Mark. The **Invisible Interface**: Increasing the Power of the Environment through Calm Technology. *In:* Conference paper, 1998. Disponível em: https://link.springer.com/chapter/10.1007/3-540-69706-3_1. Acesso em: 10 set. 2023.

WORLD ECONOMIC FORUM. The Future of Jobs Report. 2018. Centre for the New Economy and Society, Switzerland, 2018. Disponível em: http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf. Acesso em: 10 out. 2023.

XAVIER, Paulo Ramón Suárez *et al.* Polícia preditiva e "negritude": modelos para a reprodução de um estado sem direitos. Direito. **Revista da Faculdade de Direito,** v. 6, n. 3, p. 99–128. Universidade de Brasília, 2022. Disponível em: https://periodicos.unb.br/index.php/revistadedireitounb/article/view/36383. Acesso em: 09 mar. 2024.

YUSTE, R.; GENSER, J.; HERRMANN, S. It's time for neuro--rights: new human rights for the age of neurotechnology. Horizons J. Int. Relat. Sustain. Dev. v. 18, p. 154–164, 2021. Disponível em: https://www.cirsd.org/en/horizons/horizons-winter-2021-issue-no-18/its-time-for-neuro--rights. Acesso em: 20 jul. 2024.

ZANATTA, Rafael A. F. Proteção de dados pessoais como regulação do risco: uma nova moldura teórica? *In*: **I Encontro da Rede de Pesquisa em Governança da Internet**. [on-line]. 2017. Disponível em: http://www.redegovernanca.net.br. Acesso em: 10 maio 2023.

ZHU, Jichen *et al.* Explainable AI for designers: A human-centered perspective on mixed-initiative co-creation. **IEEE Conference on Computational Intelligence and Games (CIG)**. Netherlands, 2018. Disponível em: https://ieeexplore.ieee.org/abstract/document/8490433. Acesso em: 10 jul. 2024.

ZUBOFF, Shoshana. **A era do Capitalismo de Vigilância**: A luta por um futuro humano na nova fronteira do poder. Rio de Janeiro: Intrínseca, 2021.

ZUBOFF, Shoshana. Big Other: capitalismo de vigilância e perspectivas para uma civilização informação. *In:* BRUNO, Fernanda *et al.* **Tecnopolíticas da vigilância**: perspectivas da margem. São Paulo: Boitempo, 2018.